Local failure detection and diagnosis in large sensor networks

Romain Couillet[1] and Walid Hachem[2]
[1] *Centrale-Supélec-EDF Chair on System Sciences and the Energy Challenge, Gif sur Yvette, France.*
[2] *CNRS-Telecom ParisTech, Paris, France.*

arXiv:1107.1409v1 [cs.IT] 7 Jul 2011

***Abstract*—In this article, the joint fluctuations of the extreme eigenvalues and eigenvectors of a large dimensional sample covariance matrix are analyzed when the associated population covariance matrix is a finite-rank perturbation of the identity matrix, which corresponds to a so-called spiked model in random matrix theory. The asymptotic fluctuations, as the matrix size grows large, are shown to be intimately linked with matrices from the Gaussian unitary ensemble (GUE). When the spiked population eigenvalues have unit multiplicity, the fluctuations follow a central limit theorem. This result is used to develop an original framework for the detection and diagnosis of local failures in large sensor networks, for known or unknown failure magnitude. Simulations are carried out that show the strong performance of the derived failure localization algorithms.**

I. Introduction

In the field of fault detection and diagnosis, one of the elementary requests is the fast, reliable and computationally light identification of a system failure. In dynamical scenarios, these systems are composed of several fluctuating parameters whose evolutions are tracked by a mesh of sensors reporting successive correlated and noisy data measurements to a central decision unit. With the growth in size and complexity of such systems, it becomes increasingly difficult for decision units to process simultaneously and at a low computational cost the augmenting load of reported measurements. Examples of such systems are the recent cognitive radio networks [1] and smart grid technologies [2]. In the former, multiple cooperative wireless communication devices, referred to as the secondary network, exchange sensed data in order to decide collectively which communication bandwidths are left unused by the licensed, also called primary, network users. Fast detection of sudden changes, e.g. new primary user communications, is here demanded to minimize the interference generated by secondary users. In the smart-grid framework, a large dimensional graph of interconnected electricity producers, transportation systems, and consumers evolve in real-time, their behaviour being reported by diverse sensors such as voltage phasor measurements [3] at the nodes of the electricity grid to regional controllers. Fast detection of link and node failures is requested in this scenario to minimize the risk of cascaded failures leading to regional blackouts [4]. There exists a rich literature on failure detection, diagnosis and change-point estimation, ranging from off-line detection methods of uncorrelated data [5], [6] to fast change detection methods in time correlated signals [7], [8], [9]. Subspace methods were in particular proposed to detect changes from changes in the eigenstructure of sampled covariance matrices for dynamical systems [10], [11]. In this article, we propose a novel subspace approach to solve the problem of off-line detection and identification of local failures from independent or linearly time-correlated samples.

We precisely assume the observation of measurements suggesting an error has already occurred in the network. We wish the detection of a failure to be fast so we will assume that the number n of successive sensor data reports is not extremely large with respect to the size N of the network. We will also assume that the hypothetical failure scenarios are, to some extent, known in advance. In this context, calling $\mathcal{H}_0$ the hypothesis that the system does not undergo any failure and $\mathcal{H}_k$, $1 \leq k \leq K$, the hypothesis that a failure of type k occurs, the question of failure detection and localization consists in proceeding to the successive hypotheses tests: (i) decide whether the concatenation matrix of n successive network observations $\Sigma = [s_1, \ldots, s_n] \in \mathbb{C}^{N \times n}$ suggests hypothesis $\mathcal{H}_0$ or its complementary $\bar{\mathcal{H}}_0$ (i.e. the event union of the $\mathcal{H}_k$), and (ii) upon decision of $\bar{\mathcal{H}}_0$, decide what $\mathcal{H}_k$ is the most likely. Both problems are optimally solved by multi-hypothesis Neyman-Pearson tests [12] with maximum likelihood performance given the observation Σ. However, this procedure is computationally intense for large system dimensions.

The approach under consideration here follows the theory of large dimensional random matrices. Precisely, we consider the setting where both N and n grow large and such that $c_N = N/n \to c$, with $0 < c < 1$. Under this assumption, we then develop asymptotic results on the extreme eigenvalues and associated eigenvectors of a certain family of random matrices to provide novel subspace methods for failure detection and localization. Our interest is on random matrices of the type *spiked model*, introduced by Johnstone [13], specifically here of matrices modeled as $\Sigma = (I_N + P)^{\frac{1}{2}} X$, where X is a left-unitarily invariant random matrix and P is a rank-r Hermitian matrix with $r \ll N$. Such matrix models have been largely studied in the recent random matrix literature, very often in the special case where X is a *standard Gaussian matrix*, which refers in this article to a random matrix with independent $\mathcal{CN}(0, 1/n)$ entries. In [14], [15], for X a standard Gaussian matrix, it is first shown that there exists a natural mapping between the extreme (empirical) eigenvalues of $\Sigma\Sigma^*$ and the (population) eigenvalues of P. It is then proved that, almost surely, the extreme empirical eigenvalues converge to deterministic limits in the asymptotic setting, found either at

the edges of the support of the Marčenko-Pastur law [16], i.e. the weak limit of the eigenvalue distribution of XX^*, or away from them, depending on the corresponding population eigenvalues. This induces a phase transition having important consequences on fault detectability in sensor networks [17]. This observation is extended to the non-Gaussian case and generalized to a other spiked models in [18]. The fluctuations of the extreme eigenvalues are studied with different approaches depending on whether the limiting eigenvalues are found at the edge or outside the support of the Marčenko-Pastur law. When at the edge, it is proved successively in [19], [13], [20], [21], [22] that the (centered and scaled) limiting eigenvalue has Tracy-Widom fluctuations. When outside the support, those fluctuations are linked to the distribution of the eigenvalues of GUE matrices, as shown in [21]. In the specific case where the spiked eigenvalues of P have unit multiplicity, the fluctuations are Gaussian.

In this article, the properties on the extreme eigenvalues in a spiked model will be used to provide failure detection tests, in the same line as [23], [24]. For failure localization, pure eigenvalue tests will turn out to be inefficient and we therefore develop novel results on the eigenspaces associated to these eigenvalues. In [25], it is shown, in the real Gaussian case, that the projection of the eigenvectors associated with the extreme empirical eigenvalues of $\Sigma\Sigma^*$ on the subspace of the corresponding population eigenvectors of P has a positive limiting norm, which is close to 1 for c small. This remark is extended in [18] to the non Gaussian case. This property is the fundamental basis of our novel failure diagnosis method. However, the fluctuations of the eigenvector projections, fundamental here to derive test statistics for failure localization, have never been derived before for either the Gaussian or the non-Gaussian cases. The main mathematical result of this article, Theorem 4, is to derive the joint fluctuations of the eigenvalues and eigenspace projections for the eigenvalues found away from the limiting support of XX^*. Our proof technique is largely inspired by [18], [26], [27]. We also use some tools from [28] and [29]. Based on these results, we provide in this article a novel framework for local failure detection and identification in large sensor networks.

The remainder of this document unfolds as follows. Section II introduces elementary examples of sensor networks for which local failures translate simply into small rank perturbations of the identity matrix. Section III reminds important notions of random matrix theory and introduces the main mathematical results of this article. Practical application algorithms along with simulations are then carried out in Section IV. Finally, Section V concludes the article.

Notations: In this document, capital characters stand for matrices while lowercase characters stand either for scalars or vectors, with $I_N \in \mathbb{C}^{N\times N}$ the identity matrix. The i^{th} entry of a vector x is denoted $x(i)$. The symbol $(\cdot)^*$ denotes complex transpose. For a function f and a Hermitian matrix $X \in \mathbb{C}^{N\times N}$, $f(X) = U\,\mathrm{diag}(f(\lambda_1(X)),\ldots,f(\lambda_N(X)))U^*$ with $\lambda_1(X),\ldots,\lambda_N(X)$ the eigenvalues of X and $U \in \mathbb{C}^{N\times N}$ the unitary matrix of its respective eigenvectors. The symbol S_π denotes the support of the probability measure π. The notation $\mathrm{Span}(u_1,\ldots,u_k)$ designates the space generated by the vectors $u_1,\ldots,u_k$. The notation $S^\perp$ is the space orthogonal to S. We denote $\mathbb{C}^+ = \{z \in \mathbb{C},\ \Im(z) > 0\}$. The norm $\|X\|$ of a Hermitian matrix X is understood as the spectral norm, and the norm $\|x\|$ of a vector x is understood as the Euclidean norm. The notations '$\xrightarrow{\text{a.s.}}$', '$\Rightarrow$', and '$\xrightarrow{\mathbb{P}}$' denote convergence almost sure, weak, and in probability, respectively. The symbol $\mathrm{E}[\cdot]$ denotes expectation. The notation $1_A(x)$ is the indicator function of the set A.

II. Detection and localization of local failures

To motivate the study of the fluctuations of extreme eigenvalues and eigenvectors of sample covariance matrices in the context of local failures in large dimensional sensor networks, we introduce in the following two basic examples of sensor network failure scenarios, which can all be modeled as small rank perturbations of the identity matrix.

A. Node failure

Consider the following model

$$y = H\theta + \sigma w \tag{1}$$

where $H \in \mathbb{C}^{N\times p}$ is deterministic, $\theta = [\theta(1),\ldots,\theta(p)]^\mathsf{T} \in \mathbb{C}^p$, $w \in \mathbb{C}^N$ have independent and identically distributed (i.i.d.) complex standard Gaussian entries, and $\sigma > 0$. We denote $y = [y(1),\ldots,y(N)]^\mathsf{T} \in \mathbb{C}^N$. In a sensor network composed of N nodes, y represents the observation through the channel H of the vector θ, constituted of centered and normalized independent Gaussian system parameters,[1] impaired by white Gaussian noise. Therefore, $\mathrm{E}[yy^*] = HH^* + \sigma^2 I_N \triangleq R$.

In case of failure of sensor k, $y(k)$, the k^{th} entry of y, will start suddenly to return inconsistent noisy outputs, decorrelated from the rest of the network. Assuming this noise Gaussian with zero mean and variance σ_k^2 and denoting y' the observations of the network with failure at sensor k, we can write

$$y' = (I_N - e_k e_k^*)H\theta + \sigma_k e_k e_k^* \theta' + \sigma w$$

where θ' is distributed like θ and $e_k \in \mathbb{C}^N$ is such that $e_k(k) = 1$ and $e_k(i) = 0$, for all $i \neq k$.

Therefore, y' is Gaussian (as the sum of Gaussian variables) with zero mean and variance

$$\mathrm{E}[y'y'^*] = (I_N - e_k e_k^*)HH^*(I_N - e_k e_k^*) + \sigma_k^2 e_k e_k^* + \sigma^2 I_N.$$

Denoting $s = R^{-\frac{1}{2}} y'$, we have

$$\mathrm{E}[ss^*] = I_N - R^{-\frac{1}{2}} HH^* e_k e_k^* R^{-\frac{1}{2}} + R^{-\frac{1}{2}} e_k \left[(e_k^* HH^* e_k + \sigma_k^2) e_k^* R^{-\frac{1}{2}} - e_k^* HH^* R^{-\frac{1}{2}} \right].$$

Therefore, the population covariance matrix $\mathrm{E}[ss^*]$ is a perturbation of the identity matrix by

$$P_k \triangleq R^{-\frac{1}{2}} e_k \left[(e_k^* HH^* e_k + \sigma_k^2) e_k^* R^{-\frac{1}{2}} - e_k^* HH^* R^{-\frac{1}{2}} \right] - R^{-\frac{1}{2}} HH^* e_k e_k^* R^{-\frac{1}{2}}. \tag{2}$$

[1] Up to a right-product of H by a positive diagonal matrix, the variance of the entries of θ can be assumed all equal to one without loss of generality.

Notice that the image of P_k is included in the subspace $\mathrm{Span}(R^{-\frac{1}{2}}e_k, R^{-\frac{1}{2}}HH^*e_k)$ and is therefore at most of dimension two. Generalizing the above to M node failures at nodes $k_1, \ldots, k_M$, the vector s is now such that

$$\mathrm{E}[ss^*] = I_N - R^{-\frac{1}{2}}HH^*EE^*R^{-\frac{1}{2}} + R^{-\frac{1}{2}}E\left[(E^*HH^*E + \Lambda^2)E^*R^{-\frac{1}{2}} - E^*HH^*R^{-\frac{1}{2}}\right]$$

with $E = [e_{k_1}, \ldots, e_{k_M}]$, $\Lambda = \mathrm{diag}(\sigma_{k_1}, \ldots, \sigma_{k_M})$, where now (2) becomes

$$P_{k_1,\ldots,k_M} \triangleq R^{-\frac{1}{2}}E\left[(E^*HH^*E + \Lambda^2)E^* - E^*HH^*\right]R^{-\frac{1}{2}} - R^{-\frac{1}{2}}HH^*EE^*R^{-\frac{1}{2}} \quad (3)$$

which is a matrix of rank at most $2M$.

B. Sudden parameter change

Consider again the elementary model of (1) and now assume that, instead of a sensor failing, $\theta(k)$, the k^{th} entry of θ, experiences a sudden change in mean and variance. The resulting observation y' can be modeled as

$$y' = H(I_p + \alpha_k e_k e_k^*)\theta + \mu_k He_k + \sigma w$$

for some real parameters μ_k, α_k, and where $e_k \in \mathbb{C}^p$ is defined by $e_k(k) = 1$ and $e_k(i) = 0$, $i \neq k$. For this particular model, we may or may not suppose that μ_k and α_k are a priori known to the experimenter in the sense that a failure on parameter $\theta(k)$ always leads to a change in mean and standard deviation linked to μ_k, α_k. In this scenario, we now have that y' is Gaussian with zero mean and variance

$$\mathrm{E}[y'y'^*] = H(I_p + [\mu_k^2 + (1+\alpha_k)^2 - 1]e_k e_k^*)H^* + \sigma^2 I_N.$$

Denoting $R = HH^* + \sigma^2 I_N$ as in the previous example and taking $s = R^{-\frac{1}{2}}y'$, we finally have

$$\mathrm{E}[ss^*] = I_N + [\mu_k^2 + (1+\alpha_k)^2 - 1]R^{-\frac{1}{2}}He_k e_k^* H^* R^{-\frac{1}{2}}$$

which is here a rank-1 perturbation of the identity matrix by the matrix

$$P_k \triangleq \beta_k R^{-\frac{1}{2}}He_k e_k^* H^* R^{-\frac{1}{2}}$$

with $\beta_k = \mu_k^2 + (1+\alpha_k)^2 - 1$. Note that, in this scenario, the eigenvector of P_k associated with the non-zero eigenvalue is independent of μ_k and α_k. For practical applications, this has the interesting advantage that simple localization can be performed even if μ_k and α_k are unknown. This is further discussed in Section IV.

The derivation above generalizes to sudden changes of multiple parameters. If the means and variances for the sensors $k_1, \ldots, k_M$ are modified simultaneously with respective parameters $\mu_{k_1}, \ldots, \mu_{k_M}$ and $\alpha_{k_1}, \ldots, \alpha_{k_M}$, then

$$\mathrm{E}[ss^*] = I_N + R^{-\frac{1}{2}}HE\Lambda E^*H^*R^{-\frac{1}{2}}$$

with $E = [e_{k_1}, \ldots, e_{k_M}]$ and $\Lambda = \mathrm{diag}(\beta_{k_1}, \ldots, \beta_{k_M})$, $\beta_{k_i} = \mu_{k_i}^2 + (1+\alpha_{k_i})^2 - 1$, which is a rank-$M$ perturbation of the identity matrix by the matrix

$$P_{k_1,\ldots,k_M} = R^{-\frac{1}{2}}HE\Lambda E^*H^*R^{-\frac{1}{2}}.$$

Note that, contrary to the one-dimensional case, the eigenvectors of $P_{k_1,\ldots,k_M}$ depend here explicitly on the parameters μ_{k_i} and α_{k_i}.

In the following section, we introduce the novel detection and localization framework, making reference, whenever possible, to the examples described in this section.

C. Detection and localization

For either of the models above, let us assume a general scenario with K possible failure events, identified by the index $1 \leq k \leq K$ and let now $s_1, \ldots, s_n$ be n successive independent observations of the random variable s. Denote then $\Sigma \triangleq \frac{1}{\sqrt{n}}[s_1, \ldots, s_n] \in \mathbb{C}^{N\times n}$. From the fact that s is Gaussian with zero mean and covariance $(I_N + P_k)$ for a certain k, we can write

$$\Sigma = (I_N + P_k)^{\frac{1}{2}}X$$

where $X \in \mathbb{C}^{N\times n}$ is a given matrix with independent Gaussian entries of zero mean and variance $1/n$. We also denote for simplicity $P_0 = 0$ for the extra event $k = 0$ corresponding to the no-failure scenario.

The natural approach to detect and identify a failure event in a sensor network upon the observations $s_1, \ldots, s_n$ is to systematically perform a maximum likelihood test on the $K+1$ hypotheses $\mathcal{H}_0, \ldots, \mathcal{H}_K$, with $\mathcal{H}_k$ defined as the event $s \sim \mathcal{CN}(0, I_N + P_k)$. However, this optimal approach is in general computationally intensive, especially when N and K are large. Since the node failure information is carried by the perturbation matrix P_k, we provide in the following an improved maximum likelihood test relying on the properties linking P_k to the observation Σ, as the system dimensions (N, n) grow large. Precisely, based on recent advances in the field of large dimensional random matrix theory [30], we will provide a two-step approach to successively (i) decide on the existence of a failure and (ii) identify the failure event. This technique relies on the statistics of the extreme eigenvalues and eigenspace projections of $\Sigma\Sigma^*$, given the perturbation matrix P_k. The following section is dedicated to the study of those statistics as the dimensions of the matrix Σ grow large. The application of those results to failure diagnosis in sensor networks is deferred to Section IV.

III. Main results

The derivation arguments found in this section follow the ideas of [31], [26] and [27]. In our proofs, we shall also borrow some of the arguments of [29] whose context is close to ours.

A. Notations, assumptions and basic results

We start by summarizing the major notations and facts needed here. We consider a generic small rank perturbation model and define

$$\Sigma = (I_N + P)^{\frac{1}{2}}X$$

with $X \in \mathbb{C}^{N\times n}$ left-unitarily invariant, and where the rank-r Hermitian matrix P has the spectral factorization $P = U\Omega U^*$

with

$$\Omega = \begin{bmatrix} \omega_1 I_{j_1} & & \\ & \ddots & \\ & & \omega_t I_{j_t} \end{bmatrix}$$

and $\omega_1 > \ldots > \omega_s > 0 > \omega_{s+1} > \ldots > \omega_t > -1$. Of course, $j_1 + \cdots + j_t = r$. We write accordingly $U = \begin{bmatrix} U_1 \cdots U_t \end{bmatrix}$ where $U_i \in \mathbb{C}^{N \times j_i}$. We denote by $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_N$ the eigenvalues of $\Sigma\Sigma^*$. For $i \in \{1, \ldots, s\}$, we let $\mathcal{K}(i) = j_1 + \cdots + j_{i-1}$, taking by convention $j_0 = 0$. For $i \in \{s+1, \ldots, t+1\}$, we let $\mathcal{K}(i) = N - (j_i + \cdots + j_t)$. One of the purposes of this section is to establish an asymptotic relation between ω_i and the $\hat{\lambda}_{\mathcal{K}(i)+\ell}$ for $\ell = 1, \ldots, j_i$ which holds under a condition on ω_i that will be specified. We also denote by $\widehat{\Pi}_i$ the orthogonal projection matrix, when it exists, on the eigenspace of $\Sigma\Sigma^*$ associated with the eigenvalues $\{\hat{\lambda}_{\mathcal{K}(i)+\ell}\}_{\ell=1}^{j_i}$. Similarly, we denote by $\Pi_i = U_i U_i^*$ the orthogonal projection matrix on the eigenspace of P associated with the eigenvalue ω_i. Finally, we denote by $Q(z) = (XX^* - zI_N)^{-1}$ the resolvent of the matrix XX^* and by $\alpha(z) = \frac{1}{N} \operatorname{tr} Q(z)$ its normalized trace, both analytical on $\mathbb{C}^+$.

In the remainder of the paper, we shall consider the asymptotic regime where $n \to \infty$ and $N/n \to c \in (0, 1)$. The notation $n \to \infty$ will henceforth refer to this asymptotic regime.

We now state our basic assumptions:

A1 The probability law of X is invariant by left multiplication by a deterministic unitary matrix.

Thanks to the left unitary invariance of X, $Q(z)$ writes as $Q(z) = W(\Lambda - zI_n)^{-1}W^*$ where Λ is the matrix of eigenvalues of XX^*, W is a unitary random matrix Haar distributed on its unitary group, and W and Λ are independent.

A2 For every $z \in \mathbb{C}^+$, $\alpha(z)$ converges almost surely to a deterministic function $m(z)$ which is the Stieltjes transform[2] of a probability measure π with support $[a, b] \subset (0, \infty)$.

A3 We have $\|XX^*\| \xrightarrow{\text{a.s.}} b$ and $(\|(XX^*)^{-1}\|)^{-1} \xrightarrow{\text{a.s.}} a$.

This last assumption implies in particular that **A2** is satisfied for all $z \in \mathbb{C} \setminus [a, b]$.

The most classical model of a matrix X that satisfies **A1**-**A3** is when X is standard Gaussian, i.e. with independent $\mathcal{C}N(0, 1/n)$ elements, as introduced in the system models of Section II-A and Section II-B. For this model, the limiting probability distribution π is the well known Marčenko-Pastur distribution [16]. Its Stieltjes transform is

$$m(z) = \frac{1}{2zc}\left(1 - c - z + \sqrt{(1-c-z)^2 - 4zc}\right) \quad (4)$$

where the branch of the square root is the one for which $m(z) \in \mathbb{C}^+$ if $z \in \mathbb{C}^+$.

The unitary invariance of X is the basis of the following important lemma, shown in [29] using an inequality of [28] which involves Haar unitary matrices:

Lemma 1: Assume **A1**. Let $u \in \mathbb{C}^N$ and $v \in \mathbb{C}^N$ be two vectors with norm $\|u\| = \|v\| = 1$. Denote by $\sigma(XX^*)$ the eigenvalue spectrum of XX^*. Given $\varepsilon > 0$ and $z \in \mathbb{C} \setminus [a - \varepsilon, b + \varepsilon]$, denote by d_z the distance from z to $[a - \varepsilon, b + \varepsilon]$. Then for any $p > 0$,

$$\mathrm{E}\left\|1_{\sigma(XX^*) \subset [a-\varepsilon, b+\varepsilon]} u^*(Q(z) - \alpha(z)I_N)v\right\|^p \leq \frac{K_p}{d_z^p N^{p/2}}$$

where the constant K_p depends on p only. Similarly, for any $z, z' \in \mathbb{C} \setminus [a - \varepsilon, b + \varepsilon]$, we have

$$\mathrm{E}\left\|1_{\sigma(XX^*) \subset [a-\varepsilon, b+\varepsilon]} u^*\left[Q(z)Q(z') - \frac{\operatorname{tr} Q(z)Q(z')}{N} I_N\right] v\right\|^p \leq \frac{K_p}{d_z^p d_{z'}^p N^{p/2}}.$$

We now start our analysis of the extreme eigenvalues and eigenspace projections of $\Sigma\Sigma^*$ by studying the first order behavior.

B. First order behavior

1) Eigenvalues: Suppose that $x \in \mathbb{R}$ is not an eigenvalue of XX^*. We first write

$$\begin{aligned} &\det(\Sigma\Sigma^* - xI_N) \\ &= \det(I_N + P) \det(XX^* - xI_N + x[I_N - (I_N + P)^{-1}]) \\ &= \det(I_N + P) \det(XX^* - xI_N) \\ &\times \det(I_N + xP(I_N + P)^{-1}(XX^* - xI_N)^{-1}) \end{aligned}$$

after noticing that $I_N - (I_N + P)^{-1} = P(I_N + P)^{-1}$. Therefore, if x is an eigenvalue of $\Sigma\Sigma^*$ but not of XX^*, it must cancel the rightmost determinant. This determinant can be further rewritten

$$\begin{aligned} &\det(I_N + xP(I_N + P)^{-1}Q(x)) \\ &= \det(I_r + x\Omega U^*(I_N + U\Omega U^*)^{-1}Q(x)U). \end{aligned}$$

From the identity $U^*(I_N + U\Omega U^*)^{-1} = (I_r + \Omega)^{-1}U^*$, we then have

$$\det(\Sigma\Sigma^* - xI_N) = \det(I_N + P) \det(XX^* - xI_N) \det(\widehat{H}(x))$$

where $\widehat{H}(z) = I_r + z\Omega(I_r + \Omega)^{-1}U^*Q(z)U$.

Given any $z \in \mathbb{C} \setminus [a, b]$, Assumptions **A1**-**A3** in conjunction with Lemma 1 show that $\widehat{H}(z)$ is defined for all n large, almost surely, and converges almost surely to

$$H(z) = I_r + zm(z)\Omega(I_r + \Omega)^{-1} \quad (5)$$

(take u and v in Lemma 1 as any couple of columns of U, take $p > 2$ and use Borel-Cantelli's lemma [32]). We therefore expect the solutions of the equation $\det H(x) = 0$ which are outside $[a, b]$ to coincide with the limits of the isolated eigenvalues of $\Sigma\Sigma^*$.

Let us now study the behavior of the solutions of this equation. Let $h(x) = xm(x)$ on $\mathbb{R} \setminus [a, b]$. Since $a > 0$, we have

$$h'(x) = (xm(x))' = \int \frac{\lambda}{(\lambda - x)^2} d\pi(\lambda) > 0.$$

The function $h(x)$ is therefore increasing on $\mathbb{R} \setminus [a, b]$ and with limit 0 as $x \to 0$ and -1 as $x \to \infty$. Therefore, for $\omega_i > 0$, (5) leads to

$$h(\rho) + \frac{1 + \omega_i}{\omega_i} = 0 \quad (6)$$

[2]We recall that the Stieltjes transform $m(z)$ of a real measure π is defined for z outside the support of π by $m(z) = \int \frac{1}{\lambda - z} d\pi(\lambda)$.

having a unique real solution ρ_i satisfying $\rho_i > b$ if and only if $h(b^+) + (1+\omega_i)/\omega_i < 0$. When $\omega_i < 0$, (6) has a unique solution $0 < \rho_i < a$ if and only if $h(a^-) + (1+\omega_i)/\omega_i > 0$. We therefore have the following result, for which a rigorous proof is found in [31]:

Theorem 1: Assume **A1**-**A3**. Let p be zero or the maximum index such that $\omega_p > 0$ and $h(b^+) + (1+\omega_p)/\omega_p < 0$. For $i = 1, \ldots, p$, let ρ_i be the unique solution of (6) such that $\rho_i > b$. Then,

$$\hat{\lambda}_{\mathcal{K}(i)+\ell} \xrightarrow{\text{a.s.}} \rho_i$$

for $i = 1, \ldots, p$ and $\ell = 1, \ldots, j_i$, while

$$\hat{\lambda}_{\mathcal{K}(p+1)+1} \xrightarrow{\text{a.s.}} b.$$

Let q be $t+1$ or the minimum index such that $\omega_q < 0$ and $h(a^-) + (1+\omega_q)/\omega_q > 0$. For $i = q, \ldots, t$, let ρ_i be the unique solution of (6) such that $\rho_i < a$. Then,

$$\hat{\lambda}_{\mathcal{K}(i)+\ell} \xrightarrow{\text{a.s.}} \rho_i$$

for $i = q, \ldots, t$ and $\ell = 1, \ldots, j_i$, while

$$\hat{\lambda}_{\mathcal{K}(q)} \xrightarrow{\text{a.s.}} a.$$

In the remainder of the article, the variables $\omega_1, \ldots, \omega_p$ and $\omega_q, \ldots, \omega_t$ satisfying the conditions of Theorem 1 will be said to *satisfy the separation condition*.

When X is standard Gaussian, applying Theorem 1 shows after some simple derivations the following result:

Corollary 1: Consider the setting of Theorem 1. Assume additionally that X is standard Gaussian. Let p be zero or the maximum index for which $\omega_p > \sqrt{c}$ and q be $t+1$ or the minimum index such that $\omega_q < -\sqrt{c}$. Then

$$\hat{\lambda}_{\mathcal{K}(i)+\ell} \xrightarrow{\text{a.s.}} \rho_i = 1 + \omega_i + c\frac{1+\omega_i}{\omega_i} \tag{7}$$

for $i \in \{1, \ldots p, q, \ldots, t\}$, $\ell = 1, \ldots, j_i$, while

$$\hat{\lambda}_{\mathcal{K}(p+1)+1} \xrightarrow{\text{a.s.}} (1+\sqrt{c})^2$$
$$\hat{\lambda}_{\mathcal{K}(q)} \xrightarrow{\text{a.s.}} (1-\sqrt{c})^2.$$

Corollary 1 implies that, for ω_i sufficiently far from zero (either positive or negative) or, equivalently, for c sufficiently small, the spectrum of $\Sigma\Sigma^*$ exhibits j_i eigenvalues outside the support S_π of the Marčenko-Pastur law π which all converge to ρ_i. For failure detection purposes, upon observation of Σ, we may then test the null hypothesis $\Sigma = X$ (call it hypothesis $\mathcal{H}_0$) against the hypothesis $\Sigma = (I_N + P)^{\frac{1}{2}} X$ (call it hypothesis $\bar{\mathcal{H}}_0$), depending on whether eigenvalues of $\Sigma\Sigma^*$ are found outside S_π. Depending on the scenario, for c small enough, it may be that a mere evaluation of the number of eigenvalues outside the support suggests the number of simultaneous failures in the sensor network. This is the case of the two failure scenarios described in Section II-A and Section II-B. However, the information on the extreme eigenvalues of $\Sigma\Sigma^*$, if sufficient for failure detection purposes, is usually not good enough to perform accurate failure localization. This is because different failure scenarios, characterized by different perturbation matrices P, may exhibit very similar eigenvalues. Also, if the failure amplitude is a priori unknown, then eigenvalues are in general irrelevant; see the application Section IV-C. In such scenarios, we then need to consider eigenspaces properties of P. This is the target of the following section.

2) Projections on eigenspaces: Given $i \leq t$, we now assume that ω_i satisfies the separation condition. Given two deterministic vectors $b_1, b_2 \in \mathbb{C}^N$ with bounded Euclidean norms, our purpose is to study the asymptotic behavior of $b_1^* \widehat{\Pi}_i b_2$. We shall show that this bilinear form is simply related with $b_1^* \Pi_i b_2$ in the asymptotic regime.

Our starting point is to express $b_1^* \widehat{\Pi}_i b_2$ as a Cauchy integral [33]. Denoting $\mathcal{C}_i$ a positively oriented contour encompassing only the eigenvalues $\hat{\lambda}_{\mathcal{K}(i)+\ell}$ of $\Sigma\Sigma^*$ for $\ell = 1, \ldots, j_i$, we have

$$\begin{aligned} & b_1^* \widehat{\Pi}_i b_2 \\ &= -\frac{1}{2\pi\imath} \oint_{\mathcal{C}_i} b_1^* (\Sigma\Sigma^* - zI_N)^{-1} b_2 \, dz \\ &= -\frac{1}{2\pi\imath} \oint_{\mathcal{C}_i} b_1^* (I_N + P)^{-\frac{1}{2}} \\ &\times \left[XX^* - zI_N + zP(I_N+P)^{-1} \right]^{-1} (I_N + P)^{-\frac{1}{2}} b_2 \, dz. \end{aligned}$$

Using Woodbury's matrix identity, we have

$$\begin{aligned} & \left[XX^* - zI_N + zP(I_N+P)^{-1} \right]^{-1} \\ &= Q(z) - zQ(z)U \left[I_r + z\Omega(I_r+\Omega)^{-1} U^* Q(z) U \right]^{-1} \\ &\times \Omega(I_r+\Omega)^{-1} U^* Q(z) \\ &= Q(z) - zQ(z)U\widehat{H}(z)\Omega(I_r+\Omega)^{-1}U^*Q(z) \end{aligned}$$

and taking

$$\begin{aligned} \hat{a}_1(z)^* &= z b_1^* (I_N+P)^{-\frac{1}{2}} Q(z) U \\ \hat{a}_2(z) &= \Omega(I_r+\Omega)^{-1} U^* Q(z) (I_N+P)^{-\frac{1}{2}} b_2 \end{aligned}$$

we obtain

$$\begin{aligned} & b_1^* \widehat{\Pi}_i b_2 \\ &= -\frac{1}{2\pi\imath} \oint_{\mathcal{C}_i} b_1^* (I_N+P)^{-\frac{1}{2}} Q(z) (I_N+P)^{-\frac{1}{2}} b_2 \, dz \\ &+ \frac{1}{2\pi\imath} \oint_{\mathcal{C}_i} \hat{a}_1(z)^* \widehat{H}(z)^{-1} \hat{a}_2(z) \, dz. \end{aligned} \tag{8}$$

By Assumption **A3** and Theorem 1, with probability one for all large n, the first term on the right hand side is zero, while the second is equal to

$$\frac{1}{2\pi\imath} \oint_{\gamma_i} \hat{a}_1(z)^* \widehat{H}(z)^{-1} \hat{a}_2(z) dz$$

where γ_i is a deterministic positively oriented circle enclosing only ρ_i among the limits of the isolated eigenvalues specified by Theorem 1, therefore enclosing none of the ρ_j, $j \neq i$. Using Lemma 1 in conjunction with the analyticity properties of the integrand, one can show that $\hat{a}_1(z)^* \widehat{H}(z)^{-1} \hat{a}_2(z)$ converges uniformly to $a_1(z)^* H(z)^{-1} a_2(z)$ on γ_i in the almost sure sense, where

$$\begin{aligned} a_1(z)^* &= z m(z) b_1^* (I_N+P)^{-\frac{1}{2}} U \\ a_2(z) &= m(z) \Omega(I_r+\Omega)^{-1} U^* (I_N+P)^{-\frac{1}{2}} b_2. \end{aligned}$$

It results that $b_1^* \widehat{\Pi}_i b_2 - T_i \xrightarrow{\text{a.s.}} 0$, where

$$T_i \triangleq \frac{1}{2\pi\imath} \oint_{\gamma_i} a_1(z)^* H(z)^{-1} a_2(z) dz.$$

Details can be found in [29] in a similar situation. Let us find the expression of T_i. Noticing that

$$H(z)^{-1} = \sum_{\ell=1}^{t} \frac{1}{1 + zm(z)\frac{\omega_\ell}{1+\omega_\ell}} \mathcal{I}_\ell \tag{9}$$

where

$$\mathcal{I}_\ell = \begin{bmatrix} 0 & & \\ & I_{j_\ell} & \\ & & 0 \end{bmatrix}$$

we obtain

$$\begin{aligned} T_i &= \sum_{\ell=1}^{t} \frac{\omega_\ell}{(1+\omega_\ell)^2} b_1^* \Pi_\ell b_2 \frac{1}{2\pi\imath} \oint_{\gamma_i} \frac{zm^2(z)}{1 + zm(z)\frac{\omega_\ell}{1+\omega_\ell}} dz \\ &= \sum_{\ell=1}^{t} \frac{b_1^* \Pi_\ell b_2}{1+\omega_\ell} \frac{1}{2\pi\imath} \oint_{\gamma_i} \frac{zm^2(z)}{\frac{1+\omega_\ell}{\omega_\ell} + zm(z)} dz. \end{aligned}$$

From the discussion prior to Theorem 1, it is clear that the denominator of the integrand has a zero in the interior $\text{Int}(\gamma_i)$ of the disk delineated by γ_i only when $\ell = i$, and then this zero is simple due to the fact that $h'(x) = (xm(x))'$ never vanishes on $\mathbb{R} \setminus [a, b]$. Applying the residue theorem [33] and observing from (6) that $1/(1+\omega_i) = (1 + h(\rho_i))/h(\rho_i)$, we obtain the following limits.

Theorem 2: Assume **A1-A3**. Given $i \leq t$, assume that ω_i satisfies the separation condition. Let $b_1 \in \mathbb{C}^N$ and $b_2 \in \mathbb{C}^N$ be two sequences of increasing size deterministic vectors with bounded Euclidean norms. Then

$$b_1^* \widehat{\Pi}_i b_2 - \zeta_i b_1^* \Pi_i b_2 \xrightarrow{\text{a.s.}} 0$$

where

$$\zeta_i = \frac{m(\rho_i)(1 + h(\rho_i))}{h'(\rho_i)}.$$

In particular, we find after some derivations:

Corollary 2: Under the assumptions of Theorem 2, let X be standard Gaussian. Then

$$b_1^* \widehat{\Pi}_i b_2 - \frac{1 - c\omega_i^{-2}}{1 + c\omega_i^{-1}} b_1^* \Pi_i b_2 \xrightarrow{\text{a.s.}} 0.$$

This result is consistent with [25] derived in the real Gaussian case for eigenvalues with unit multiplicity.

Theorem 4 and Corollary 4 provide an interesting characterization of the eigenspaces of P through limiting projections in the large dimensional setting. In the context of local failure in large sensor networks, it is therefore possible to detect and diagnose one or multiple failures by comparing eigenspace projection patterns associated with each failure type. Precisely, an appropriate diagnosis consists in determining the most likely failure type among all hypothetical failures, given the extreme eigenvalues and associated eigenspace projections of $\Sigma\Sigma^*$. To this end though, not only first order limits but also second order behaviour need be characterized precisely. This is the target of the following section.

C. Second order behavior

Before studying the fluctuations of $\hat{\lambda}_{\mathcal{K}(i)+\ell}$, $\ell = 1, \ldots, j_i$, when ω_i satisfies the separation condition, we first remind for later use in our applicative framework the fluctuations of $\hat{\lambda}_{\mathcal{K}(i)+\ell}$ when ω_i does *not* satisfy the separation condition, and when X is a standard Gaussian matrix. For this, we have the following theorem [20], [22], [21].

Theorem 3: Let X be standard Gaussian, then if $0 < \omega_i < \sqrt{c}$,

$$N^{\frac{2}{3}} \frac{\hat{\lambda}_{\mathcal{K}(i)+\ell} - (1+\sqrt{c})^2}{(1+\sqrt{c})^{\frac{4}{3}}\sqrt{c}} \Rightarrow T_2$$

and, if $-\sqrt{c} < \omega_i < 0$,

$$N^{\frac{2}{3}} \frac{\hat{\lambda}_{\mathcal{K}(i)+\ell} - (1-\sqrt{c})^2}{-(1-\sqrt{c})^{\frac{4}{3}}\sqrt{c}} \Rightarrow T_2$$

for $\ell = 1, \ldots, j_i$, as $n \to \infty$, where T_2 is the complex Tracy-Widom distribution function [19].

The tools used to derive Theorem 3 are much different from those exploited here and will not be discussed. Similar to [24], Theorem 3 will be used to derive tests to decide on the presence of eigenvalues outside the support of the Marčenko-Pastur law. For failure detection purposes in sensor networks, this will be used to declare a failure prior to diagnose the fault. Then, to diagnose a failure, second order statistics of both eigenvalue and eigenspace projections when the separation property arises are needed. This is the aim of the remainder of the section.

We now turn to the second order analysis of the eigenspectrum of $\Sigma\Sigma^*$ when ω_i satisfies the separation condition, and when X is only assumed to satisfy **A1-A3**. We first need the following additional assumption:

A4 For all $z \in \mathbb{C} \setminus [a, b]$,

$$\sqrt{N}\left(\alpha(z) - m(z)\right) \xrightarrow{\mathbb{P}} 0$$

as $n \to \infty$.

In practice, this assumption means that the fluctuations of the spectral measure of XX^* are negligible with respect to those of the $\hat{\lambda}_k$ and $b_1^* \widehat{\Pi}_i b_2$, which will show to be of order $\sqrt{N}$. This assumption is satisfied by most of the random matrix models of practical importance in our context, provided $\sqrt{N}(N/n - c) \to 0$. The classical illustrating example in this regard concerns the standard Gaussian case. Denote by $m_n(z)$ the Stieltjes transform (4) where c is replaced with N/n, and let π_n the Marčenko-Pastur law associated with $m_n(z)$. For any $z \in \mathbb{C} \setminus [a, b]$, the function $f(x) = (x - z)^{-1}$ is analytic outside the support of π_n for n sufficiently large. As a consequence, Theorem 1.1 of [34] shows that $\sqrt{N}(\alpha(z) - m_n(z)) \xrightarrow{\mathbb{P}} 0$. Assuming in addition that $\sqrt{N}(N/n - c) \to 0$, it is not difficult to show from (4) that $\sqrt{N}(m_n(z) - m(z)) \to 0$.[3]

For practical purposes, we shall also assume:

A5 Each ω_i, $1 \leq i \leq t$, satisfies the separation condition.

[3]It is useful to note that when the convergence rate of N/n is slower than $1/\sqrt{N}$, our analysis remains true if we replace $m(z)$ with a finite horizon deterministic equivalent [35], [36]. For simplicity, we have chosen not to enter into these details here.

The main result of this section is the following theorem.

Theorem 4: Assume **A1-A5**. For $i = 1, \ldots, t$, let

$$V_{i,n} = \sqrt{N} U_i^* \left(\widehat{\Pi}_i - \zeta_i I_N \right) U_i$$

and

$$L_{i,n} = \sqrt{N} \begin{bmatrix} \hat{\lambda}_{\mathcal{K}(i)+1} - \rho_i \\ \vdots \\ \hat{\lambda}_{\mathcal{K}(i)+j_i} - \rho_i \end{bmatrix}.$$

For $\rho \in \mathbb{R} \setminus [a, b]$, let

$$D(\rho) = \begin{bmatrix} \frac{h(\rho)(1+h(\rho))h''(\rho)}{h'(\rho)^3} & -\frac{h(\rho)(1+h(\rho))}{h'(\rho)^2} \\ -\frac{\rho}{h'(\rho)} & 0 \end{bmatrix}$$

and

$$R(\rho) = \begin{bmatrix} m'(\rho) - m(\rho)^2 & m''(\rho)/2 - m(\rho)m'(\rho) \\ m''(\rho)/2 - m(\rho)m'(\rho) & m^{(3)}(\rho)/6 - m'(\rho)^2 \end{bmatrix}$$

where $m^{(3)}$ is the third derivative of m. Consider the matrices

$$\begin{bmatrix} G_i \\ K_i \end{bmatrix} = \left((D(\rho_i)R(\rho_i)D(\rho_i)^*)^{1/2} \otimes I_{j_i} \right) \begin{bmatrix} M_{1,i} \\ M_{2,i} \end{bmatrix}$$

where $M_{1,1}, M_{2,1}, \ldots, M_{1,t}, M_{2,t}$ are independent GUE matrices such that $M_{1,i}$ and $M_{2,i}$ are both $j_i \times j_i$ matrices.[4] Let L_i be the $\mathbb{R}^{j_i}$-valued vector of eigenvalues of K_i arranged in decreasing order. Then[5]

$$((V_{i,n}, L_{i,n}))_{i=1}^t \Rightarrow ((G_i, L_i))_{i=1}^t . \tag{10}$$

Remark 1: To be more precise, $(V_{i,n}, L_{i,n})$ can be cast into an $\mathbb{R}^{j_i^2 + j_i}$-valued random vector after rearranging the real and imaginary parts of the elements of $V_{i,n}$ and taking into account the Hermitian symmetry constraint. Hence, the convergence in law specified by (10) acts on the space of probability measures on $\mathbb{R}^{r + j_1^2 + \cdots + j_t^2}$.

Theorem 4 provides a very general expression of the joint limiting fluctuations of both eigenvalues and eigenspace projections. It is particularly interesting to note that the fluctuations of $(V_{i,n}, L_{i,n})$ are asymptotically independent across i.

The theoretical is an interesting new result to the field of large dimensional random matrix theory, although it might be difficult to use in practice since one needs to derive explicitly the joint density of $(G_i, L_i)_{i=1}^t$. Nonetheless, Theorem 4 can be immediately put to practice in two scenarios. The first scenario corresponds to the case where only the fluctuations of the eigenspace projection vector $(V_{i,n})_{i=1}^t$ is of interest. In this case, $(G_i)_{i=1}^t$ is a correlated Gaussian random variable. Reminding that $M_{1,i}$ can be seen as an $\mathbb{R}^{j_i^2}$-valued Gaussian vector with entries of zero mean and unit variance, $(G_i)_{i=1}^t$ can then be seen as a random Gaussian vector in $\mathbb{R}^{j_1^2 + \cdots + j_t^2}$ with entries $(j_{i-1}^2 - 1)$ to j_i^2 of zero mean and variance $[D(\rho_i)R(\rho_i)D(\rho_i)^*]_{11}$, for each $i \in \{1, \ldots, t\}$ with $j_0 = 0$. The second scenario of interest, which is discussed at length in the following, corresponds to the case where the multiplicities of the eigenvalues of P are all equal to one. In this case, $M_{1,i}$ and $M_{2,i}$ are independent Gaussian variables and we immediately have the following corollary:

Corollary 3: Assume the setting of Theorem 4. Assume in addition that $j_i = 1$ for all i. Then

$$((V_{i,n}, L_{i,n}))_{i=1,\ldots,r} \Rightarrow \mathcal{N}(0, R)$$

with

$$R = \begin{bmatrix} D(\rho_1)R(\rho_1)D(\rho_1)^* & & \\ & \ddots & \\ & & D(\rho_r)R(\rho_r)D(\rho_r)^* \end{bmatrix}.$$

After some calculus, in the standard Gaussian case, we further have:

Corollary 4: Under the assumptions of Corollary 3, if X is a standard Gaussian matrix, then $D(\rho_i)R(\rho_i)D(\rho_i)^* = C(\rho_i)$, where

$$C(\rho_i) \triangleq \begin{bmatrix} \frac{c^2(1+\omega_i)^2}{(c+\omega_i)^2(\omega_i^2 - c)} \left(c\frac{(1+\omega_i)^2}{(c+\omega_i)^2} + 1 \right) & \frac{(1+\omega_i)^3 c^2}{(\omega_i + c)^2 \omega_i} \\ \frac{(1+\omega_i)^3 c^2}{(\omega_i + c)^2 \omega_i} & \frac{c(1+\omega_i)^2(\omega_i^2 - c)}{\omega_i^2} \end{bmatrix}.$$

Due to its simple expression, Corollary 4 is particularly handy to use in the context of failure diagnosis when hypothetical failures are characterized by distinct values of ω_i, as will be shown in Section IV.

In Figure 1, the histogram of a simulation of 10 000 realizations of the projection $V_{1,n} = \sqrt{N}(|\hat{u}_1^* u_1|^2 - \zeta_1)$, with $u_1 = U_1 \in \mathbb{C}^N$, $\hat{u}_1 \hat{u}_1^* = \widehat{\Pi}_1$ of unit rank, and X standard Gaussian, is depicted against the asymptotic Gaussian law derived in Corollary 4, for $c = 1/8$, $r = 1$, $N/n = 1/8$, $N = 256$ and ω_1 successively equal to 1 and 0.5. In this scenario, $\sqrt{N/n} \simeq 0.35$. For $\omega_1 = 1$, the simulation shows a rather accurate fit between asymptotic theory and simulation. For $\omega_1 = 0.5$, the Gaussian approximation is much less accurate. This is due to the value $\omega_1 = 0.5$, which is rather close to $\sqrt{N/n} \simeq 0.35$. The value $N = 256$ is here insufficient for the large dimensional behaviour of the fluctuations of $|\hat{u}_1^* u_1|^2$ to arise. This behaviour will have important consequences for the question of diagnosing failures which are difficult to observe.

The remainder of this section is devoted to the proof of Theorem 4. We start with the following lemma, which deals with the asymptotic behavior of the $V_{i,n}$. This lemma will be proved in Appendix A-A.

Lemma 2: Assume **A1-A5**. Let

$$\bar{V}_{i,n} = \sqrt{N} \left(\frac{h(\rho_i)(1 + h(\rho_i))h''(\rho_i)}{h'(\rho_i)^3} U_i^* (Q(\rho_i) - m(\rho_i) I) U_i \right.$$
$$\left. - \frac{h(\rho_i)(1 + h(\rho_i))}{h'(\rho_i)^2} U_i^* (Q'(\rho_i) - m'(\rho_i) I) U_i \right).$$

Then for any $i \in \{1, \ldots, t\}$,

$$V_{i,n} - \bar{V}_{i,n} \xrightarrow{\mathbb{P}} 0.$$

We now consider the isolated extreme eigenvalues. In order to study the asymptotic behavior of these eigenvalues, we shall adapt to our situation the approach of [26]. For $i = 1, \ldots, t$, consider real numbers $x_1(i) > y_1(i) > x_2(i) > y_2(i) > \cdots > y_{j_i}(i)$. Since the separation condition is satisfied by

[4] We remind that matrices from the Gaussian unitary ensemble, or GUE matrices, are random Hermitian matrices with independent standard real Gaussian diagonal entries and independent standard complex Gaussian upper-diagonal entries [37].

[5] It is clear that if **A5** is not fulfilled, Theorem 4 generalizes by letting the index i in (10) span only the subset $\{1, \ldots, p, q, \ldots, t\}$.

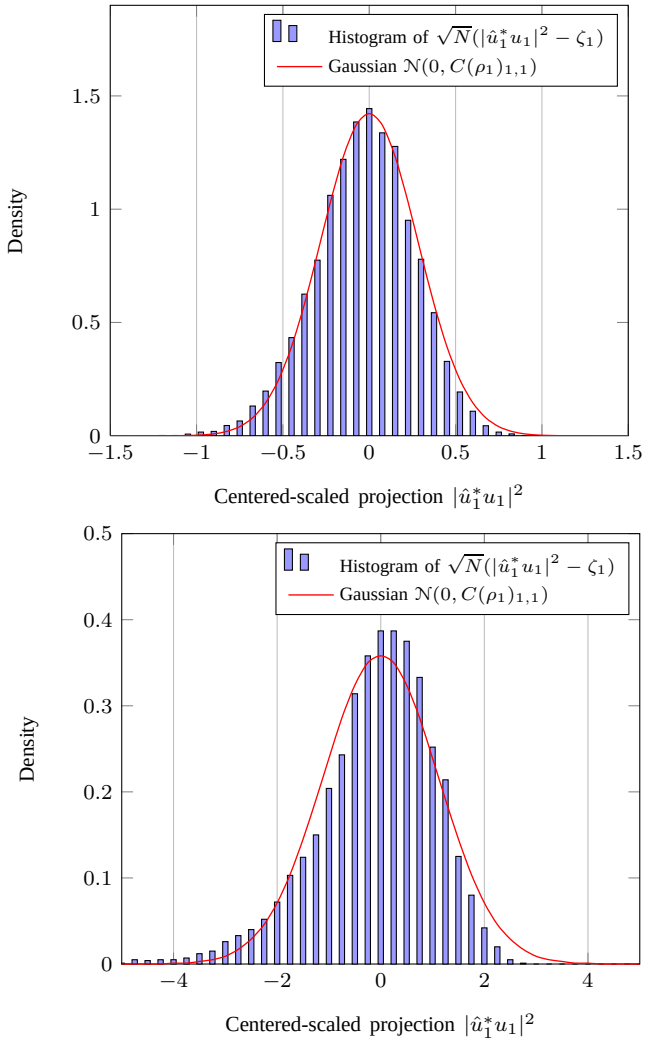

Fig. 1. Empirical and theoretical distribution of the fluctuations of $\hat{u}_1$ with $r = 1$, X has i.i.d. zero mean variance $1/n$ entries, $N/n = 1/8$, $N = 256$ and $\omega_1 = 1$ (top) or $\omega_1 = 0.5$ (bottom).

assumption for each $i = 1, \ldots, t$, the equation $\det \widehat{H}(x) = 0$ has r roots outside $[a, b]$ with probability one for all n large. Therefore, we have the equivalence relation

$$\begin{aligned}\Big(x_\ell(i) > \sqrt{N}(\hat{\lambda}_{\mathcal{K}(i)+\ell} - \rho_i) > y_\ell(i), \\ i = 1, \ldots, t, \quad \ell = 1, \ldots, j_i\Big) \\ \Leftrightarrow \\ \Big(N^{j_i} \det \widehat{H}\Big(\rho_i + \frac{x_\ell(i)}{\sqrt{N}}\Big) \det \widehat{H}\Big(\rho_i + \frac{y_\ell(i)}{\sqrt{N}}\Big) < 0, \\ i = 1, \ldots, t, \quad \ell = 1, \ldots, j_i\Big). \end{aligned} \tag{11}$$

This equivalence leads us to study the limits of the finite dimensional distributions of the $\mathbb{R}^t$-valued random process $\left[N^{j_i/2} \det \widehat{H}(\rho_i + x/\sqrt{N})\right]_{i=1}^t$ in the parameter x. This is given by the following lemma, proved in Appendix A-B.

Lemma 3: Assume **A1-A5**. Define the $\mathbb{R}^t$-valued random process

$$\chi_n(x) = \Big[N^{j_i/2} \det \widehat{H}(\rho_i + x/\sqrt{N}) - \Big(\prod_{\ell \neq i} (1 - \beta_\ell/\beta_i)^{j_\ell}\Big) \times \det\Big(\sqrt{N}\beta_i \rho_i U_i^* (Q(\rho_i) - m(\rho_i) I) U_i + \beta_i h'(\rho_i) x I\Big)\Big]_{i=1}^t$$

where $\beta_i = \omega_i/(1 + \omega_i)$. Then

$$(\chi_n(x_1), \ldots, \chi_n(x_p)) \xrightarrow{\mathbb{P}} 0$$

for every finite sequence $(x_1, \ldots, x_p)$.

Let B be a rectangle of $\mathbb{R}^{j_1^2 + \cdots + j_t^2}$. The real and imaginary parts of the elements of the $V_{i,n}$ defined in Theorem 4 can be stacked into a $\mathbb{R}^{j_1^2 + \cdots + j_t^2}$-valued vector V_n when taking into account the Hermitian symmetry constraint. A vector $\bar{V}_n$ with the same size can be constructed similarly from $\bar{V}_{i,n}$ defined in Lemma 2. Let L_n be the $\mathbb{R}^r$-valued vector $L_n = [L_{1,n}^\mathsf{T}, \ldots, L_{t,n}^\mathsf{T}]^\mathsf{T}$ (see Theorem 4) and let C be the rectangle of $\mathbb{R}^r$ determined by the left hand side of (11), so that this event is written $[L_n \in C]$. Let $\bar{L}_n$ be the vector obtained by arranging the eigenvalues of the matrices

$$\bar{T}_{i,n} = -\frac{\rho_i}{h'(\rho_i)} \sqrt{N} U_i^* (Q(\rho_i) - m(\rho_i) I) U_i$$

for $i = 1, \ldots, t$, similarly to the elements of L_n. From Lemma 2, Lemma 3, and the discussion preceding Lemma 3, we have

$$\mathbb{P}\left[V_n \in B, L_n \in C\right] - \mathbb{P}\left[\bar{V}_n \in B, \bar{L}_n \in C\right] \to 0$$

as $n \to \infty$, for arbitrary rectangles B and arbitrary rectangles C specified at the left hand side of (11). Observe that

$$\begin{bmatrix} \bar{V}_{i,n} \\ \bar{T}_{i,n} \end{bmatrix} = \sqrt{N} \left(D(\rho_i) \otimes I_{j_i}\right) \begin{bmatrix} U_i^*(Q(\rho_i) - m(\rho_i) I) U_i \\ U_i^*(Q'(\rho_i) - m'(\rho_i) I) U_i \end{bmatrix}.$$

In order to terminate the proof of Theorem 4, we shall make use of the following lemma, that we state in a slightly more general form than needed here.

Lemma 4: Assume **A1-A4**. Let $f_1, \ldots, f_t$ and $g_1, \ldots, g_t$ be real functions analytical on a neighborhood of $[a, b]$. Let S_n be the t-uple of random matrices

$$S_n = \left(\sqrt{N} \begin{bmatrix} U_i^* f_i(XX^*) U_i - \left(\int f_i(\lambda) d\pi(\lambda)\right) I_{j_i} \\ U_i^* g_i(XX^*) U_i - \left(\int g_i(\lambda) d\pi(\lambda)\right) I_{j_i} \end{bmatrix}\right)_{i=1}^t.$$

For $i = 1, \ldots, t$, define the covariance matrices

$$R_i = \int \left(\begin{bmatrix} f_i - \int f_i d\pi \\ g_i - \int g_i d\pi \end{bmatrix} \begin{bmatrix} (f_i - \int f_i d\pi) & (g_i - \int g_i d\pi) \end{bmatrix} \right) d\pi.$$

Then S_n converges in distribution towards

$$\left(\left(R_i^{1/2} \otimes I_{j_i}\right) \begin{bmatrix} M_{1,i} \\ M_{2,i} \end{bmatrix}\right)_{i=1}^t \tag{12}$$

where the matrices $(M_{1,1}, M_{2,1}, \ldots, M_{1,t}, M_{2,t})$ are independent GUE matrices such that $M_{1,i}$ and $M_{2,i}$ have dimensions $j_i \times j_i$.

The proof is provided in Appendix A-C.

Applying this lemma with with $f_i(\lambda) = 1/(\lambda - \rho_i)$ and $g_i(\lambda) = 1/(\lambda - \rho_i)^2$, R_i takes the value $R(\rho_i)$ provided in the statement of Theorem 4. It results that

$$\left(\begin{bmatrix} \bar{V}_{i,n} \\ \bar{T}_{i,n} \end{bmatrix}\right)_{i=1}^t \Rightarrow \left(\begin{bmatrix} G_i \\ K_i \end{bmatrix}\right)_{i=1}^t$$

where the convergence takes place on the space of probability measures on $\mathbb{R}^{2(j_1^2+\cdots+j_t^2)}$. This completes the proof of Theorem 4.

IV. Application

In this section, we provide a general framework for local failure detection and diagnosis in large sensor networks, such as the examples proposed in Section II, based on the results of Section III. This framework is a two-step approach for successively (i) detecting failures within a given maximally acceptable false alarm rate and (ii) upon positive detection, diagnosing the failures with high probability. Simulations are then run to validate the proposed algorithms.

We recall that we assume a number K of failure scenarios indexed by $1 \leq k \leq K$. Scenario k is characterized by the population covariance matrix $I_N + P_k$. We additionally denote $P_0 = 0$ for convenience. We consider precisely the detection and localization to be made for failure models of the type $\Sigma = (I_N+P_k)^{\frac{1}{2}}X$, $1 \leq k \leq K$, with $P_k = \sum_{i=1}^{t_k} \omega_{k,i} U_{k,i} U_{k,i}^*$ of rank $r_k = \sum_{i=1}^{t_k} j_{k,i}$, where $U_{k,i} \in \mathbb{C}^{N \times j_{k,i}}$ and $\omega_{k,1} > \ldots > \omega_{k,s_k} > 0 > \omega_{k,s_k+1} > \ldots > \omega_{k,t_k}$, and X standard Gaussian. We denote $\mathcal{H}_0$ the null hypothesis for the model $\Sigma = X$ and $\mathcal{H}_k$ the hypothesis for a failure of type k. For the failure scenario k, we also take p_k to be zero or the smallest index i such that $\omega_{k,i} > \sqrt{c}$, and q_k to be $t_k + 1$ or the largest index i such that $\omega_{k,i} < -\sqrt{c}$.

Remark 2: This model (and in particular the models of Section II-A and Section II-B) may be extended by introducing deterministic linear time correlation to the observations $s_1, \ldots, s_n$. That is, we can write $\Sigma = (I_N + P_k)^{\frac{1}{2}} X T^{\frac{1}{2}}$ for some time-correlation matrix T. In this scenario, XTX^* being unitarily invariant for X Gaussian, the localization scheme proposed extends naturally to this scenario, as long as the eigenvalue distribution of T converges weakly to a compactly supported distribution as N grows large and that T has asymptotically no eigenvalue outside this support. However, a result similar to the Tracy-Widom fluctuations in Theorem 3 does not exist yet for the fluctuations of the extreme eigenvalues in the model XTX^*. As such, although failure localization is possible from Theorem 4, no failure detection test is yet available for such a model.[6]

A. Detection algorithm

As stated in the introduction, the detection phase relies on existing results, and more specifically on the fluctuations of the largest eigenvalues given by Theorem 3. The detection algorithms proposed here parallel that introduced in [24] in the context of collaborative signal sensing. The objective is to decide between hypothesis $\mathcal{H}_0$ and its complementary $\bar{\mathcal{H}}_0$.

First assume that all P_k only have non-negative eigenvalues. From Theorem 1, the largest eigenvalue $\hat{\lambda}_1$ of $\Sigma\Sigma^*$ tends to the right edge of the support S_π of the Marčenko-Pastur law for all large n under $\mathcal{H}_0$, while $\hat{\lambda}_1$ is found away from this edge under $\bar{\mathcal{H}}_0$ *if* the largest eigenvalue $\omega_{k,1}$ of P_k exceeds $\sqrt{c}$. We will therefore assume in the following that $\omega_{k,1} > \sqrt{c}$ is verified for all k. That is, we assume that $c_N \triangleq N/n < c_+$ where c_+ is defined as

$$c_+ \triangleq \inf \left\{ \omega_{k,1}^2,\ 1 \leq k \leq K \right\}.$$

This condition allows for a theoretically almost sure error detection, as $N, n \to \infty$. We then rely on Theorem 3 to design an appropriate hypothesis test. Our test consists in rejecting hypothesis $\mathcal{H}_0$ if the probability in favor of $\mathcal{H}_0$ is sufficiently low. That is, for a given acceptable *false alarm rate* η,[7] the statistical test is defined as

$$\hat{\lambda}_1' \underset{\bar{\mathcal{H}}_0}{\overset{\mathcal{H}_0}{\lessgtr}} (T_2)^{-1}(1-\eta) \tag{13}$$

where $\hat{\lambda}_1'$ is given by

$$\hat{\lambda}_1' \triangleq N^{\frac{2}{3}} \frac{\hat{\lambda}_1 - (1+\sqrt{c_N})^2}{(1+\sqrt{c_N})^{\frac{4}{3}} c_N^{\frac{1}{2}}}.$$

That is, the test verifies whether $\hat{\lambda}_1$ exceeds some threshold above which the probability for $\mathcal{H}_0$ is less than η.

If the matrices P_k are now all non-positive definite then, symmetrically, we need to set c_N such that the smallest eigenvalue $\hat{\lambda}_N$ of $\Sigma\Sigma^*$ is visible on the left-hand side of S_π. That is, we take N, n to be such that $c_N < c_-$, with c_- defined as

$$c_- \triangleq \inf \left\{ \omega_{k,t_k}^2,\ 1 \leq k \leq K \right\}.$$

The decision test is in that case given by

$$\hat{\lambda}_N' \underset{\bar{\mathcal{H}}_0}{\overset{\mathcal{H}_0}{\lessgtr}} (T_2)^{-1}(1-\eta) \tag{14}$$

where $\hat{\lambda}_N'$ is defined as

$$\hat{\lambda}_N' \triangleq N^{\frac{2}{3}} \frac{\hat{\lambda}_N - (1-\sqrt{c_N})^2}{-(1-\sqrt{c_N})^{\frac{4}{3}} c_N^{\frac{1}{2}}}.$$

The above test is particularly suited to the model of Section II-B in which the matrices P_k are non-positive definite when $\mu_k = 0$ and $\alpha_k = -1$ for all k, corresponding to a sudden drop of a zero mean random parameter $\theta(k)$ to zero.

When the matrices P_k have both positive and negative eigenvalues, then a deterministic choice has to be made by the experimenter. In the most general setting, to ensure a false alarm rate lower than η, one has to choose two scalars $a(\eta), b(\eta) \in \mathbb{R} \cup \{-\infty, \infty\}$ such that

$$P\left(\left\{ \hat{\lambda}_1' > a(\eta) \right\} \cup \left\{ \hat{\lambda}_N' > b(\eta) \right\} \right) \leq \eta. \tag{15}$$

The choice of $a(\eta), b(\eta)$ depends primarily on the structure of P_k and will impact the correct detection rate for fixed false alarm rates.

Following the work of [38], in which the asymptotic independence of the fluctuations of the largest and the smallest

[6] Alternatively, a whitening procedure may be used prior to failure detection using $\Sigma' = \Sigma T^{-\frac{1}{2}}$ as the random matrix under study, for T invertible.

[7] We recall that the false alarm rate is the probability of declaring $\bar{\mathcal{H}}_0$ under true hypothesis $\mathcal{H}_0$.

eigenvalues of GUE matrices is proved, it is natural to conjecture that the same result holds for the eigenvalues $\hat{\lambda}_1$ and $\hat{\lambda}_N$ under $\mathcal{H}_0$. In this case, (15) would become asymptotically

$$T_2(a(\eta))T_2(b(\eta)) > 1-\eta.$$

For any fixed b, taking $b(\eta) = b$, the hypothesis test now becomes

$$\min\left\{b - \hat{\lambda}'_N,\ (T_2)^{-1}\left(\frac{1-\eta}{T_2(b)}\right) - \hat{\lambda}'_1\right\} \underset{\bar{\mathcal{H}}_0}{\overset{\bar{\mathcal{H}}_0}{\lessgtr}} 0.$$

In particular, for $b = \infty$, $T_2(b) = 1$ and then the test reduces to

$$\hat{\lambda}'_1 \underset{\bar{\mathcal{H}}_0}{\overset{\bar{\mathcal{H}}_0}{\lessgtr}} (T_2)^{-1}(1-\eta)$$

which is the same test as proposed in (13). Taking instead $a(\eta) = -\infty$, we obtain the test (14).

For rather symmetrical distributions of the eigenvalues $\omega_{k,i}$ of P_k around zero, it may be interesting to set $b(\eta) = a(\eta)$, in which case

$$b(\eta) = (T_2)^{-1}\left(\sqrt{1-\eta}\right).$$

In this setting, the decision test is now

$$\max\left\{\hat{\lambda}'_N, \hat{\lambda}'_1\right\} \underset{\bar{\mathcal{H}}_0}{\overset{\bar{\mathcal{H}}_0}{\lessgtr}} (T_2)^{-1}\left(\sqrt{1-\eta}\right).$$

In the following section, we assume that the procedure of failure detection was achieved successfully and that we are now interested in localizing the failures.

B. Localization algorithm

In this section, we wish to detect all possible failure events from a set of failures indexed by $k \in \{1,\dots,K\}$. The index set $\{1,\dots,K\}$ may gather all events accounting for a single, as well as multiple, local failures. Similar to the previous sections, we denote $\rho_{k,i} = 1 + \omega_{k,i} + c\frac{1+\omega_{k,i}}{\omega_{k,i}}$ and $\zeta_{k,i} = \frac{1-c\omega_{k,i}^{-2}}{1+c\omega_{k,i}^{-1}}$, we define the mapping $\mathcal{K}_k$ to be such that $\mathcal{K}_k(i) = j_{k,1} + \ldots + j_{k,i-1}$ if $1 \le i \le s_k$ and $\mathcal{K}_k(i) = N - (j_{k,1} + \ldots + j_{k,t_k})$ if $s_k + 1 \le i \le t_k$. Finally, we denote $\widehat{\Pi}_{k,i}$ any projector on the subspace generated by the eigenvalues $\hat{\lambda}_{\mathcal{K}_k(i)+1}, \ldots, \hat{\lambda}_{\mathcal{K}_k(i)+j_{k,i}}$.

Based on the fact that different P_k's have in general very distinct eigenspaces (and sometimes very different eigenvalues), we then propose the following subspace localization test, which decides on the hypothesis $\mathcal{H}_{k^\star}$ for which $k^\star$ is given by

$$k^\star = \arg\max_{k\in S} g_k\left(\left(V^k_{i,n}, L^k_{i,n}\right)_{i\in\mathcal{L}(p_k,q_k)}\right) \tag{16}$$

with g_k the true density of the vector $\left(V^k_{i,n}, L^k_{i,n}\right)_{i\in\mathcal{L}(p_k,q_k)}$, $\mathcal{L}(p_k,q_k) = \{1,\dots,p_k,q_k,\dots,r_k\}$, S the set of indexes k such that $\mathcal{L}(p_k,q_k)$ is non-empty, and where

$$V^k_{i,n} \triangleq \sqrt{N}U^*_{k,i}\left(\widehat{\Pi}_{k,i} - \zeta_{k,i}I_{j_{k,i}}\right)U_{k,i}$$

and

$$L^k_{i,n} \triangleq \sqrt{N}\begin{bmatrix}\hat{\lambda}_{\mathcal{K}_k(i)+1} - \rho_{k,i}\\ \vdots \\ \hat{\lambda}_{\mathcal{K}_k(i)+j_{k,i}} - \rho_{k,i}\end{bmatrix}.$$

Note that we need here to specify the indexation $i \in \mathcal{L}(p_k,q_k)$ since we do not assume **A5** for any k.

From Theorem 4, this probability can be approximated for large n, which provides immediately a maximum likelihood test for the most asymptotically likely $\mathcal{H}_k$ hypothesis. In the particular case where the $\omega_{k,i}$ all have multiplicity one, according to Corollary 4, as N, n grow large, the vectors in the test (16) are asymptotically independent and Gaussian. We therefore substitute the test (16) by the following test, leading to the estimator $\hat{k}$ defined as

$$\hat{k} = \arg\max_{k\in S}\prod_{i\in\mathcal{L}(p_k,q_k)} f\left((V^k_{i,n}, L^k_{i,n}); C(\rho_{k,i})\right) \tag{17}$$

where $f(x;\Omega)$, $x \in \mathbb{C}^m$, $\Omega \in \mathbb{C}^{m\times m}$, is the m-variate real normal density of zero mean and covariance C at point x, and $C(\rho_{k,i})$ is defined similarly as in Corollary 4, with ω_i replaced by $\omega_{k,i}$. Taking a log-likelihood notation, this is explicitly

$$\hat{k} = \arg\min_{k\in S}\sum_{i\in\mathcal{L}(p_k,q_k)} \left[(V^k_{i,n}, L^k_{i,n})C(\rho_{k,i})^{-1}(V^k_{i,n}, L^k_{i,n})^\mathsf{T} + \log\det C(\rho_{k,i}) + 2\log(2\pi)\right].$$

In the general case where $\omega_{k,i}$ has multiplicity $j_{k,i}$, as discussed previously, it is simpler to restrict the detection test to the eigenspace projections $(V^k_{i,n})_{i\in\mathcal{L}(p_k,q_k)}$ only. In this case, denoting $\bar{V}^k_{i,n} = (V^k_{i,n,D}, \sqrt{2}V^k_{i,n,U}) \in \mathbb{R}^{j^2_{k,i}}$ with $V^k_{i,n,D} \in \mathbb{R}^{j_{k,i}}$ the vector of the diagonal entries of $V^k_{i,n}$ and $V^k_{i,n,U} \in \mathbb{R}^{(j^2_{k,i}-j_{k,i})}$ the vector of the real and imaginary parts of the upper-diagonal entries of $V^k_{i,n}$, we obtain the test

$$\hat{k} = \arg\min_{k\in S}\sum_{i\in\mathcal{L}(p_k,q_k)} \Big[\frac{1}{[C(\rho_{k,i})]_{11}}(\bar{V}^k_{i,n})^\mathsf{T}\bar{V}^k_{i,n} + j^2_{k,i}\log([C(\rho_{k,i})]_{11}) + j^2_{k,i}\log(2\pi)\Big]. \tag{18}$$

Remark 3: We discuss below the advantages of the detection tests proposed in Section IV-A and the localization algorithms (17)–(18) compared to the optimum maximum likelihood approach, as well as the extension to more elaborate scenarios:

- the detection algorithms proposed in Section IV-A are very versatile, as they adapt to multiple failure scenarios showing small rank perturbations in the population covariance, and provide a theoretical expression of the minimum ratio $c_N = N/n$ necessary for detectability.
- unlike the traditional maximum-likelihood approach which tests the joint distribution of Σ for all hypotheses $1 \le k \le K$, and therefore leads to calculus of the order of at least N^3 for each k, the proposed localization algorithm (17) is based on a test requiring for each k a calculus complexity of order r_k, where $r_k \ll N$.
- we may decide not to consider the joint fluctuations of all eigenvalues found outside S_π, but only some

of them. This leads to an asymptotically less efficient, although much faster, algorithm, where $\mathcal{L}(p_k, q_k)$ in (17) is replaced by $\mathcal{L}(p', q')$ for given $p' \leq p_k$, $q' \leq q_k$, for all k. For N not too large, it is in fact preferable to consider only a few eigenvalues and eigenspace projections simultaneously, due to convergence speed limitations.

- the entries $L^k_{i,n}$ of the vector $(V^k_{i,n}, L^k_{i,n})$ may also be discarded, especially in scenarios where eigenvalues of P_k are very similar for each hypothesis $\mathcal{H}_k$. This may again increase the convergence speed of the asymptotic approximation for not-too-large N, while it is expected to perform worse for large N.

So far, we have performed failure detection under the important assumption that the failure scenarios form a discrete set $\{1, \ldots, K\}$. This assumes in particular that the failure amplitudes are known prior to detection and localization. In the next section, we use Corollary 4 to improve this approach in the particularly simple example of Section II-B, when the failure amplitude is a priori unknown.

C. Extension to unknown failure amplitude

In this section, we assume the scenario where the eigenvectors of the perturbation matrix P_k are independent of the amplitude of the failure parameters, in the sense that a change in magnitude of the failure of type k does not affect the eigenspaces of P_k. This is for instance the case of the single-failure scenario of Section II-B, for which we recall that P_k expresses as $P_k = \beta_k R^{-\frac{1}{2}} H e_k e_k^* H^* R^{-\frac{1}{2}}$ with β_k the failure parameter. We now assume β_k unknown, which is a more realistic assumption than assuming it perfectly known in advance. We also suppose that X has i.i.d. Gaussian entries. Based on a simple extension of the algorithm presented in Section IV-B to unknown ω_k, we provide hereafter a second localization algorithm.

For notational convenience, we assume $P_k = \omega_k u_k u_k^*$ for each k and that $\omega_k > \sqrt{c}$, unknown. We then denote $\hat{\lambda}$ the largest eigenvalue of $\Sigma\Sigma^*$ and $\hat{u}$ its associated eigenvector.

Obviously, since ω_k is not known, neither is ρ_k. Therefore, we cannot proceed here to localization based on the fluctuations of $\hat{\lambda}$. Instead, we will use $\hat{\lambda}$ precisely as an estimate of ρ_k, which we know is consistent with growing N, n. From $\hat{\lambda}$, assumed larger than $(1+\sqrt{c})^2$, we want to derive an estimate $\hat{\omega}$ of ω_k (k is the effective failure index). This is obtained from an inversion of the relation (7). Precisely, we obtain

$$\hat{\omega} \triangleq \frac{1}{2}(\hat{\lambda} - (1+c)) + \frac{1}{2}\sqrt{(\hat{\lambda} - (1+c))^2 - 4c}$$

if $\hat{\lambda} > (1+\sqrt{c})^2$ and

$$\hat{\omega} \triangleq \frac{1}{2}(\hat{\lambda} - (1+c)) - \frac{1}{2}\sqrt{(\hat{\lambda} - (1+c))^2 - 4c}$$

if $\hat{\lambda} < (1-\sqrt{c})^2$.

From this estimate, we then obtain an estimate $\hat{\zeta}$ of ζ_k as follows

$$\hat{\zeta} = \frac{1 - c\hat{\omega}^{-2}}{1 + c\hat{\omega}^{-1}}.$$

A natural object to consider for the failure localization is now $|u_k^* \hat{u}|^2 - \hat{\zeta}$. To provide a diagnosis test, we need to derive the fluctuations of this random variable. From Theorem 4, the fluctuations of $\sqrt{N}(|u_k^* \hat{u}|^2 - \zeta_k)$ depend on ω_k but not on u_k. From the expression of $\hat{\zeta}$, it is immediate that the fluctuations of $\sqrt{N}(\hat{\zeta} - \zeta_k)$ also depend on ω_k only. But since ω_k is estimated by $\hat{\omega}$, irrespective of the failure index k, the diagnosis test leads to finding the most likely argument $\hat{k}'$ among K variables with same Gaussian statistics. This therefore simplifies the estimator $\hat{k}'$ of the most likely index k to the following minimum-distance estimator

$$\hat{k}' = \arg \min_{k \in \{1,\ldots,K\}} \left| |u_k^* \hat{u}|^2 - \hat{\zeta} \right|.$$

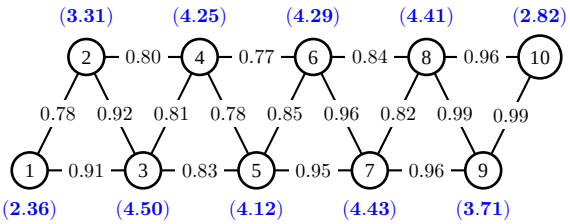

Fig. 2. Network of $N = 10$ sensors. The correlation $\mathrm{E}[y(i)^* y(j)]$ between data on sensors i and j, $i \neq j$, can be read on the link (i, j), while $\mathrm{E}[|y(i)|^2]$ variances are shown in parentheses.

In the next section, we provide simulation results for single failure localization for the detection and localization algorithm assuming the failure amplitude known or unknown, applied to the scenarios of Section II-A and Section II-B, respectively.

D. Simulations

In this section, we focus on the application of the algorithms designed in Sections IV-A and IV-B for single node failure in the scenario of Section II-A and single parameter change in the scenario of Section II-B.

1) Node failure in the scenario of Section II-A: Our first application example relates to the sensor network model $y = H\theta + \sigma w$ of Section II-A for $N = 10$ nodes, $p = N$, and $\sigma^2 = -20$ dB. This is depicted in Figure 2, where the entries of $HH^* + \sigma^2 I_N$ are presented. We also take $\sigma_k^2 = \sum_{i=1}^N (HH^*)_{ki}$, which is a natural assumption to avoid that a mere energy detector on $y(k)$ provide a simpler solution to our problem. This failure amplitude is assumed known by the experimenter. In practical scenarios, this may arise if a sensor starts returning time delayed data, supposedly uncorrelated with real-time data but with same variance. We assume a single failure scenario. In this context, it appears that, for all k, $\omega_{k,1} > 0$, $\omega_{k,2} < 0$ and $\omega_{k,1}$ is much larger than $|\omega_{k,2}|$. It is therefore more interesting only to consider the largest eigenvalue of $\Sigma\Sigma^*$ to detect and locate an hypothetical node failure. Under these conditions, the theoretical threshold for $c_N = N/n$ (if N, n were large) is 0.8 with the worst-case failure corresponding to a failure of node 10. We therefore carry out $100\,000$ Monte Carlo simulations of node 10 failures for n varying from 8 to 140 and under false alarm rates varying from 10^{-2} to 10^{-4}. This is depicted in Figure 3, where it can be observed that, for $n = 8$, detection and localization are barely possible, although it is clearly the starting point where detection becomes feasible. For not

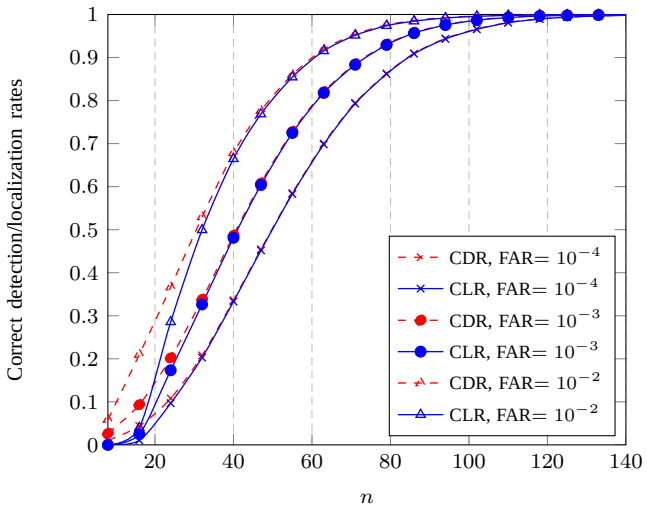

Fig. 3. Correct detection (CDR) and localization (CLR) rates for different levels of false alarm rates (FAR) and different values of n, for node 10 failure in the sensor network of Figure 2. The minimal theoretical n for observability is $n = 8$.

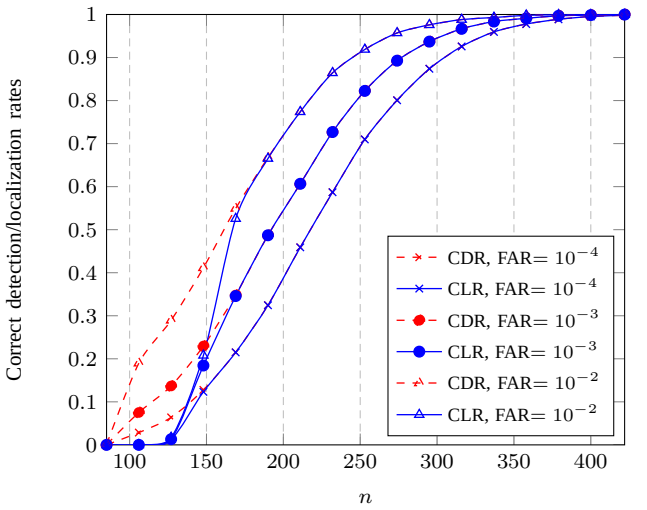

Fig. 4. Correct detection (CDR) and localization (CLR) rates for different levels of false alarm rates (FAR) and different values of n, for worst case node failure in a 100-node sensor network. The minimal theoretical n for observability is $n = 85$.

too large n, while detection rates increase, we observe that localization capabilities are still unsatisfying. This is mainly due to the inappropriate fit of the large dimensional model with $N = 10$ and with the eigenvectors corresponding to the extreme eigenvalues of $\Sigma\Sigma^*$ being too loosely correlated to their associated population eigenvectors. Larger values of n show much better performance with miss localization probability going to zero as $n \to \infty$. In particular, about five times the ratio n/N is required for localization to be very efficient. In this case, the large dimensional model for the fluctuations of the eigenvalues and eigenvectors is more adapted.

The same conditions are simulated for a system with $N = 100$ nodes in which each node has eight neighbors and with correlation values of the same order of magnitude as in Figure 2. The detectability threshold for N/n is here 0.85 and we still consider the worst case failure scenario. This is depicted in Figure 4, where one can see that smaller ratios N/n under the threshold than in the case $N = 10$ are demanded for high detectability and localization ability to appear.

2) Sudden unknown parameter change: In this section, we consider the parameter change scenario of Section II-B. We still consider the network of Figure 2, and $\sigma^2 = -20$ dB as above. We now consider a sudden change of parameter $\theta(10)$ with $\beta_{10} = 2$, being the worst case scenario for failure identification if $\beta_k = 2$ for all k. We depict the performance of the failure detection and localization algorithms and compare the settings where β_k is known or unknown in advance to the experimenter. In the former scenario, we apply the localization algorithm of Section IV-B based on the joint fluctuations of the extreme eigenvalues and eigenspace projections, while in the latter, we apply the localization algorithm of Section IV-C, where a prior step of eigenvalue inference is performed before the study of the fluctuations of the eigenspace projections. The results are presented in Figure 5.

It appears from Figure 5 that the suboptimal algorithm of Section IV-C performs only slightly worse than the algorithm of Section IV-B for large n, and that it even performs better for small n. This last observation is explained by the inadequacy of the theoretical value of ζ_k for too small values of n. It is therefore interesting to see that, for practical purposes, the absence of prior knowledge on the amplitude of the failure does not severely reduce the efficiency of the localization algorithm.

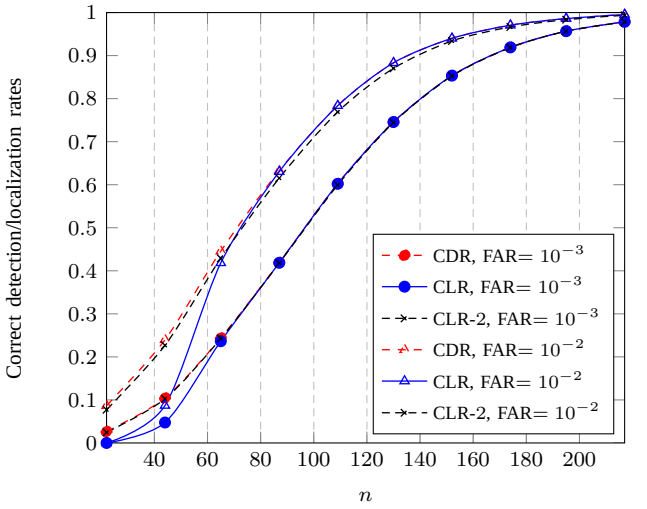

Fig. 5. Correct detection (CDR) and localization rates for different levels of false alarm rates (FAR) and different values of n, for sudden change of parameter 10 in the Scenario of Section II-B. Comparison is made between localization assuming β_k known (CLR) and localization assuming β_k unknown (CLR-2). The minimal theoretical n for observability is $n = 22$.

V. Conclusion

In this article, a characterization of the joint fluctuations of the extreme eigenvalues and corresponding eigenspace projections of a certain class of random matrices is provided. This characterization was used to perform fast and computationally

reasonable detection and localization of multiple failures in large sensor networks through a general hypothesis testing framework. The main practical outcomes of this article lie first in a characterization of the minimum number of observations necessary to ensure failure detectability in large networks and second in the design of flexible but simple algorithms that can be adapted to multiple types of failure scenarios consistent with the small rank perturbation random matrix model. We also extend the detection and diagnosis approach to scenarios where the amplitudes of the hypothetical failures are not a priori known. Practical simulations suggest that the proposed algorithms allow for high failure detection and localization performance even for networks of small sizes, although for those much more observations than theoretically predicted are in general demanded.

APPENDIX A
PROOFS OF RESULTS OF SECTION III

A. Proof of Lemma 2

We shall assume without loss of generality that $i = 1$. In Section III-B2, we saw that

$$U_1^* \widehat{\Pi}_1 U_1 = \frac{1}{2\pi\imath} \oint_{\gamma_1} \widehat{A}_1(z)^* \widehat{H}(z)^{-1} \widehat{A}_2(z) dz$$

with probability one, where

$$\begin{aligned} \widehat{A}_1(z)^* &= z U_1^* (I_N + P)^{-\frac{1}{2}} Q(z) U \\ &= \frac{z}{(1+\omega_1)^{1/2}} U_1^* Q(z) U \end{aligned}$$

and

$$\begin{aligned} \widehat{A}_2(z) &= \Omega (I_r + \Omega)^{-1} U^* Q(z) (I_N + P)^{-\frac{1}{2}} U_1 \\ &= (1+\omega_1)^{-1/2} \Omega (I_r + \Omega)^{-1} U^* Q(z) U_1 \end{aligned}$$

(take b_1 and b_2 as any two columns of U_1 in (8)). Similarly,

$$\zeta_1 I_{j_1} = \frac{1}{2\pi\imath} \oint_{\gamma_1} A_1(z)^* H(z)^{-1} A_2(z) dz$$

where

$$\begin{aligned} A_1(z)^* &= z m(z) U_1^* (I + P)^{-1/2} U \\ &= \frac{z m(z)}{(1+\omega_1)^{1/2}} \begin{bmatrix} I_{j_1} & 0 \end{bmatrix}, \\ A_2(z) &= m(z) \Omega (I + \Omega)^{-1} U^* (I + P)^{-1/2} U_1 \\ &= \frac{\omega_1 m(z)}{(1+\omega_1)^{3/2}} \begin{bmatrix} I_{j_1} \\ 0 \end{bmatrix}. \end{aligned}$$

Fixing $z \in \gamma_1$, we have

$$\widehat{H}(z)^{-1} = H(z)^{-1} - H(z)^{-1} E(z) H(z)^{-1} + \mathcal{O}(\|E(z)\|^2)$$

where

$$E(z) = \widehat{H}(z) - H(z) = z\Omega (I_r + \Omega)^{-1} U^* (Q(z) - m(z) I) U.$$

Guided by this observation, we write

$$\begin{aligned} V_{1,n} &= \frac{\sqrt{N}}{2\pi\imath} \oint_{\gamma_1} \Big(\widehat{A}_1(z)^* \widehat{H}(z)^{-1} \widehat{A}_2(z) \\ &\qquad - A_1(z)^* H(z)^{-1} A_2(z) \Big) \, dz \quad \text{(a.s., } n \text{ large)} \\ &= \frac{\sqrt{N}}{2\pi\imath} \oint_{\gamma_1} (\widehat{A}_1(z)^* - A_1(z)^*) H(z)^{-1} A_2(z) \, dz \\ &\quad + \frac{\sqrt{N}}{2\pi\imath} \oint_{\gamma_1} A_1(z)^* H(z)^{-1} (\widehat{A}_2(z) - A_2(z)) \, dz \\ &\quad - \frac{\sqrt{N}}{2\pi\imath} \oint_{\gamma_1} A_1(z)^* H(z)^{-1} E(z) H(z)^{-1} A_2(z) \, dz \\ &\quad + \mathcal{E} \\ &= Z_1 + Z_2 + Z_3 + \mathcal{E} \end{aligned}$$

where $\mathcal{E}$ contains all the higher order terms that appear when we develop the integrand at the right hand side of the first equality.

In what follows, we successively study each of the terms at the right hand side of this equation. Recalling (9), the term Z_1 writes

$$Z_1 = \frac{\sqrt{N}}{2\pi\imath} \oint_{\gamma_1} \frac{z m(z) \, U_1^* (Q(z) - m(z) I) U_1}{(1+\omega_1) \left((1+\omega_1)/\omega_1 + z m(z) \right)} dz.$$

The denominator has one simple zero in $\mathrm{Int}(\gamma_1)$. With probability one, the numerator has no zero in $\mathrm{Int}(\gamma_1)$. Using the residue theorem and the identity $(1+\omega_1)^{-1} = (1 + h(\rho_1))/h(\rho_1)$, we obtain

$$Z_1 = \frac{1 + h(\rho_1)}{h'(\rho_1)} \sqrt{N} U_1^* (Q(\rho_1) - m(\rho_1) I) U_1.$$

Similarly, the term Z_2 writes

$$Z_2 = Z_1 = \frac{1 + h(\rho_1)}{h'(\rho_1)} \sqrt{N} U_1^* (Q(\rho_1) - m(\rho_1) I) U_1.$$

Turning to Z_3, we have

$$Z_3 = -\frac{\sqrt{N}}{2\pi\imath} \oint_{\gamma_1} \frac{z^2 m(z)^2 \; U_1^* (Q(z) - m(z)) U_1}{(1+\omega_1) \left((1+\omega_1)/\omega_1 + z m(z) \right)^2} dz$$

which shows that we have a pole with degree 2 in $\mathrm{Int}(\gamma_1)$. Write the integrand as $G(z)/g(z)$ and recall that the residue of a meromorphic function $f(z)$ associated with a degree 2 pole at z_0 is $\lim_{z \to z_0} d\left((z - z_0)^2 f(z) \right)/dz$. After some simple calculations, this results in

$$\begin{aligned} Z_3 &= \frac{G(\rho_1) g''(\rho_1)}{g'(\rho_1)^3} - \frac{G'(\rho_1)}{g'(\rho_1)^2} \\ &= \frac{1 + h(\rho_1)}{h'(\rho_1)} \left(\frac{h(\rho_1) h''(\rho_1)}{h'(\rho_1)^2} - 2 \right) \\ &\quad \times \sqrt{N} U_1^* (Q(\rho_1) - m(\rho_1) I) U_1 \\ &\quad - \frac{h(\rho_1)(1 + h(\rho_1))}{h'(\rho_1)^2} \sqrt{N} U_1^* (Q'(\rho_1) - m'(\rho_1) I) U_1. \end{aligned}$$

We now show briefly that the last term $\mathcal{E}$ in the expression of $V_{1,n}$ converges to zero in probability. A more detailed argument is given in [29]. Recall that $\mathcal{E}$ accounts for all the higher order terms that show up when we expand the integrand

$\widehat{A}_1\widehat{H}^{-1}\widehat{A}_2 - A_1H^{-1}A_2$. Let us focus on one of these terms, namely

$$\frac{\sqrt{N}}{2\pi\imath}\oint_{\gamma_1} A_1(z)^* \times \left(\widehat{H}(z)^{-1} - H(z)^{-1} + H(z)^{-1}E(z)H(z)^{-1}\right) A_2(z)dz$$

and show that it converges in probability to zero. The other terms can be treated similarly. First, we can show that

$$\|\widehat{H}(z)^{-1} - H(z)^{-1} + H(z)^{-1}E(z)H(z)^{-1}\| \leq K\|E(z)\|^2$$

on γ_1 where K is some constant. Now we write

$$\begin{aligned} E(z) &= z\Omega(I_r+\Omega)^{-1}U^*(Q(z)-\alpha(z)I)U \\ &\quad + z(\alpha(z)-m(z))\Omega(I_r+\Omega)^{-1} \\ &= E_1(z) + E_2(z). \end{aligned}$$

Noticing that $A_1(z)$ and $A_2(z)$ are bounded on γ_1, and writing $z = \rho_1 + R\exp(2\imath\pi\theta)$ on γ_1, the result is shown if we show that

$$\sqrt{N}\int_0^1 \|E_i(\rho_1 + Re^{2\imath\pi\theta})\|^2 d\theta \xrightarrow{\mathbb{P}} 0 \tag{19}$$

for $i = 1, 2$. Lemma 1 shows that $\mathrm{E}\|E_1(z)\|^2 \leq K'/n$ on γ_1 where the constant K' is independent of z. By Markov's inequality, (19) is true for $i = 1$. Convergence for $i = 2$ is obtained from Assumption **A4** in conjunction with the analyticity of $\alpha(z) - m(z)$, as shown in [29].

Taking the sum $Z_1 + Z_2 + Z_3$, we obtain the desired result.

B. Proof of Lemma 3

Set $i = 1$ and write

$$\Omega(I_r+\Omega)^{-1} = \begin{bmatrix} \beta_1 I_{j_1} & \\ & B_1 \end{bmatrix}.$$

Let $y_n = \rho_1 + x/\sqrt{N}$. Write $U = \begin{bmatrix} U_1 & \widetilde{U}_1 \end{bmatrix}$ and

$$\begin{aligned} \widehat{H}(y_n) &= \begin{bmatrix} I_{j_1} + \beta_1 y_n U_1^* Q(y_n) U_1 & \beta_1 y_n U_1^* Q(y_n)\widetilde{U}_1 \\ y_n B_1 \widetilde{U}_1^* Q(y_n) U_1 & I_{r-j_1} + y_n B_1 \widetilde{U}_1^* Q(y_n)\widetilde{U}_1 \end{bmatrix} \\ &= \begin{bmatrix} \widehat{H}_{11} & \widehat{H}_{12} \\ \widehat{H}_{21} & \widehat{H}_{22} \end{bmatrix}. \end{aligned}$$

Let us study

$$N^{\frac{j_1}{2}} \det \widehat{H}(y_n) = \det(\widehat{H}_{22}) \det(\sqrt{N}\widehat{H}_{11} - \sqrt{N}\widehat{H}_{12}\widehat{H}_{22}^{-1}\widehat{H}_{21}).$$

By Lemma 1,

$$\widehat{H}_{22} \xrightarrow{\text{a.s.}} \begin{bmatrix} (1-\beta_2/\beta_1)I_{j_2} & & \\ & \ddots & \\ & & (1-\beta_t/\beta_1)I_{j_t} \end{bmatrix}$$

hence

$$\det \widehat{H}_{22} \xrightarrow{\text{a.s.}} \prod_{\ell>1}(1-\beta_\ell/\beta_1)^{j_\ell}.$$

By the same lemma,

$$\mathrm{E}\left[\sqrt{N}1_{\sigma(XX^*)\in[a-\varepsilon,b+\varepsilon]}\|\widehat{H}_{12}\|^2\right] \leq K/\sqrt{N}$$

and

$$\mathrm{E}\left[1_{\sigma(XX^*)\in[a-\varepsilon,b+\varepsilon]}\|\widehat{H}_{21}\|\right] \leq K'/\sqrt{N}$$

where K and K' are some constants, hence

$$\det(\sqrt{N}\widehat{H}_{11} - \sqrt{N}\widehat{H}_{12}\widehat{H}_{22}^{-1}\widehat{H}_{21}) - \det(\sqrt{N}\widehat{H}_{11}) \xrightarrow{\mathbb{P}} 0.$$

Let us study this last term. We write

$$\begin{aligned} \sqrt{N}\widehat{H}_{11} &= \sqrt{N}\beta_1 y_n U_1^*(Q(y_n) - \alpha(y_n)I_N)U_1 \\ &\quad + \sqrt{N}\beta_1(y_n\alpha(y_n) - \rho_1\alpha(\rho_1))I_{j_1} \\ &\quad + \sqrt{N}\beta_1(\rho_1\alpha(\rho_1) - \rho_1 m(\rho_1))I_{j_1} \\ &= Z_1 + Z_2 + Z_3. \end{aligned}$$

Writing

$$\begin{aligned} Q(y_n) - Q(\rho_1) &= Q(y_n)(Q(\rho_1)^{-1} - Q(y_n)^{-1})Q(\rho_1) \\ &= N^{-1/2}xQ(y_n)Q(\rho_1) \end{aligned}$$

we have

$$\begin{aligned} &Z_1 - \sqrt{N}\beta_1 y_n U_1^*(Q(\rho_1) - \alpha(\rho_1)I)U_1 \\ &= x\beta_1 y_n U_1^* Q(y_n)Q(\rho_1) - N^{-1}(\operatorname{tr} Q(y_n)Q(\rho_1))I)U_1 \\ &\xrightarrow{\mathbb{P}} 0 \end{aligned}$$

by Lemma 1. From **A4**, we further have

$$Z_1 - \sqrt{N}\beta_1\rho_1 U_1^*(Q(\rho_1) - m(\rho_1)I)U_1 \xrightarrow{\mathbb{P}} 0.$$

Turning to the second term, we can show using **A4** that

$$Z_2 = \beta_1 x \frac{y_n\alpha(y_n) - \rho_1\alpha(\rho_1)}{x/\sqrt{N}} \xrightarrow{\text{a.s.}} \beta_1 x(\rho_1 m(\rho_1))'.$$

Again by **A4**, $Z_3 \xrightarrow{\text{a.s.}} 0$. This results in

$$\begin{aligned} &N^{j_1/2}\det\widehat{H}(\rho_1 + x/\sqrt{N}) - \left(\prod_{\ell>1}(1-\beta_\ell/\beta_1)^{j_\ell}\right) \\ &\times \det\left(\sqrt{N}\beta_1\rho_1 U_1^*(Q(\rho_1) - m(\rho_1)I)U_1 + \beta_1 x h'(\rho_1)\right) \\ &\xrightarrow{\mathbb{P}} 0. \end{aligned}$$

The same argument for $i > 1$ leads to the result.

C. Proof of Lemma 4

Recall that XX^* admits the spectral factorization $XX^* = W\Lambda W^*$ where W and $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_N)$ are independent, and where W is Haar distributed on the group of $N \times N$ unitary matrices.

From Assumption **A4** and the analyticity of f_1, we can show as in [34] that

$$\sqrt{N}\left(\frac{1}{N}\sum_{k=1}^N f_1(\lambda_k) - \int f_1(\lambda)d\pi(\lambda)\right) \xrightarrow{\mathbb{P}} 0$$

hence the lemma is shown if we show the result on

$$\underline{S}_n = \left(\sqrt{N}\begin{bmatrix} U_i^* f_i(XX^*)U_i - \left(\frac{1}{N}\sum_{k=1}^N f_i(\lambda_k)\right)I_{j_i} \\ U_i^* g_i(XX^*)U_i - \left(\frac{1}{N}\sum_{k=1}^N g_i(\lambda_k)\right)I_{j_i} \end{bmatrix}\right)_{i=1}^t.$$

Let Z be a $N \times r$ random matrix with independent $\mathcal{CN}(0,1)$ elements, chosen independently of Λ. Write $Z = [Z_1 \ldots Z_t]$

where the blocks Z_ℓ have the same dimensions as the U_ℓ. Then $\underline{S}_n$ is equal in distribution to

$$\sqrt{N}\times \left(\begin{bmatrix} (Z_i^*Z_i)^{-1/2}Z_i^* \left(f_i(\Lambda) - \frac{\operatorname{tr} f_i(\Lambda)}{N} I_{j_i}\right) Z_i (Z_i^*Z_i)^{-1/2} \\ (Z_i^*Z_i)^{-1/2}Z_i^* \left(g_i(\Lambda) - \frac{\operatorname{tr} g_i(\Lambda)}{N} I_{j_i}\right) Z_i (Z_i^*Z_i)^{-1/2} \end{bmatrix}\right)_{i=1}^t .$$

By the law of large numbers, $N^{-1}Z_i^*Z_i \xrightarrow{\text{a.s.}} I_{j_i}$, hence it will be enough to show the result on

$$\begin{aligned}\bar{S}_n &= \frac{1}{\sqrt{N}} \left(\begin{bmatrix} Z_i^* \left(f_i(\Lambda) - N^{-1}\operatorname{tr} f_i(\Lambda) I_{j_i}\right) Z_i \\ Z_i^* \left(g_i(\Lambda) - N^{-1}\operatorname{tr} g_i(\Lambda) I_{j_i}\right) Z_i \end{bmatrix}\right)_{i=1}^t \\ &= \frac{1}{\sqrt{N}} \sum_{k=1}^N \left(\begin{bmatrix} \left(f_i(\lambda_k) - \frac{\operatorname{tr} f_i(\Lambda)}{N}\right) \begin{pmatrix} z_{i,k}z_{i,k}^* - I_{j_i} \end{pmatrix} \\ \left(g_i(\lambda_k) - \frac{\operatorname{tr} g_i(\Lambda)}{N}\right) \begin{pmatrix} z_{i,k}z_{i,k}^* - I_{j_i} \end{pmatrix} \end{bmatrix}\right)_{i=1}^t\end{aligned}$$

where we have written $Z_i^* = [z_{i,1} \dots z_{i,N}]$. Write $c_{i,k} = f_i(\lambda_k) - N^{-1}\operatorname{tr} f_i(\Lambda)$ and $d_{i,k} = g_i(\lambda_k) - N^{-1}\operatorname{tr} g_i(\Lambda)$. Up to an element rearrangement, $\bar{S}_n$ can be rewritten as the $\mathbb{R}^{2(j_1^2+\cdots+j_t^2)}$-valued vector

$$\frac{1}{\sqrt{N}} \sum_{k=1}^N \left[v_{i,k} \otimes \begin{bmatrix} c_{i,k} \\ d_{i,k} \end{bmatrix} \right]_{i=1}^t \tag{20}$$

where for every $i = 1, \dots, t$, the N vectors $v_{i,k}$ are valued in $\mathbb{R}^{j_i^2}$. We shall show that this sum converges in law to a Gaussian $\mathbb{R}^{2(j_1^2+\cdots+j_t^2)}$-valued random vector whose covariance matrix is equal to the covariance matrix of (12) rearranged similarly to $\bar{S}_n$.

We observe that the summands of (20) are centered and are independent conditionally to Λ. Observe also that for every k, vectors $(v_{i,k})_{i=1}^t$ are independent and that the elements of each of these vectors are decorrelated. Based on **A2** and **A3**, we have

$$\begin{aligned}&\frac{1}{N}\sum_{k=1}^N \mathrm{E}\left[\left[v_{i,k} \otimes \begin{bmatrix} c_{i,k} \\ d_{i,k} \end{bmatrix} \right]_{i=1}^t \left(\left[v_{i,k} \otimes \begin{bmatrix} c_{i,k} \\ d_{i,k} \end{bmatrix} \right]_{i=1}^t \right)^{\mathsf{T}} \middle| \Lambda \right] \\ &= \operatorname{diag}\left(I_{j_i} \otimes \left(\frac{1}{N}\sum_{k=1}^N \begin{bmatrix} c_{i,k}^2 & c_{i,k}d_{i,k} \\ c_{i,k}d_{i,k} & d_{i,k}^2 \end{bmatrix} \right) \right)_{i=1}^t \\ &\xrightarrow{\text{a.s.}} \operatorname{diag}\left(I_{j_i} \otimes R_i\right)_{i=1}^t\end{aligned}$$

which coincides with the covariance matrix of (12) after the rearrangement.

Furthermore, thanks to **A3**, it is easy to see that the Lyapunov condition

$$\frac{1}{N^{1+\eta}}\sum_{k=1}^N \mathrm{E}\left[\left\| \left[v_{i,k} \otimes \begin{bmatrix} c_{i,k} \\ d_{i,k} \end{bmatrix} \right]_{i=1}^t \right\|^{2+2\eta} \middle| \Lambda \right] \xrightarrow{\text{a.s.}} 0$$

is valid for any $\eta > 0$, hence (20) satisfies the conditions of the central limit theorem, which proves the lemma.

REFERENCES

[1] J. Mitola III and G. Q. Maguire Jr, "Cognitive radio: making software radios more personal," *IEEE Personal Commun. Mag.*, vol. 6, no. 4, pp. 13–18, 1999.

[2] M. Amin and B. F. Wollenberg, "Toward a smart grid: power delivery for the 21st century," *IEEE Power and Energy Magazine*, vol. 3, no. 5, pp. 34–41, 2005.

[3] A. G. Phadke, "Synchronized phasor measurements in power systems," *IEEE Computer Applications in Power*, vol. 6, no. 2, pp. 10–15, 1993.

[4] R. Kinney, P. Crucitti, R. Albert, and V. Latora, "Modeling cascading failures in the North American power grid," *The European Physical Journal B-Condensed Matter and Complex Systems*, vol. 46, no. 1, pp. 101–107, 2005.

[5] M. Basseville and I. V. Nikiforov, *Detection of abrupt changes: theory and application*. Citeseer, 1993, vol. 10.

[6] I. V. Nikiforov, "A generalized change detection problem," *IEEE Trans. Inf. Theory*, vol. 41, no. 1, pp. 171–187, 1995.

[7] J. Unnikrishnan, V. V. Veeravalli, and S. Meyn, "Minimax robust quickest change detection," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1604–1614, 2011.

[8] V. V. Veeravalli, "Decentralized quickest change detection," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1657–1665, 2001.

[9] S. Dayanik and C. Goulding, "Sequential detection and identification of a change in the distribution of a markov-modulated random sequence," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3323–3345, 2009.

[10] S. W. Choi, C. Lee, J. M. Lee, J. H. Park, and I. B. Lee, "Fault detection and identification of nonlinear processes based on kernel pca," *Chemometrics and intelligent laboratory systems*, vol. 75, no. 1, pp. 55–67, 2005.

[11] M. Basseville, M. Abdelghani, and A. Benveniste, "Subspace-based fault detection algorithms for vibration monitoring," *Automatica*, vol. 36, no. 1, pp. 101–109, 2000.

[12] J. Neyman and E. S. Pearson, "On the use and interpretation of certain test criteria for purposes of statistical inference part i," *Biometrika*, vol. 20, no. 1-2, p. 175, 1928.

[13] I. M. Johnstone, "On the distribution of the largest eigenvalue in principal components analysis," *Annals of Statistics*, vol. 99, no. 2, pp. 295–327, 2001.

[14] Z. D. Bai and J. W. Silverstein, "No Eigenvalues Outside the Support of the Limiting Spectral Distribution of Large Dimensional Sample Covariance Matrices," *Annals of Probability*, vol. 26, no. 1, pp. 316–345, Jan. 1998.

[15] J. Baik and J. W. Silverstein, "Eigenvalues of large sample covariance matrices of spiked population models," *Journal of Multivariate Analysis*, vol. 97, no. 6, pp. 1382–1408, 2006.

[16] V. A. Marčenko and L. A. Pastur, "Distributions of eigenvalues for some sets of random matrices," *Math USSR-Sbornik*, vol. 1, no. 4, pp. 457–483, Apr. 1967.

[17] R. Rao and J. W. Silverstein, "Fundamental limit of sample generalized eigenvalue based detection of signals in noise using relatively few signal-bearing and noise-only samples," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 4, no. 3, pp. 468–480, 2010.

[18] F. Benaych-Georges and R. Rao, "The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices," *arXiv Preprint, arXiv:0910.2120*, 2009.

[19] C. A. Tracy and H. Widom, "On orthogonal and symplectic matrix ensembles," *Communications in Mathematical Physics*, vol. 177, no. 3, pp. 727–754, 1996.

[20] K. Johansson, "Shape Fluctuations and Random Matrices," *Communications of Mathematical Physics*, vol. 209, pp. 437–476, 2000.

[21] J. Baik, G. Ben Arous, and S. Péché, "Phase transition of the largest eigenvalue for non-null complex sample covariance matrices," *The Annals of Probability*, vol. 33, no. 5, pp. 1643–1697, 2005.

[22] O. N. Feldheim and S. Sodin, "A universality result for the smallest eigenvalues of certain sample covariance matrices," *Geometric And Functional Analysis*, vol. 20, no. 1, pp. 88–123, 2010.

[23] L. S. Cardoso, M. Debbah, P. Bianchi, and J. Najim, "Cooperative spectrum sensing using random matrix theory," in *IEEE Pervasive Comput. (ISWPC'08)*, Santorini, Greece, May 2008, pp. 334–338.

[24] P. Bianchi, J. Najim, M. Maida, and M. Debbah, "Performance of Some Eigen-based Hypothesis Tests for Collaborative Sensing," *IEEE Trans. Inf. Theory*, 2010, to appear.

[25] D. Paul, "Asymptotics of sample eigenstructure for a large dimensional spiked covariance model," *Statistica Sinica*, vol. 17, no. 4, p. 1617, 2007.

[26] F. Benaych-Georges, A. Guionnet, and M. Maïda, "Fluctuations of the extreme eigenvalues of finite rank deformations of random matrices," *arXiv preprint*, 2011.

[27] F. Benaych-Georges and N. R. Rao, "The singular values and vectors of low rank perturbations of large rectangular random matrices," *arXiv preprint*, 2011.

[28] L. Pastur and V. Vasilchuk, “On the law of addition of random matrices: covariance and the central limit theorem for traces of resolvent,” in *Probability and mathematical physics*, ser. CRM Proc. Lecture Notes. Providence, RI: Amer. Math. Soc., 2007, vol. 42, pp. 399–416.
[29] W. Hachem, P. Loubaton, X. Mestre, J. Najim, and P. Vallet, “A subspace estimator of finite rank perturbations of large random matrices,” *arXiv preprint*, 2011.
[30] R. Couillet and M. Debbah, *Random matrix methods for wireless communications*, 1st ed. New York, NY, USA: Cambridge University Press, 2011, to appear.
[31] F. Benaych-Georges and N. R. Rao, “The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices,” *to be published in Advances in Mathematics*, 2011, arXiv.
[32] P. Billingsley, *Probability and Measure*, 3rd ed. Hoboken, NJ: John Wiley & Sons, Inc., 1995.
[33] W. Rudin, *Real and complex analysis*, 3rd ed. McGraw-Hill Series in Higher Mathematics, May 1986.
[34] Z. D. Bai and J. W. Silverstein, “CLT of linear spectral statistics of large dimensional sample covariance matrices,” *Annals of Probability*, vol. 32, no. 1A, pp. 553–605, 2004.
[35] W. Hachem, P. Loubaton, and J. Najim, “Deterministic Equivalents for Certain Functionals of Large Random Matrices,” *Annals of Applied Probability*, vol. 17, no. 3, pp. 875–930, 2007.
[36] ——, “A CLT for Information Theoretic Statistics of Gram Random Matrices with a Given Variance Profile,” *Annals of Probability*, vol. 18, no. 6, pp. 2071–2130, Dec. 2008.
[37] R. J. Muirhead, *Aspects of multivariate statistical theory*. Wiley Online Library, 1982.
[38] P. Bianchi, M. Debbah, and J. Najim, “Asymptotic independence in the spectrum of the Gaussian Unitary Ensemble,” *Electronic Communications in Probability*, no. 15, pp. 376–395, 2010. [Online]. Available: http://arxiv.org/abs/0811.0979
[39] A. W. Van der Vaart, *Asymptotic statistics*. New York: Cambridge University Press, 2000.