Information and Inference: A Journal of the IMA (2019) **8**, 289–312
doi:10.1093/imaiai/iay010
Advance Access publication on 3 July 2018

Ensemble-based estimates of eigenvector error for empirical covariance matrices

Dane Taylor[†]
Department of Mathematics, University at Buffalo, State University of New York, Buffalo, NY 14260, USA, Statistical and Applied Mathematical Sciences Institute, Research Triangle Park, NC 27709, USA and Department of Mathematics, University of North Carolina, Chapel Hill, NC 27599, USA
[†]Corresponding author. Email: danet@buffalo.edu

Juan G. Restrepo
Department of Applied Mathematics, University of Colorado, Boulder, CO 80309, USA
juanga@colorado.edu

and

François G. Meyer
Department of Electrical, Computer and Energy Engineering, University of Colorado, Boulder, CO 80309, USA
fmeyer@colorado.edu

[Received on 27 December 2016; revised on 28 February 2018; accepted on 6 April 2018]

Covariance matrices are fundamental to the analysis and forecast of economic, physical and biological systems. Although the eigenvalues $\{\lambda_i\}$ and eigenvectors $\{\boldsymbol{u}_i\}$ of a covariance matrix are central to such endeavours, in practice one must inevitably approximate the covariance matrix based on data with finite sample size n to obtain empirical eigenvalues $\{\tilde{\lambda}_i\}$ and eigenvectors $\{\tilde{\boldsymbol{u}}_i\}$, and therefore understanding the error so introduced is of central importance. We analyse eigenvector error $\|\boldsymbol{u}_i - \tilde{\boldsymbol{u}}_i\|^2$ while leveraging the assumption that the true covariance matrix having size p is drawn from a matrix ensemble with known spectral properties—particularly, we assume the distribution of population eigenvalues weakly converges as $p \to \infty$ to a spectral density $\rho(\lambda)$ and that the spacing between population eigenvalues is similar to that for the Gaussian orthogonal ensemble. Our approach complements previous analyses of eigenvector error that require the full set of eigenvalues to be known, which can be computationally infeasible when p is large. To provide a scalable approach for uncertainty quantification of eigenvector error, we consider a fixed eigenvalue λ and approximate the distribution of the expected square error $r = \mathbb{E}\left[\|\boldsymbol{u}_i - \tilde{\boldsymbol{u}}_i\|^2\right]$ across the matrix ensemble for all $\boldsymbol{u}_i$ associated with $\lambda_i = \lambda$. We find, for example, that for sufficiently large matrix size p and sample size $n > p$, the probability density of r scales as $1/nr^2$. This power-law scaling implies that the eigenvector error is extremely heterogeneous—even if r is very small for most eigenvectors, it can be large for others with non-negligible probability. We support this and further results with numerical experiments.

Keywords: covariance matrix; empirical eigenvector; Wigner surmise; Wishart distribution; graphical model.

1. Introduction

The spectral properties of covariance matrices are a central topic in mathematics, probability and statistics (Mehta, 1991; Anderson, 2003; Hastie *et al.*, 2009; Golub & Loan, 2012) and provide a cornerstone to applications in physics, biology, economics and social science (Mantegna & Stanley, 2000; Elton *et al.*, 2009; Volkov *et al.*, 2009; Weigt *et al.*, 2009; Delvenne *et al.*, 2010; Gatti *et al.*, 2010; Bassett *et al.*, 2011). The estimation of eigenvectors of a sample covariance matrix remains a fundamental tool for these and numerous other application domains. Sample covariance matrices can be computed locally if the data set lies along a manifold, or globally if the data are organized along a linear subspace (Hastie *et al.*, 2009). Often, the practitioner has access to a generative stochastic model for the covariance matrix that can be derived from first principles or domain knowledge, and he/she needs to estimate the accuracy of eigenvectors calculated from a sample covariance matrix.

We consider the 'classical' (large sample, $n > p$) framework where one has access to n measurements of a p-dimensional vector $\boldsymbol{x}$ with covariance $\boldsymbol{C}$, which are encoded as the columns of a matrix $\boldsymbol{X}$ of size $p \times n$. The sample covariance matrix $\tilde{\boldsymbol{C}} = n^{-1}(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])^T$ is an unbiased estimator to the population covariance matrix $\boldsymbol{C}$, and the main motivation for our work is to estimate how well the sample eigenvectors $\{\tilde{\boldsymbol{u}}_i\}$ of $\tilde{\boldsymbol{C}}$ approximate the population (i.e., true) eigenvectors $\{\boldsymbol{u}_i\}$ of $\boldsymbol{C}$ in the limit when both p and n are large. If we further assume that $\tilde{\boldsymbol{C}}$ is distributed according to a Wishart distribution $W(\boldsymbol{C}, n)$ centred at $\boldsymbol{C}$ (which occurs, for example, when $\boldsymbol{x}$ follows a multivariate normal distribution), then for fixed p and $n \to \infty$, the expected error between $\tilde{\boldsymbol{u}}_i$ and $\boldsymbol{u}_i$ for $\boldsymbol{C}$ for $i \in \{1, \ldots, p\}$ is given asymptotically by (Anderson, 2003, Theorem 13.5.1)

$$\mathbb{E}\left[n\|\boldsymbol{u}_i - \tilde{\boldsymbol{u}}_i\|^2\right] \to h_i, \tag{1.1}$$

where

$$h_i \triangleq \sum_{j=1; j\neq i}^{p} \frac{\lambda_i \lambda_j}{(\lambda_i - \lambda_j)^2} \tag{1.2}$$

and $\lambda_1, \ldots, \lambda_p$ are the population eigenvalues of $\boldsymbol{C}$ (which we assume to be simple and in ascending order). One important application of the asymptotic result (1.1) is that it provides an estimate for the expected residual error between the sample and the population eigenvectors for large n,

$$\mathbb{E}\left[\|\boldsymbol{u}_i - \tilde{\boldsymbol{u}}_i\|^2\right] \approx \frac{1}{n} h_i. \tag{1.3}$$

The usefulness of (1.1)–(1.3), however, is limited by the fact that h_i requires knowledge of all p eigenvalues, which can be problematic—even computationally infeasible—when p is large. Moreover, the values $\{\lambda_i\}$ are typically unknown for empirical data, and in practice one often approximates (1.2) using $\tilde{\lambda}_i \approx \lambda_i$, where $\{\tilde{\lambda}_i\}$ are the corresponding sample eigenvalues of $\tilde{\boldsymbol{C}}$.

Thus motivated, we study (1.1)–(1.3) for the limit of large p, seeking to avoid the computation of p distinct eigenvalues by considering situations in which the right-hand side of (1.2) converges with increasing p. Defining such an extension, however, comes with several complications. One difficulty is that by allowing p to increase, one ceases to study a single population covariance matrix $\boldsymbol{C}$, but instead studies a sequence of population covariance matrices of growing size p. One must therefore make an assumption about the origin of these population covariance matrices, and herein we assume that they are

drawn from a random matrix ensemble. Moreover, because $\lim_{p\to\infty} \|\boldsymbol{u}_i - \widetilde{\boldsymbol{u}}_i\|^2$ is identical for any fixed i (i.e. since $i/p \to 0$ for fixed i), we find it more interesting to study the $p \to \infty$ limiting behaviour of (1.1)–(1.3) for fixed $\lambda_i = \lambda$, examining the associated eigenvectors $\{\widetilde{\boldsymbol{u}}_i\}$ across the ensemble. (Note that the index i can vary from one population covariance matrix to another.) Finally, in this research we will assume (1.1) as a starting point—that is, we assume $n \to \infty$ much faster than $p \to \infty$. In Section 4, we study the necessary relative scaling behaviour of p and n, finding that $n = \mathcal{O}(p^2)$ is a necessary relative scaling for the ensemble of population covariance matrices that we study.

The first main contribution of this paper is an asymptotic $p \to \infty$ estimate, $\hat{h}_i \approx h_i$, for when the population eigenvalues $\{\lambda_i\}$ are distributed according to a known limiting $p \to \infty$ spectral density $\rho(\lambda)$. The idea of taking advantage of existing *a priori* knowledge about the spectral density $\rho(\lambda)$ has led to novel insights and improved inference for covariance analyses (Bickel & Levina, 2008; Lam & Fan, 2009; Ledoit & Péché, 2011). In practice, the probability distribution $\rho(\lambda)$ can be estimated from empirical data or can sometimes be derived analytically (Mehta, 1991; Kuhn, 2008). A situation of particular interest is when the covariance matrix follows a graphical (i.e. network-based) model in which complex network properties can give rise to different spectral densities (c.f., Farkas *et al.*, 2001; Goh *et al.*, 2001; Chung *et al.*, 2003; Dorogovtsev *et al.*, 2003; Benaych-Georges & Nadakuditi, 2011; Peixoto, 2013; Zhang *et al.*, 2014; Taylor *et al.*, 2016, 2017).

Note that values h_i given by (1.2) depend on the consecutive right and left eigengaps around each eigenvalue λ_i,

$$s_i^+ \triangleq \lambda_{i+1} - \lambda_i \text{ and } s_i^- \triangleq \lambda_i - \lambda_{i-1}. \tag{1.4}$$

In the context of quantum physics, these eigengaps are often referred to as level spacings since the eigenvalues typically represent energy levels (Guhr *et al.*, 1998). Herein, we assume the population covariance matrices have eigengap statistics consistent with the Gaussian orthogonal ensemble (GOE) of random matrices, thereby allowing us to take advantage of existing theory for GOE eigengap statistics. In particular, we leverage the Wigner surmise (Wigner, 1958, 1993)

$$P(s) = \frac{\pi p^2 \rho^2(\lambda)}{2} s \exp\left(-\frac{\pi p^2 \rho^2(\lambda)}{4} s^2\right) \tag{1.5}$$

for $s \in \{s_i^+, s_i^-\}$, which is a celebrated result obtained by Eugene Wigner in the 1950s to describe the distribution of eigengaps for GOE matrices of size $p = 2$. Equation (1.5) has had an enormous impact in physics (Brody, 1973; Shklovskii *et al.*, 1993; Ellegaard *et al.*, 1995; Abul-Magd & Simbel, 1999; Pimpinelli *et al.*, 2005; Schierenberg *et al.*, 2012) and economics (Plerou *et al.*, 2002; Akemann *et al.*, 2010). It was originally introduced as a 'surmise' because it was believed to offer an accurate approximation to the eigengaps for large GOE matrices. Remarkably, over the past 5 decades there has been considerable numerical support validating the approximation's accuracy for large GOE matrices in which $p \gg 2$. Moreover, (1.5) has been observed to accurately predict the eigengap distribution for numerous empirical covariance matrices describing real-world datasets (Plerou *et al.*, 2002; Akemann *et al.*, 2010).

Our second and third main contributions utilize an extension to the Wigner surmise that approximates the joint distribution $J(s^-, s^+)$ and was derived for GOE matrices of size 3×3 (Herman *et al.*, 2007). While developing theory based on such approximations introduces error into our analysis, as we shall show, the simplicity of these *surmises* allows us to make insights that may otherwise be

unobtainable. For example, it is easy to show from (1.5) that an expected eigengap should have size

$$\mathbb{E}[s^{\pm}] = \mathscr{O}\left(\frac{1}{p\rho(\lambda_i)}\right) \tag{1.6}$$

as $p \to \infty$. See also Pastur *et al.* (2011, p. 16) in which this scaling is obtained as the 'typical spacing unit' for a random matrix with convergent spectral density. In this work, we use (1.6) to study the large-p scaling behaviour for h_i as well as the necessary relative scaling between p and n. Our approach involves introducing and estimating a probability density function $f_H(h)$ of h_i in (1.2), which describes the distribution of h_i (keeping λ_i fixed) across the population covariance matrix ensemble. We obtain estimates for $f_H(h)$ in terms of λ_i, $\rho(\lambda_i)$ and p, which are of great consequence, because they describe the expected uncertainty associated with sample eigenvectors across the ensemble of population covariance matrices associated with $\rho(\lambda)$. That is, our second and third main results offer estimates for the expected eigenvector error $\mathbb{E}\left[\|\boldsymbol{u}_i - \widetilde{\boldsymbol{u}}_i\|^2\right]$ that neither require a covariance matrix nor its eigenvalues. Importantly, the ensemble-based approach that we develop herein provides a new direction for uncertainty quantification of empirical eigenvectors that is scalable for high-dimensional (large p) data.

The paper is organized as follows. We state our main results in Section 2. In Section 3, we provide numerical simulations to support these results. In Section 4, we describe conditions in which (1.1), and thus our main results, are valid. The Appendix contains the derivations of our main results.

2. Main results

In this section, we present asymptotic ($n \to \infty$ and $p \to \infty$) approximations for the expected residual error $\mathbb{E}\left[\|\boldsymbol{u}_i - \widetilde{\boldsymbol{u}}_i\|^2\right]$ of sample eigenvectors as well as their distribution across an ensemble of population covariance matrices. We first provide preliminary discussion in Section 2.1. In Section 2.2, we present main result 1, which provides a $p \to \infty$ estimate for the right-hand side of (1.2) using the assumption that the distribution of population eigenvalues weakly converges to a spectral density $\rho(\lambda)$. In Sections 2.3 and 2.4, we present main results 2 and 3, which additionally assume that the distribution of eigengaps for population covariance matrices is the same as that for the GOE random matrix ensemble.

2.1 *Model specification and assumptions*

We consider a sequence of population covariance matrices (each denoted $\boldsymbol{C}$) of growing size $p \to \infty$ such that each is drawn from a random matrix ensemble. Let $\{\lambda_i\}_{i=1}^p$ and $\{\boldsymbol{u}_i\}_{i=1}^p$ denote the population eigenvalues and corresponding eigenvectors, respectively, for each $\boldsymbol{C}$. We make the following two assumptions regarding the eigenspectra for the matrix ensemble.

Assumption 2.1 We assume that the population eigenvalues $\{\lambda_i\}$ are simple and that the spectral density $\rho_p(\lambda) = p^{-1}\sum_{i=1}^p \delta_{\lambda_i}(\lambda)$ weakly converges as $p \to \infty$ to a limiting spectral density $\rho_p(\lambda) \to \rho(\lambda)$ that has compact support $[\lambda_{\min}, \lambda_{\max}] \subset \mathbb{R}^+$, and is continuous and differentiable on the interior of its support, $(\lambda_{\min}, \lambda_{\max})$.

Many ensembles of symmetric random matrices satisfy Assumption 2.1 (see Mehta, 1991; Anderson, 2003; Kuhn, 2008) including, for example, those described by the semicircle law (Pastur *et al.*, 2011, see Sections 2 and 6). For some applications, it may also be beneficial to posit a parametric model for $\rho(\lambda)$, which can be estimated for small p and n and extended to the entire dataset.

ASSUMPTION 2.2 We assume that the joint probability distribution $J(s^-, s^+)$ of the left and right gaps, $s_i^{\pm} = |\lambda_i - \lambda_{i\pm1}|$, around each eigenvalue λ_i is given by the following generalized Wigner surmise for the GOE:

$$J(s^-, s^+) \approx \frac{3^7 \left[p\rho(\lambda_i)\right]^5}{32\pi^3} \left[s^+ s^- (s^+ + s^-)\right] \exp\left(-\frac{\left[3p\rho(\lambda_i)\right]^2}{4\pi} \left[(s^+)^2 + (s^-)^2 + s^+ s^-\right] \right). \quad (2.1)$$

The joint distribution given by (2.1) was derived in Herman *et al.* (2007) (see equation (15)) using 3×3 GOE matrices and can be constructed as a generalization of the Wigner surmise (see (1.5)), which approximates marginal distributions for $J(s^-, s^+)$. In addition to establishing the scaling $\mathbb{E}[s^{\pm}] = \mathcal{O}\left(1/p\rho(\lambda_i)\right)$ (see (1.6)), Assumption 2.2 also implies that the p population eigenvalues are simple (i.e. distinct), $\lambda_1 < \cdots < \lambda_p$, and is akin to the 'level repulsion of eigenvalues' observed in large random matrices that states that the eigengap probability is 0 for $s^{\pm} = 0$ (Bourgade *et al.*, 2014).

As shown in Herman *et al.* (2007), (2.1) gives a very good approximation to the exact distribution, which may be expressed as an infinite-dimensional integral, and can be approximated using numerical integration and Toeplitz determinants. We note in passing that it would be interesting in the future research to connect this approach to (1.1)–(1.3); however, it is unclear whether or not this approach would allow for the type of results as we present here. In contrast, while the surmises (i.e. (1.5) and (2.1)) introduce an approximation error—which remains an open topic of great interest in random matrix theory—their simplicity allows us to gain insights that may be otherwise unobtainable. In addition, as demonstrated in our numerical simulations (see Figs 1 and 2), (2.1) provides a good approximation for the ensemble of population covariance matrices that we study.

For each population covariance matrix $\boldsymbol{C}$, we consider a sample covariance matrix $\tilde{\boldsymbol{C}} = n^{-1}(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])^T$ computed from n observations, $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$, of a random vector $\boldsymbol{x} \in \mathbb{R}^p$ and $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n]$. We denote by $\widetilde{\lambda}_1 \leqslant \cdots \leqslant \widetilde{\lambda}_p$ the p sample eigenvalues of $\tilde{\boldsymbol{C}}$ and $\widetilde{\boldsymbol{u}}_1, \cdots, \widetilde{\boldsymbol{u}}_p$ the corresponding sample eigenvectors. We assume each sample covariance matrix $\tilde{\boldsymbol{C}}$ is Wishart-distributed around $\boldsymbol{C}$, as in the case when $\boldsymbol{x}$ follows a multivariate-Gaussian distribution (Anderson, 2003).

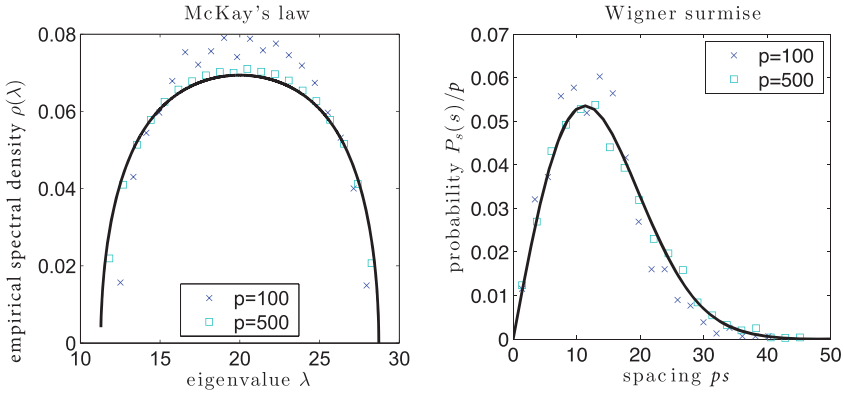

FIG. 1. Left: Empirical spectral density $\rho_p(\lambda)$ for a k-regular graphical model converges towards McKay's law (McKay, 1981) given by (3.1) (black curve) in the limit $p \to \infty$. Right: Distribution of normalized eigengaps $\{ps_i^+\}$ for eigenvalues $|\lambda_i - 20| < 1$ is well described by the Wigner surmise for GOE matrices, which is given by (1.5) (black curve).

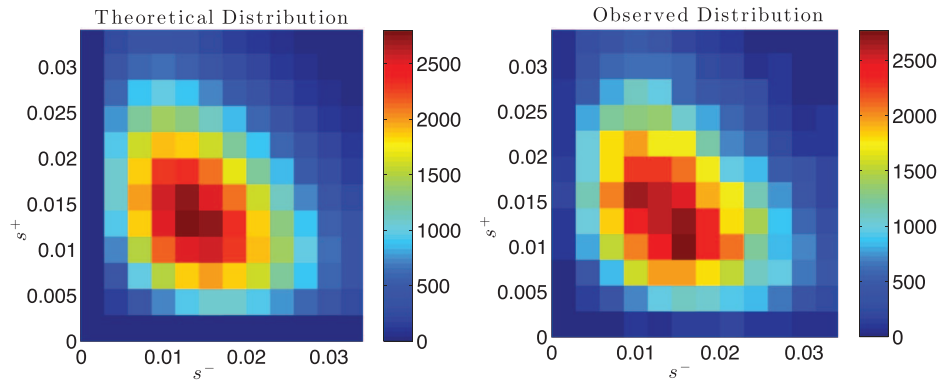

FIG. 2. Joint distribution of consecutive eigengaps $J(s^-, s^+)$: theoretical distribution (left) given by (2.1) and numerically observed distribution (right). The colour indicates the unnormalized counting measure.

2.2 *Main result* 1*: estimate of* h_i *for large* p

We may now present our first main result, an estimate $\hat{h}_i$ for h_i (see (1.2)) for high-dimensional (large p) covariance matrices. Under Assumption 2.1, we find the following asymptotic $p \to \infty$ approximation for h_i:

$$\hat{h}_i = \lambda_i^2 \left[\left(\frac{1}{(s_i^-)^2} + \frac{1}{(s_i^+)^2} \right) + p\rho(\lambda_i) \left(\frac{1}{s_i^-} + \frac{1}{s_i^+} \right) \right]. \tag{2.2}$$

See Appendix A for the derivation.

Equation (2.2) is an asymptotic approximation in that $h_i/\hat{h}_i \to 1$ as $p \to \infty$, and we can explain the role of the different terms as follows. The left term in the squared brackets approximates the terms in (1.2) that involve the nearest-neighbour eigenvalues of λ_i, which are respectively located at $\lambda_{i-1} = \lambda_i - s_i^-$ and $\lambda_{i+1} = \lambda_i + s_i^+$ and dominate the summation in (1.2) when p is large. This term does not require the knowledge of the probability distribution $\rho(\lambda)$. The second term accounts for the remaining terms in (1.2), which involve the remaining eigenvalues, $\{\lambda_j : |j-i| > 1\}$. Finally, as explained for (1.6), since $1/p\rho(\lambda_i)$ is the same order as the expected gap between two population eigenvalues (Pastur *et al.*, 2011, p. 16), all terms in the right-hand side of (2.2) can potentially obtain similar magnitudes.

Note that main result 1 does not depend on n. It approximates h_i for population covariance matrices drawn from a matrix ensemble with a convergent spectral density $\rho(\lambda)$. By combining (2.2) with (1.3), we obtain a large-p approximation to the expected error of a sample eigenvector $\widetilde{\boldsymbol{u}}_i$:

$$\mathbb{E}\left[\|\boldsymbol{u}_i - \widetilde{\boldsymbol{u}}_i\|^2 \right] \approx \hat{h}_i/n. \tag{2.3}$$

Importantly, this result uses knowledge of the population eigenvalue distribution $\rho(\lambda)$ and nearest-neighbour eigengaps $s_i^{\pm}$. Thus, it does not require knowledge of the full set of sample eigenvalues $\{\widetilde{\lambda}_i\}$ and is therefore scalable for high-dimensional (large p) data. However, we stress that approximation (2.3) is valid only when p and n are sufficiently large, which we will explore in Section 4.

2.3 *Main result 2: estimate of the probability density of h_i across matrix ensemble*

Equation (2.2) gives an asymptotic estimate for h_i that uses λ_i, s_i^+ and s_i^-, which can be calculated for a given population covariance matrix $\boldsymbol{C}$ (or estimated from $\tilde{\boldsymbol{C}}$). We now turn our attention to studying the distribution of h_i—denoted by $f_H(h_i)$—across all the population covariance matrices in the matrix ensemble for which $\lambda_i = \lambda$ is an eigenvalue. That is, we consider fixed λ_i and approximate $f_H(h_i)$ by allowing (s_i^+, s_i^-) to be distributed according to Assumption 2.2. Note that, once λ is fixed, the index i can differ from one population covariance matrix to another, and so from here on we will drop the subscript i whenever describing the distribution of a variable across the matrix ensemble.

We again consider the case where p is large, and we study the limit of the distribution of $\hat{h}$ given by (2.2) for $p \to \infty$. By combining (2.2) with (2.1), we obtain the following semi-analytical expression for the limiting probability density of the approximation $\hat{h}$ to h,

$$f_H(h) = -\int_{s^0(h)}^{\infty} J\left(s^*(h, s^+), s^+\right) \frac{\partial s^*(h, s^+)}{\partial h}\, \mathrm{d}s^+, \tag{2.4}$$

where the variables $s^0(h)$ and $s^*(h, s^+)$ depend on the eigenvalue λ around which h is computed and are given by

$$s^0(h) = \frac{\lambda^2 p \rho(\lambda)}{2h} \left\{1 + \sqrt{1 + \frac{4h}{[\lambda p \rho(\lambda)]^2}}\right\}, \tag{2.5}$$

and

$$s^*(h, s^+) = \lambda^2 p \rho(\lambda) \frac{1 + \sqrt{1 + \frac{4}{[\lambda p \rho(\lambda)]^2}\left(h - \frac{\lambda^2}{(s^+)^2} - \frac{\lambda p \rho(\lambda)}{s^+}\right)}}{2\left(h - \frac{\lambda^2}{(s^+)^2} - \frac{\lambda p \rho(\lambda)}{s^+}\right)}, \tag{2.6}$$

respectively. See Appendix B for the derivation.

The significance of (2.4) stems from the fact that it allows one to approximate the distribution of h, and therefore the distribution of expected eigenvector errors using the approximation (2.3), which again assumes sufficiently large p and n. Specifically, $f_H(h)$ estimates the distribution of expected residual error across the covariance matrix ensemble associated, that is, as opposed to (1.1), which is an estimate for a single covariance matrix from the ensemble.

2.4 *Main result 3: asymptotic behaviour of $f_H(h)$ for large h*

Keeping λ and $\rho(\lambda)$ fixed, in the limit when the left gap, s^-, or right gap, s^+, goes to zero, then h goes to infinity, and we find the following scaling behaviour for the probability density function:

$$f_H(h) = \mathscr{O}\left(\frac{p^2}{h^2}\right). \tag{2.7}$$

See Appendix C for the derivation.

We point out that the limit of large h is especially interesting because it corresponds to the case where the sample estimates $\tilde{\boldsymbol{u}}_i$ of the eigenvectors $\boldsymbol{u}_i$ are the least accurate. The observation that $f_H(h)$ has a power-law decay for large h implies that the error associated with sample eigenvectors is very heterogeneous, and one should expect situations in which the error $\mathbb{E}\left[\|\boldsymbol{u}_i - \widetilde{\boldsymbol{u}}_i\|^2\right] \approx h_i/n$ is small for

many $\widetilde{\boldsymbol{u}}_i$; but on the rare occasions in which s_i^+ and/or s_i^- are small, $\mathbb{E}\left[\|\boldsymbol{u}_i - \widetilde{\boldsymbol{u}}_i\|^2\right]$ can be orders-of-magnitude larger. Later, in Section 4 we show that $n \geqslant h_i/2$ is a necessary (but not sufficient) condition for (1.3) to offer an accurate approximation. Because $h_i = \mathcal{O}(p^2)$, we find $n = \mathcal{O}(p^2)$ to be a necessary (but not sufficient) relative scaling for (1.3) when $n, p \to \infty$. This should be further explored in a future work.

3. Numerical validation of main results

We now report the results of numerical experiments to validate the theoretical predictions given by the main results described in Section 2. In Section 3.1, we describe the matrix ensemble for population covariance matrices used for these experiments. In Section 3.2, we support main result 1. We support main results 2 and 3 in Sections 3.3 and 3.4, respectively.

3.1 *Population covariance matrix ensemble: Laplacians of k-regular graphs*

We seek to study the error of sample eigenvectors in the limit of large p and $n > p$, focusing on the scenario in which the population covariance matrices are drawn from a matrix ensemble satisfying Assumptions 2.1 and 2.2. All covariance matrices must also be positive semi-definite (Anderson, 2003) so that $\lambda_i \geqslant 0$ for all i. In addition, to help mitigate the computational cost of studying the eigenspectra for high-dimensional (large p) covariance matrices, we would like to study a sparse random matrix ensemble in which most matrix entries are zero.

Thus motivated, we study a graphical model (Hastie *et al.*, 2009). We let the population covariance matrices be given by the unnormalized—also called combinatorial (Bapat, 2010)—Laplacian matrices[1] of random k-regular graphs, which we generate using the configuration model (Newman, 2003). For k-regular graphs with fixed $k \gg 1$, the spectral density $\rho_p(\lambda) = \sum_i \delta_{\lambda_i}(\lambda)$ weakly converges as $p \to \infty$ to a semicircle distribution

$$\rho_p(\lambda) \to \rho(\lambda) = \begin{cases} \frac{k\sqrt{4(k-1)-(\lambda-k)^2}}{2\pi(k^2-(\lambda-k)^2)}, & \text{if } |(\lambda - k)| \leqslant 2\sqrt{k-1}, \\ 0, & \text{otherwise.} \end{cases} \tag{3.1}$$

Equation (3.1) is known as McKay's law (McKay, 1981). While McKay obtained (3.1) for fixed $k \gg 1$, it also describes the case for increasing k, provided that k grows sufficiently slowly with p (Dumitriu & Pal, 2012).

Numerous empirical Laplacian matrices have been observed to give rise to eigengap statistics consistent with the Wigner surmise given by (1.5) (Plerou *et al.*, 2002; Akemann *et al.*, 2010), and we therefore believe the extended surmise given by (2.1) will also be widely applicable. Importantly, our assumption that $k \gg 1$ ensures all graphs are strongly connected, which has been observed to be an important requirement for the eigengap statistics to behave similarly to that for the GOE (Murphy *et al.*, 2017). Understanding the relation between eigengap statistics and graph topology remains an important

[1] Any Laplacian matrix $\boldsymbol{C}$ is positive semi-definite: $\mathbf{v}^T\boldsymbol{C}\mathbf{v} \geqslant 0$ for any vector $\mathbf{v}$. Moreover, Laplacian matrices arise for many types of random processes on graphs and are related, for example, to the autocovariance matrices of random walks on graphs (Delvenne *et al.*, 2010). We also note that a Laplacian matrix $\boldsymbol{C}$ can be written as $\boldsymbol{C} = \boldsymbol{X}\boldsymbol{X}^T$, where $\boldsymbol{X}$ is a random incidence matrix that describes the connectivity of a random graph $G = (V, E)$, with an arbitrary orientation of the edges. Each entry $x_{e,v}$ of $\boldsymbol{X}$ can take one of three values: 1 if v is the head of the oriented edge e, -1 if v is the tail of the oriented edge e, or 0 otherwise.

open topic (Murphy *et al.*, 2017; Taylor *et al.*, 2017). In future work, it would be interesting to allow for graphs with more complicated structure—often called complex networks—and there is a large body of work exploring spectral densities for these graphs (Farkas *et al.*, 2001; Goh *et al.*, 2001; Chung *et al.*, 2003; Dorogovtsev *et al.*, 2003; Benaych-Georges & Nadakuditi, 2011; Peixoto, 2013; Zhang *et al.*, 2014; Taylor *et al.*, 2016, 2017).

We now illustrate that this graphical model satisfies Assumptions 2.1 and 2.2. Figure 1 (left) displays the empirical spectral density computed over 50 population covariance matrices sampled from the graphical model, using matrix sizes $p = 100$ (blue crosses) and $p = 500$ (cyan squares). Note that p indicates both the matrix size and the number of vertices in the k-regular graph. As p increases from 100 to 500, we observe the convergence of the empirical spectral density towards (3.1) (black curve). Figure 1 (right) displays the empirical probability density of the normalized spacing ps^+ for the set of eigenvalues $\{\lambda_i\}$ such that $|\lambda_i - 20| < 1$. We used approximately $2p\rho(\lambda)$ eigengaps to estimate the empirical densities. As expected, the eigengap distributions appear to be consistent with the Wigner surmise given by (1.5) (black curve). Note that the agreement improves with increasing p. To gain further insight into the gap distribution and to validate Assumption 2.2, we compared the unnormalized counting measure with the joint eigengap distribution $J(s^-, s^+)$ for the eigenvalues $\{\lambda_i\}$ such that $|\lambda_i - 20| < 1$. Figure 2 (left) displays the level sets of $p^2 J(s^-, s^+)$ according to (2.1) with $p = 1,000$. Figure 2 (right) shows the unnormalized counting measure computed across 100 covariance matrices of size $p = 1,000$. These are in very good agreement.

Before continuing, we need to make two clarifying points. First, while we could use (3.1) to test our approximations (see Section 2), in the numerical experiments to follow, we instead estimate the limiting distribution of eigenvalues using the data, as this approach would be more relevant for empirical data. That is, we estimate $\rho(\lambda)$ by numerically computing the average spectral density of random population covariance matrices drawn from the graphical model for a given p.

Secondly, main results 2 and 3 describe the distribution $f_H(h_i)$ across the random matrix ensemble from which population covariance matrices are drawn. That is, we consider the distribution of h_i associated with a particular eigenvalue $\lambda_i = \lambda$. However, if one fixes λ, then one is confronted with an undersampling issue since it is unlikely that the Laplacian of a randomly generated k-regular graph will have λ as a particular eigenvalue. To overcome this issue, we fix λ and numerically study the distribution $f_H(h)$ for values $\{h_i\}$ associated with eigenvalues $\{\lambda_i : |\lambda_i - \lambda| < \delta\}$ for small $\delta > 0$. For each p, we choose δ to be sufficiently small so that the resulting distribution appears to not depend on δ. We note that this represents a compromise between undersampling the random matrix ensemble and the error introduced by allowing λ_i lie within a small neighbourhood (rather than remain fixed at λ).

3.2 *Experimental validation of main result* 1

We first compared the estimate $\widehat{h}_i$, given by (2.2), with the true values of h_i, defined by (1.2), for covariance matrices ensemble described in Section 3.1. We considered graphs of fixed degree $k = 20$ and $p = 100$ vertices. In Fig. 3 (left), we compare (2.2) with the true value of h_i computed directly from the eigenvalues. The points lie close to the diagonal (dashed line), which validates the accuracy of the approximations. To illustrate the effect of the terms $p\rho(\lambda_i)\lambda_i^2/s_i^{\pm}$ in (2.2), we plot our approximation with (red plus symbols) and without (blue crosses) these corrections. One can observe that these terms improve the estimate for small h_i and have little effect for large h_i. This is expected since large h_i corresponds to very small $s_i^{\pm}$. In this limit, the correction terms become negligible as $(s_i^{\pm})^{-2} \gg (s_i^{\pm})^{-1}$.

In the next experiment, we compare $\widehat{h}_i$ given by (2.2) with a bootstrap estimate of the mean sample error, $n\widehat{\mathbb{E}}\left[\|\boldsymbol{u}_i - \widetilde{\boldsymbol{u}}_i\|^2\right]$, for Wishart distribution $W(\boldsymbol{C}, n)$. Specifically, we generated a population

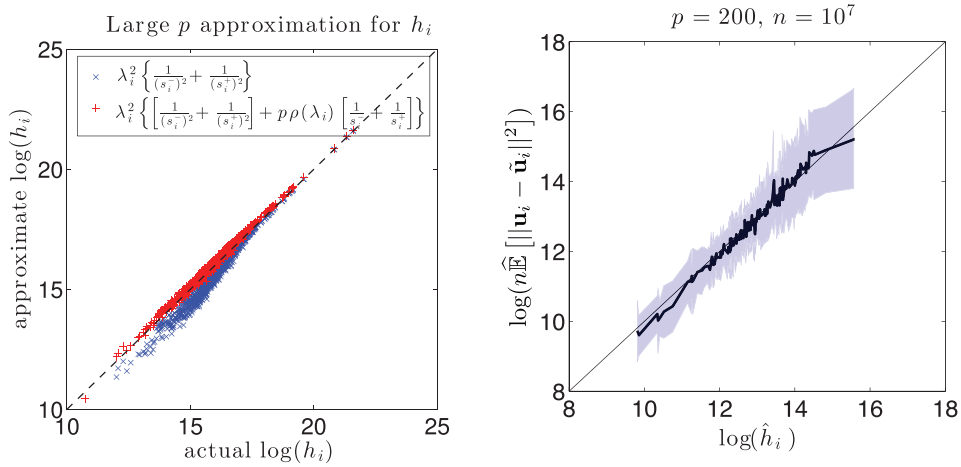

FIG. 3. Support for main result 1. Left: Approximation $\widehat{h}_i$ given by (2.2) as function of the true h_i given by (1.2). Results indicate $\hat{h}_i$ and h_i for a single population covariance matrix $\boldsymbol{C}$ of size $p = 100$ and $i \in \{1, \ldots, p\}$. We show $\widehat{h}_i$ with (red plus symbols) and without (blue crosses) the correction terms. Right: Bootstrap estimate of the sample mean error, $n\widehat{\mathbb{E}}\left[\|\boldsymbol{u}_i - \widetilde{\boldsymbol{u}}_i\|^2\right]$, which is computed from 100 samples from Wishart distribution $W(\boldsymbol{C}, n)$ with $n = 10^7$, vs. approximation $\widehat{h}_i$ given by (2.2). The mean is plotted by the black curve, and the standard deviation is shown in blue. See text for details.

covariance matrix $\boldsymbol{C}$ with $k = 20$ and $p = 200$. We then generated 100 random realizations $\widetilde{\boldsymbol{C}}$ from $W(\boldsymbol{C}, n)$ with $n = 10^7$. Let $\{\boldsymbol{u}_i\}_{i=1}^p$ be the eigenvectors of $\boldsymbol{C}$. For each random realization $\widetilde{\boldsymbol{C}}$, we calculated its eigenvectors $\{\widetilde{\boldsymbol{u}}_i\}_{i=1}^p$ and computed the residual error $\boldsymbol{u}_i - \widetilde{\boldsymbol{u}}_i$ between the sample eigenvectors and the population eigenvectors. We then computed a bootstrap estimate, $\widehat{\mathbb{E}}\left[\|\boldsymbol{u}_i - \widetilde{\boldsymbol{u}}_i\|^2\right]$, indicating the observed mean eigenvector error across the 100 realizations of $\widetilde{\boldsymbol{C}}$. In Fig. 3 (right), we plot the observed values $n\widehat{\mathbb{E}}\left[\|\boldsymbol{u}_i - \widetilde{\boldsymbol{u}}_i\|^2\right]$ vs. our prediction given by (2.3). The mean is plotted in black, and the standard deviation is shown in blue. We note that the solid curves lie very close to the diagonal indicating the accuracy of (2.3).

In these experiments, the sample size n was chosen to be sufficiently large so that (1.3) and (2.3) are accurate. Recall that (1.1) is an asymptotic $n \to \infty$ limit for $\mathbb{E}\left[n\|\boldsymbol{u}_i - \widetilde{\boldsymbol{u}}_i\|^2\right]$, and we numerically observe that n must be very large for the asymptotic result to provide an accurate approximation. We discuss in Section 4 a simple and practical bound that can be used to choose appropriate values of n.

3.3 *Experimental validation of main result* 2

We now describe experiments that validate the second main result presented in Section 2.3. We confirm that the approximation of $f_H(h)$ given by (2.4) is in good agreement with the empirical distribution of h_i. Furthermore, we show experimentally that $f_H(h)$ in (2.4) also approximates the distribution of the expected residual error $\mathbb{E}\left[n\|\boldsymbol{u}_i - \widetilde{\boldsymbol{u}}_i\|^2\right]$, provided that n and p are sufficiently large.

We generated 50 unweighted graphs of fixed degree $k = 20$ and fixed size $p = 1,000$. For each graph, we constructed the population covariance matrix, $\boldsymbol{C}$, as explained in Section 3.1. For each $\boldsymbol{C}$, we generated 10 sample covariance matrices $\widetilde{\boldsymbol{C}}$ from Wishart distribution $W(\boldsymbol{C}, n)$ with $n = 10^{10}$. For each $\widetilde{\boldsymbol{C}}$, we calculated its eigenvectors $\{\widetilde{\boldsymbol{u}}_i\}_{i=1}^p$ and computed the residual error, $\boldsymbol{u}_i - \widetilde{\boldsymbol{u}}_i$, between the sample and population eigenvectors.We consider all eigenvectors such that their associated eigenvalues satisfy

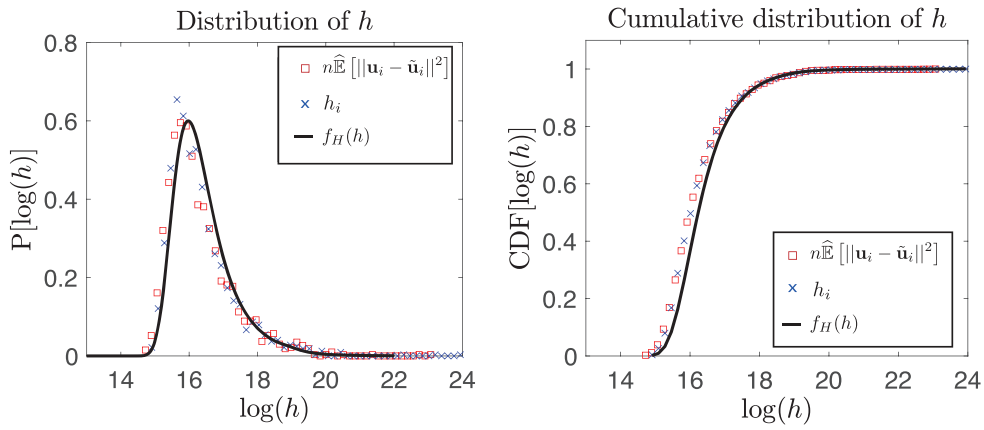

FIG. 4. Support for main result 2. Accuracy of probability density function $f_H(h)$ of h given by (2.4) (left) and its associated cumulative density (right). We depict the following: (black curves) semi-analytical expression of the probability distribution $f_H(h)$ given by (2.4), (blue crosses, ×) an empirically observed distribution of h_i computed for $\{h_i : |\lambda_i - \lambda| < 1\}$ with $\lambda = 20$, (red squares, □) empirically observed distribution of bootstrap estimate, $n\widehat{\mathbb{E}}\left[\|\boldsymbol{u}_i - \widetilde{\boldsymbol{u}}_i\|^2\right]$.

$|\lambda_i - \lambda| < 1$. We then computed a bootstrap estimate, $n\widehat{\mathbb{E}}\left[\|\boldsymbol{u}_i - \widetilde{\boldsymbol{u}}_i\|^2\right]$, of the mean sample error for each $\boldsymbol{C}$ using the 10 realizations of $\widetilde{\boldsymbol{C}}$.

In Fig. 4 (left), we use a solid black curve to represent the semi-analytical expression of the probability distribution $f_H(h)$ given by (2.4). We plot its corresponding cumulative distribution in Fig. 4 (right). We plot with blue crosses in both panels a numerically observed distribution of h_i, which we estimate using 50 covariances $\boldsymbol{C}$ drawn from the graphical model described in Section 3.1. We plot with red squares an empirical distribution of bootstrap estimates, $n\widehat{\mathbb{E}}\left[\|\boldsymbol{u}_i - \widetilde{\boldsymbol{u}}_i\|^2\right]$. As expected, the probability density function $f_H(h)$ provides a good approximation of the empirical distribution of h_i as well as the distribution of $n\widehat{\mathbb{E}}\left[\|\boldsymbol{u}_i - \widetilde{\boldsymbol{u}}_i\|^2\right]$ (that is, provided n and p are both sufficiently large). However, we note that the distribution $f_H(h)$ is shifted slightly to the right. This is in agreement with Fig. 4, where one can observe that $\hat{h}_i$ typically overestimates h_i by a very small amount (i.e. the red + symbols tend to be just above the diagonal).

3.4 *Experimental validation of main result* 3

We conclude with numerical validation of main result 3, $f_H(h) \propto h^{-2}$ for large h, which we presented in Section 2.4. We generated 500 covariance matrices $\boldsymbol{C}$ using the graphical model described in Section 3.1, with $k = 20$ and $p = 2{,}000$. Figure 5 displays $P[\log(h)]$ using our theoretical distribution $f_H(h)$ given by (2.4) (dashed red curve) as a function of $\log(h)$. We also display as a solid black line the limiting scaling behaviour, $f_H(h) \propto h^{-2}$, given by (2.7). Finally, we compare these two probability density functions with the empirical distribution of $\log(h)$, shown as blue crosses. We note for large h that all distributions are parallel in this log–log plot, indicating that they have the same asymptotic power-law scaling.

4. Discussion

A central motivator for our research has been equation (1.1), which describes the limiting $n \to \infty$ expected sample error $\|\boldsymbol{u}_i - \tilde{\boldsymbol{u}}_i\|^2$ of a sample eigenvector $\tilde{\boldsymbol{u}}_i$ for a covariance matrix drawn from a

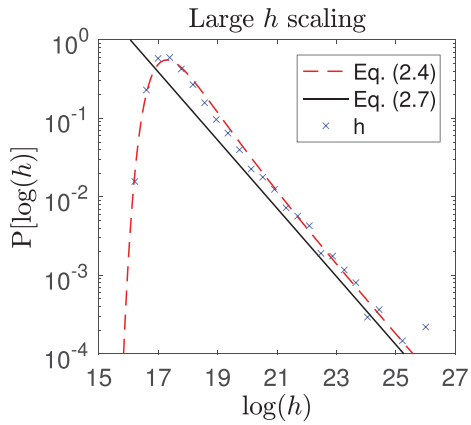

FIG. 5. Support for main result 3. $\mathcal{O}(p^2/h^2)$ scaling of $f_H(h)$ in the limit of large h. We plot P[log(h)] as a function of log(h): semi-analytical expression $f_H(h)$ computed from (2.4) in red (– –); limiting approximation to $f_H(h)$ for large h, given by (2.7) in black (–); empirical distribution, shown as blue crosses (×).

Wishart distribution. However, this equality only holds asymptotically. In this discussion, we describe the conditions in which the approximation (1.3) is expected to be accurate. That is, when is the sample size n sufficiently large for given covariance matrix size p?

The standard approach to this problem usually involves a tail bound. Instead, we use here a simple argument that yields a lower bound that works very well in practice. Indeed, we provide a necessary (but not sufficient) lower bound on n such that (1.1) and (2.2) are valid. Since both $\boldsymbol{u}_i$ and $\tilde{\boldsymbol{u}}_i$ are normalized and we assume $\boldsymbol{u}_i \approx \tilde{\boldsymbol{u}}_i$, we have

$$\|\boldsymbol{u}_i - \tilde{\boldsymbol{u}}_i\|^2 = 2[1 - \langle \boldsymbol{u}_i, \tilde{\boldsymbol{u}}_i \rangle] \leqslant 2. \tag{4.1}$$

Under the approximation $\|\boldsymbol{u}_i - \tilde{\boldsymbol{u}}_i\|^2 \approx h_i/n$ given by (3), it follows that

$$n \geqslant h_i/2. \tag{4.2}$$

We now provide numerical support for this bound using the graphical model described in Section 3.1. We first generate a k-regular graph with p vertices and compute the unnormalized Laplacian matrix, $\boldsymbol{C}$, which we treat as a covariance matrix. Let $\{\boldsymbol{u}_i\}_{i=1}^p$ be the eigenvectors of $\boldsymbol{C}$. In order to study the convergence of the empirical eigenvectors, we generate 100 random matrices $\widetilde{\boldsymbol{C}}$ from the Wishart distribution $W(\boldsymbol{C}, n)$. For each random realization $\widetilde{\boldsymbol{C}}$, we calculate its eigenvectors $\{\widetilde{\boldsymbol{u}}_i\}_{i=1}^p$ and compute the residual error $\boldsymbol{u}_i - \widetilde{\boldsymbol{u}}_i$ between the sample eigenvectors and the population eigenvectors.

Figure 6 (left) displays $\log(h_i)$ as a function of $\log(n\|\boldsymbol{u}_i - \widetilde{\boldsymbol{u}}_i\|^2)$ for each random realization of a Wishart matrix $\widetilde{\boldsymbol{C}}$ for $p = 200$, $k = 5$ and several choices of n. For each value of n, we plot the bound given by (4.2), $\log(2n)$, as a vertical solid line. Figure 6 (right) displays a scatterplot of $\log(h_i)$ as a function of $\log(n\|\boldsymbol{u}_i - \widetilde{\boldsymbol{u}}_i\|^2)$ for $k = 5$, $n = 10^5$ and several values of p. We also plot $\log(2n)$ as a vertical solid line. Both panels illustrate (4.2) as a useful bound for considering when the approximation $h_i \approx \mathbb{E}\left[n\|\boldsymbol{u}_i - \widetilde{\boldsymbol{u}}_i\|^2\right]$ given by (1.3) will be valid. Specifically, we require $\mathbb{E}\left[\|\boldsymbol{u}_i - \widetilde{\boldsymbol{u}}_i\|^2\right] < 2$ and observe considerable discrepancy as $\mathbb{E}\left[\|\boldsymbol{u}_i - \widetilde{\boldsymbol{u}}_i\|^2\right] \to 2$.

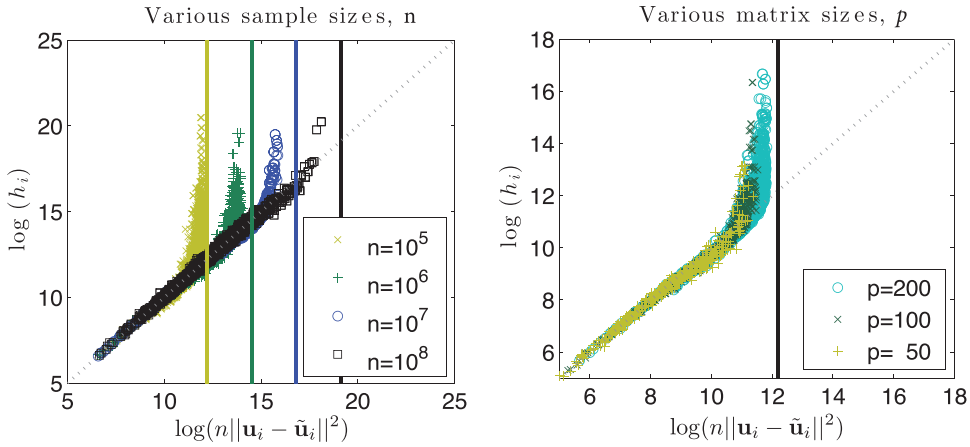

FIG. 6. True values of h_i given by (1.2) vs. residual eigenvector error $n\|\boldsymbol{u}_i - \widetilde{\boldsymbol{u}}_i\|^2$ across 100 Wishart-distributed sample covariance matrices $\widetilde{\boldsymbol{C}}$ with expectation $\boldsymbol{C}$ with (left) $p = 200$ and various n and (right) $n = 10^5$ and various p. The vertical lines indicate $h_i \leqslant 2n$ is a necessary (but not sufficient) condition for accuracy of the approximation $\mathbb{E}\left[n\|\boldsymbol{u}_i - \widetilde{\boldsymbol{u}}_i\|^2\right] \approx h_i/n$ given by (1.3).

We conclude by exploring the effect of inaccuracy for (1.3) on the distribution of the bootstrap estimates, $n\widehat{\mathbb{E}}\left[\|\boldsymbol{u}_i - \widetilde{\boldsymbol{u}}_i\|^2\right]$, of the mean residual error across the covariance matrix ensemble. We identify two sources of discrepancy: (i) choosing n too small so that the bound (4.2) is violated and (ii) using insufficiently many samples from the Wishart distribution $W(\boldsymbol{C}, n)$ to provide an accurate bootstrap estimate, $n\widehat{\mathbb{E}}\left[\|\boldsymbol{u}_i - \widetilde{\boldsymbol{u}}_i\|^2\right]$. That is, the bootstrap estimate $\widehat{\mathbb{E}}\left[\|\boldsymbol{u}_i - \widetilde{\boldsymbol{u}}_i\|^2\right]$ is only a reliable estimate of $\mathbb{E}\left[\|\boldsymbol{u}_i - \widetilde{\boldsymbol{u}}_i\|^2\right]$ if we generate enough random samples $\widetilde{\boldsymbol{C}}$ from the Wishart distribution $W(\boldsymbol{C}, n)$.

We highlight these two sources of discrepancy with a numerical experiment similar to the one described in Section 3.3. We generated 50 covariance matrices, $\boldsymbol{C}$, with $k = 20$ and $p = 1,000$, as explained in Section 3.1. For each $\boldsymbol{C}$, we generated R random realizations $\widetilde{\boldsymbol{C}}$ from the Wishart distribution $W(\boldsymbol{C}, n)$. For each $\widetilde{\boldsymbol{C}}$, we calculated its eigenvectors $\{\widetilde{\boldsymbol{u}}_i\}_{i=1}^p$ and computed the residual error $\boldsymbol{u}_i - \widetilde{\boldsymbol{u}}_i$ between the sample eigenvectors and the population eigenvectors. We then computed a bootstrap estimate, $\widehat{\mathbb{E}}\left[\|\boldsymbol{u}_i - \widetilde{\boldsymbol{u}}_i\|^2\right]$, of the mean sample error for each $\boldsymbol{C}$ using the R realizations of $\widetilde{\boldsymbol{C}}$.

In Fig. 7 (left) and Fig. 7 (right), we plot as solid black curves the probability distribution $f_H(h)$ given by (2.4) and its corresponding cumulative distribution, respectively. We plot by blue crosses the empirical distribution of h_i, which we estimate using the 50 covariances $\boldsymbol{C}$. Note that $f_H(h)$ accurately predicts the observed distribution of h_i, since p is sufficiently large. In addition, we plot the distribution of the bootstrap estimates $\widehat{\mathbb{E}}\left[n\|\boldsymbol{u}_i - \widetilde{\boldsymbol{u}}_i\|^2\right]$ of the mean sample error for $R = 10$ and $n = 10^7$ (green circles) as well as $R = 1$ and $n = 10^{10}$ (red squares). Note that when $R = 1$, the bootstrap estimate $\widehat{\mathbb{E}}\left[\|\boldsymbol{u}_i - \widetilde{\boldsymbol{u}}_i\|^2\right]$ is actually just the sample error $\|\boldsymbol{u}_i - \widetilde{\boldsymbol{u}}_i\|^2$.

Observe that both distributions disagree with $f_H(h)$ for different reasons: for $R = 10$ and $n = 10^7$, the distribution of $\widehat{\mathbb{E}}\left[n\|\boldsymbol{u}_i - \widetilde{\boldsymbol{u}}_i\|^2\right]$ is expected to differ because $n = 10^7$ is too small and does not satisfy the bound given by (4.2) (vertical dashed lines). On the other hand, for $R = 1$ and $n = 10^{10}$, the sample error does not provide a good bootstrap estimate for the mean sample error $\mathbb{E}\left[n\|\boldsymbol{u}_i - \widetilde{\boldsymbol{u}}_i\|^2\right]$, which is the relevant quantity that is described in (1.1) and (1.3). We observe in Fig. 7 that using too few samples (i.e. small R) affects the distribution of $\widehat{\mathbb{E}}\left[n\|\boldsymbol{u}_i - \widetilde{\boldsymbol{u}}_i\|^2\right]$ by shifting it towards small values of h (see red squares).

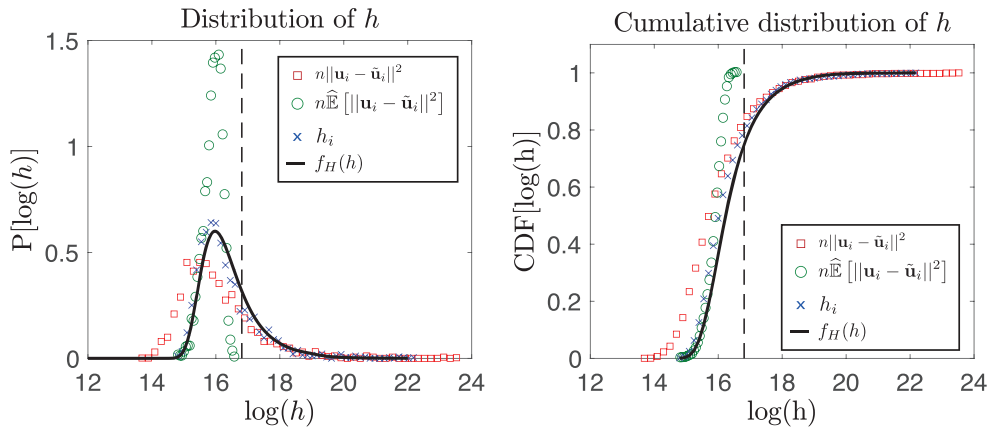

FIG. 7. Discrepancy between theoretical distribution $f_H(h)$ of h given by (2.4) and a distribution of bootstrap estimates, $\widehat{\mathbb{E}}\left[n\|\boldsymbol{u}_i - \widetilde{\boldsymbol{u}}_i\|^2\right]$, due to error for (1.1) and (1.3). We plot results for two sources of error (green circles, ○). Sample size n is too small and does not satisfy bound (4.2), which is shown by the vertical dashed lines (red squares, □). Insufficiently, many (particularly, $R = 1$) samples are used to provide a reliable bootstrap estimate $\widehat{\mathbb{E}}\left[n\|\boldsymbol{u}_i - \widetilde{\boldsymbol{u}}_i\|^2\right]$ to $\mathbb{E}\left[n\|\boldsymbol{u}_i - \widetilde{\boldsymbol{u}}_i\|^2\right]$. (See text for details.)

Funding

This material was based upon work partially supported by the National Science Foundation under Grant DMS-1127914 to the Statistical and Applied Mathematical Sciences Institute. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

ABUL-MAGD, A. & SIMBEL, M. (1999) Wigner surmise for high-order level spacing distributions of chaotic systems. *Phys. Rev. E*, **60**, 5371.

AKEMANN, G., FISCHMANN, J. & VIVO, P. (2010) Universal correlations and power-law tails in financial covariance matrices. *Physica A*, **389**, 2566–2579.

ANDERSON, T. W. (2003) *An Introduction to Multivariate Statistical Analysis*, 3rd edn. Hoboken, New Jersey: John Wiley, Inc.

BAPAT, R. B. (2010) *Graphs and Matrices*. London: Springer.

BASSETT, D. S., WYMBS, N. F., PORTER, M. A., MUCHA, P. J., CARLSON, J. M. & GRAFTON, S. T. (2011) Dynamic reconfiguration of human brain networks during learning. *Proc. Natl. Acad. Sci.*, **108**, 7641–7646.

BENAYCH-GEORGES, F. & NADAKUDITI, R. R. (2011) The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Adv. Math.*, **227**, 494–521.

BICKEL, P. J. & LEVINA, E. (2008) Covariance regularization by thresholding. *Ann. Stat.*, **36**, 2577–2604.

BOURGADE, P., ERDÖS, L. & YAU, H.-T. (2014) Edge universality of beta ensembles. *Comm. Math. Phys.*, **332**, 261–353.

BRODY, T. (1973) A statistical measure for the repulsion of energy levels. *Lett. Nuovo Cimento (1971–1985)*, **7**, 482–484.

CHUNG, F., LU, L. & VU, V. (2003) The spectra of random graphs with given expected degrees. *Proc. Natl. Acad. Sci.*, **100**, 6313–6318.

Delvenne, J.-C., Yaliraki, N., Sophia, N. & Barahona, M. (2010) Stability of graph communities across time scales. *Proc. Natl. Acad. Sci.*, **107**, 12755–12760.
Dorogovtsev, S. N., Goltsev, A. V., Mendes, J. F. F. & Samukhin, A. N. (2003) Spectra of complex networks. *Phys. Rev. E*, **68**, 046109.
Dumitriu, I. & Pal, S. (2012) Sparse regular random graphs: spectral density and eigenvectors. *Ann. Probab.*, **40**, 2197–2235.
Ellegaard, C., Guhr, T., Lindemann, K., Lorensen, H., Nygård, J. & Oxborrow, M. (1995) Spectral statistics of acoustic resonances in aluminum blocks. *Phys. Rev. Lett.*, **75**, 1546.
Elton, E. J., Gruber, M. J., Brown, S. J. & Goetzmann, W. N. (2009) *Modern Portfolio Theory and Investment Analysis*. Hoboken, New Jersey: John Wiley.
Farkas, I. J., Derényi, I., Barabasi, A.-L. & Vicsek, T. (2001) Beyond the semicircle law. *Phys. Rev. E*, **64**, 026704.
Gatti, D. M., Barry, W. T., Nobel, A. B., Rusyn, I. & Wright, F. A. (2010) Heading down the wrong pathway: on the influence of correlation within gene sets. *BMC Genomics*, **11**, 574.
Goh, K.-I., Kahng, B. & Kim, D. (2001) Spectra and eigenvectors of scale-free networks. *Phys. Rev. E*, **64**, 051903.
Golub, G. H. & Loan, C. F. V. (2012) *Matrix Computations*, vol. **3**. Baltimore, MD: JHU Press.
Guhr, T., Müller-Groeling, A. & Weidenmüller, H. A. (1998) Random-matrix theories in quantum physics: common concepts. *Phys. Rep.*, **299**, 189–425.
Hastie, T., Tibshirani, R. & Friedman, J. (2009) *The Elements of Statistical Learning*, vol. **2**. New York: Springer.
Herman, D., Ong, T. T., Usaj, G., Mathur, H. & Baranger, H. U. (2007) Level spacings in random matrix theory and Coulomb blockade peaks in quantum dots. *Phys. Rev. B*, **76**, 195448.
Kuhn, R. (2008) Spectra of sparse random matrices. *J. Phys. A*, **41**, 295002.
Lam, C. & Fan, J. (2009) Sparsistency and rates of convergence in large covariance matrix estimation. *Ann. Stat.*, **37**, 4254.
Ledoit, O. & Péché, S. (2011) Eigenvectors of some large sample covariance matrix ensembles. *Probab. Theory Related Fields*, **151**, 233–264.
Mantegna, R. & Stanley, H. (2000) *An Introduction to Econophysics*. Cambridge, UK: Cambridge University Press.
McKay, B. (1981) The expected eigenvalue distribution of a large regular graph. *J. Linear Algebra Appl.*, **40**, 203–216.
Mehta, M. L. (1991) *Random Matrices*, 3rd edn. New York: Academic Press.
Murphy, N. B., Cherkaev, E. & Golden, K. M. (2017) Anderson transition for classical transport in composite materials. *Phys. Rev. Lett.*, **118**, 036401.
Newman, M. E. J. (2003) The structure and function of complex networks. *SIAM Rev.*, **45**, 167–256.
Pastur, L. A., Shcherbina, M. & Shcherbina, M. (2011) *Eigenvalue Distribution of Large Random Matrices*, vol. **171**. Providence, RI: American Mathematical Society.
Peixoto, T. P. (2013) Eigenvalue spectra of modular networks. *Phys. Rev. Lett.*, **111**, 098701.
Pimpinelli, A., Gebremariam, H. & Einstein, T. (2005) Evolution of terrace-width distributions on vicinal surfaces: Fokker-Planck derivation of the generalized Wigner surmise. *Phys. Rev. Lett.*, **95**, 246101.
Plerou, V., Gopikrishnan, P., Rosenow, B., Amaral, L. A. N., Guhr, T. & Stanley, H. E. (2002) Random matrix approach to cross correlations in financial data. *Phys. Rev. E*, **65**, 066126.
Schierenberg, S., Bruckmann, F. & Wettig, T. (2012) Wigner surmise for mixed symmetry classes in random matrix theory. *Phys. Rev. E*, **85**, 061130.
Shklovskii, B., Shapiro, B., Sears, B., Lambrianides, P. & Shore, H. (1993) Statistics of spectra of disordered systems near the metal-insulator transition. *Phys. Rev. B*, **47**, 11487.
Taylor, D., Caceres, R. S. & Mucha, P. J. (2017) Super-resolution community detection for layer-aggregated multilayer networks. *Phys. Rev. X*, **7**, 031056.
Taylor, D., Shai, S., Stanley, N. & Mucha, P. J. (2016) Enhanced detectability of community structure in multilayer networks through layer aggregation. *Phys. Rev. Lett.*, **116**, 228301.

Volkov, I., Banavar, J. R., Hubbell, S. P. & Maritan, A. (2009) Inferring species interactions in tropical forests. *Proc. Natl. Acad. Sci.*, **106**, 13854–13859.
Weigt, M., White, R. A., Szurmant, H., Hoch, J. A. & Hwa, T. (2009) Identification of direct residue contacts in protein–protein interaction by message passing. *Proc. Natl. Acad. Sci.*, **106**, 67–72.
Wigner, E. P. (1958) On the distribution of the roots of certain symmetric matrices. *Ann. Math.*, **67**, 325–327.
Wigner, E. P. (1993) On a class of analytic functions from the quantum theory of collisions. *The Collected Works of Eugene Paul Wigner*. Berlin: Springer, pp. 409–440.
Zhang, X., Nadakuditi, R. R. & Newman, M. E. J. (2014) Spectra of random graphs with community structure and arbitrary degrees. *Phys. Rev. E*, **89**, 042816.

A. Derivation of main result 1

In this Appendix, we approximate h_i in (1.1) in terms of the nearest-neighbour eigenvalue gaps. By doing so, we will be able to exploit the knowledge of the $p \to \infty$ limiting distribution of the eigenvalues. We begin by dividing the summation into two parts so that

$$h_i = h_i^- + h_i^+, \tag{A.1}$$

with

$$h_i^- = \sum_{j=1}^{i-1} \frac{\lambda_i \lambda_j}{(\lambda_i - \lambda_j)^2}, \tag{A.2}$$

$$h_i^+ = \sum_{j=i+1}^{p} \frac{\lambda_i \lambda_j}{(\lambda_i - \lambda_j)^2}. \tag{A.3}$$

Our numerical experiments show that typically the nearest-neighbour terms dominate the others. Taking this into account, we isolate the first spacing and rewrite $h_i^{\pm}$ as

$$h_i^- = \frac{\lambda_i \lambda_{i-1}}{(\lambda_i - \lambda_{i-1})^2} + \sum_{j=1}^{i-2} \frac{\lambda_i \lambda_j}{(\lambda_i - \lambda_j)^2}, \tag{A.4}$$

$$h_i^+ = \frac{\lambda_i \lambda_{i+1}}{(\lambda_i - \lambda_{i+1})^2} + \sum_{j=i+2}^{p} \frac{\lambda_i \lambda_j}{(\lambda_i - \lambda_j)^2}. \tag{A.5}$$

We study the large p behaviour of (A.4) and (A.5) by separately considering the nearest-neighbour terms and the summations. In particular, we will obtain approximations that rely only on the right and left nearest-neighbour eigenvalue gaps,

$$s_i^{\pm} = |\lambda_i - \lambda_{i\pm1}|. \tag{A.6}$$

We first consider the isolated terms

$$\frac{\lambda_i \lambda_{i\pm1}}{(\lambda_i - \lambda_{i\pm1})^2} = \frac{\lambda_i(\lambda_i \pm s_i^{\pm})}{(s_i^{\pm})^2} \tag{A.7}$$

$$= \frac{\lambda_i^2}{(s_i^{\pm})^2}\left[1 + \mathcal{O}(s_i^{\pm})\right]. \tag{A.8}$$

Using that $s_i^{\pm} \to 0$ as $p \to \infty$ (which is established by Assumption 2.2 and convergences, in expectation, with rate $s_i^{\pm} = \mathcal{O}(1/p)$), we find the asymptotic estimate

$$\frac{\lambda_i \lambda_{i\pm1}}{(\lambda_i - \lambda_{i\pm1})^2} \to \frac{\lambda_i^2}{(s_i^{\pm})^2}. \tag{A.9}$$

We now turn our attention to the summations, which we will approximate using the limiting $p \to \infty$ spectral density $\rho(\lambda)$ of the normalized empirical counting measure of the eigenvalues. More precisely, consider a sequence of size-p symmetric covariance matrices, each having eigenvalues $\{\lambda_j\}$ for $j \in \{1, \ldots, p\}$. We define for each matrix the empirical spectral density

$$\rho_p(\lambda) = p^{-1} \sum_j \delta(\lambda_j), \tag{A.10}$$

where $\delta(\lambda)$ is the Dirac delta function and $\lambda \in \mathbb{R}$. We assume the covariance matrices are drawn from an ensemble such that the sequence $\{\rho_p(\lambda)\}$ weakly converges, implying that

$$\int_{-\infty}^{\infty} \rho_p(\lambda) f(\lambda)\, \mathrm{d}\lambda \to \int_{-\infty}^{\infty} \rho(\lambda) f(\lambda)\, \mathrm{d}\lambda \tag{A.11}$$

as $p \to \infty$ for any continuous and bounded function $f(\lambda)$. We assume that $\rho(\lambda)$ is continuous, is bounded, has compact support (denoted by $\mathrm{supp}(\rho)$) and is differentiable on $\mathrm{supp}(\rho)$. For notational convenience, we assume $\mathrm{supp}(\rho) = (\alpha, \beta)$ for some $\alpha, \beta \in \mathbb{R}$, allowing us to replace the limits of integration in (A.11) by (α, β). However, our analysis can be easily extended to unions of such intervals.

We begin be rewriting the summations in (A.4) and (A.5) as the integration of function

$$f_{\lambda_i}(\lambda) = \tfrac{\lambda_i \lambda}{(\lambda_i - \lambda)^2} \tag{A.12}$$

with probability measure $\rho_p(\lambda)$ given by (A.10),

$$\frac{1}{p} \sum_{j=1}^{i-2} \frac{\lambda_i \lambda_j}{(\lambda_i - \lambda_j)^2} = \int_{\alpha}^{\lambda_{i-1}} \rho_p(\lambda) f_{\lambda_i}(\lambda)\, \mathrm{d}\lambda, \tag{A.13}$$

$$\frac{1}{p} \sum_{j=i+2}^{p} \frac{\lambda_i \lambda_j}{(\lambda_i - \lambda_j)^2} = \int_{\lambda_{i+1}}^{\beta} \rho_p(\lambda) f_{\lambda_i}(\lambda)\, \mathrm{d}\lambda. \tag{A.14}$$

Because $f_{\lambda_i}(\lambda)$ is unbounded at the singularity $\lambda = \lambda_i$, (A.11) does not describe the behaviour of integral $\int_{\alpha}^{\beta} \rho_p(\lambda) f_{\lambda_i}(\lambda)\, \mathrm{d}\lambda$, which we find to diverge with p for any $\lambda_i \in \mathrm{supp}(\rho)$. Fortunately, (A.13) and (A.14) do not require integration across the singularity at $\lambda = \lambda_i$; however, the limits of integration, i.e. λ_{i-1} in (A.13) and λ_{i+1} in (A.14), depend on p (and converge to the singularity at λ_i). Thus, (A.11) is also not directly applicable to (A.13) and (A.14).

To proceed, we restrict our attention to (A.13) since analogous results can be obtained for (A.14). We consider, for the moment, (A.13) with fixed upper limit $\lambda_i - \varepsilon$ for $\varepsilon > 0$ and $\epsilon \approx 0$. Equation (A.11) implies the $p \to \infty$ limit

$$\int_{\alpha}^{\lambda_i - \varepsilon} f_{\lambda_i}(\lambda) \rho_p(\lambda)\, \mathrm{d}\lambda \to \int_{\alpha}^{\lambda_i - \varepsilon} f_{\lambda_i}(\lambda) \rho(\lambda)\, \mathrm{d}\lambda. \tag{A.15}$$

We now study how the right-hand side of (A.15) scales with ε. Using that both $\rho(\lambda)$ and $f_{\lambda_i}(\lambda)$ are differentiable for $\lambda \in \text{supp}(\rho) \setminus \{\lambda_i\}$, we implement integration by parts, treating the numerator and denominator separately, to obtain

$$\int_\alpha^{\lambda_i-\varepsilon} f_{\lambda_i}(\lambda)\rho(\lambda)\,\mathrm{d}\lambda \;=\; \lambda_i \frac{(\lambda_i-\varepsilon)\rho(\lambda_i-\varepsilon)}{\varepsilon} - \lambda_i \int_\alpha^{\lambda_i-\varepsilon} \frac{\rho(\lambda)+\lambda\rho'(\lambda)}{\lambda_i-\lambda}\,\mathrm{d}\lambda. \tag{A.16}$$

The first term in the right-hand side of (A.16) has the $\varepsilon \to 0$ asymptotic estimate

$$\lambda_i \frac{(\lambda_i-\varepsilon)\rho(\lambda_i-\varepsilon)}{\varepsilon} \to \frac{\lambda_i^2\rho(\lambda_i)}{\varepsilon}. \tag{A.17}$$

The second term on the right-hand side of (A.16) is bounded as

$$\begin{aligned}\left|\lambda_i \int_\alpha^{\lambda_i-\varepsilon} \frac{\left[\rho(\lambda)+\lambda\rho'(\lambda)\right]}{\lambda_i-\lambda}\,\mathrm{d}\lambda\right| &\leqslant \lambda_i \left[\sup_{\lambda\in(\alpha,\lambda_i-\varepsilon]} |\rho(\lambda)+\lambda\rho'(\lambda)|\right] \int_\alpha^{\lambda_i-\varepsilon} \frac{1}{|\lambda_i-\lambda|}\,\mathrm{d}\lambda \\ &= \lambda_i \left[\sup_{\lambda\in(\alpha,\lambda_i-\varepsilon]} |\rho(\lambda)+\lambda\rho'(\lambda)|\right] \ln\left(\frac{\lambda_i-\alpha}{\varepsilon}\right).\end{aligned} \tag{A.18}$$

It follows that the second term in the right-hand side of (A.16) has scaling $\mathcal{O}(\ln(1/\varepsilon))$ and is dominated in the limit $\varepsilon \to 0$ by the first term, which is $\mathcal{O}(1/\varepsilon)$. We combine (A.17) and (A.18) to obtain the $\varepsilon \to 0$ asymptotic estimate

$$\int_\alpha^{\lambda_i-\varepsilon} f_{\lambda_i}(\lambda)\rho(\lambda)\,\mathrm{d}\lambda \to \frac{\lambda_i^2\rho(\lambda_i)}{\varepsilon}. \tag{A.19}$$

We finally note that in the case where $\rho'(\lambda)$ is unbounded, it is straightforward to separate the integral on the left-hand side of (A.18) into two domains: one containing all values λ, where $\rho'(\lambda)$ is unbounded and the second domain having the upper limit $\lambda_i-\varepsilon$. The first integral will converge to zero due to (A.11); whereas the second satisfies the bound given by (A.18), implying that the integral term in (A.16) is $\mathcal{O}(\ln(1/\varepsilon))$, provided that $\rho(\lambda)$ is differentiable in a small neighbourhood containing λ_i.

We study the $p \to \infty$ limiting behaviour for the right-hand side of (A.13) by considering the following identity:

$$\int_\alpha^{\lambda_{i-1}} f_{\lambda_i}(\lambda)\rho_p(\lambda)\,\mathrm{d}\lambda = \int_\alpha^{\lambda_i-s_i^-} f_{\lambda_i}(\lambda)\rho(\lambda)\,\mathrm{d}\lambda + \int_\alpha^{\lambda_i-s_i^-} f_{\lambda_i}(\lambda)\left[\rho_p(\lambda)-\rho(\lambda)\right]\mathrm{d}\lambda. \tag{A.20}$$

The first term on the right-hand side grows linearly with p, which is straightforward to show by setting $\varepsilon = s_i^-$ in (A.19) and using that $s_i^- = \mathcal{O}(1/p)$. Turning our attention to the second term on the right-hand side of (A.20), recall that it would converge to zero if the upper limit of integration was fixed. However, $\lambda_i - s_i^-$ limits to λ_i and the $p \to \infty$ behaviour of the second term therefore depends on the rate of weak convergence for $\rho_p(\lambda) \to \rho(\lambda)$. We assume that this term scales sublinearly with p and is therefore dominated by the first term on the right-hand side of (A.20). Under this assumption (and by conducting a similar analysis for (A.5)), we obtain the asymptotic estimates

$$p\int_\alpha^{\lambda_{i-1}} f_{\lambda_i}(\lambda)p\rho_p(\lambda)\,\mathrm{d}\lambda \to \frac{\lambda_i^2 p\rho(\lambda_i)}{s_i^-}, \tag{A.21}$$

$$p\int_{\lambda_{i+1}}^{\beta} f_{\lambda_i}(\lambda)\rho_p(\lambda)\,\mathrm{d}\lambda \to \frac{\lambda_i^2 p\rho(\lambda_i)}{s_i^+}. \tag{A.22}$$

In summary, we combine (A.21), (A.22) and (A.17) to obtain the asymptotic large p approximation,

$$h_i^{\pm} \approx \frac{\lambda_i^2}{\left(s_i^{\pm}\right)^2} + \frac{p\rho(\lambda_i)\lambda_i^2}{s_i^{\pm}}, \tag{A.23}$$

which gives approximation (2.2). We stress that this approximation assumes a sufficiently high rate of weak convergence for the spectral density so that the second term on the right-hand side of (A.20) is sublinear.

B. Derivation of main result 2

In this section, we take a different perspective and consider h_i, defined by (1.2), to be the realization of a random variable that is a function of the corresponding family of random covariance matrices. Using the approximation $\hat{h}$ of h provided by (2.2), we derive an estimate for the probability distribution, $P(h)$ of h. Let us denote by H the random variable for which h_i is a realization.

Our goal is to remove the dependency on the random variables s^+ and s^- in (2.2), so that $\hat{h}$ becomes a function of only λ, which is distributed according to the density $\rho(\lambda)$. The only missing ingredients are the probability distributions of s^+ and s^-. We note that these two random variables are correlated, and thus our line of attack involves using an approximation to the joint probability for the eigenvalue gaps, $J(s^-, s^+)$, and derive an expression for the limiting probability density of approximation $\hat{h}$. In this section, we keep the discussion general and derive an expression that is valid for all $\rho(\lambda)$.

We assume that the joint probability distribution $J(s^-, s^+)$ of the left and right gaps around each eigenvalue λ can be approximated by (2.1), which is reproduced below for ease of presentation:

$$J(s^-, s^+) \approx \frac{3^7\,[p\rho(\lambda)]^5}{32\pi^3}\left[s^+s^-(s^+ + s^-)\right]\exp\left(-\frac{[3p\rho(\lambda)]^2}{4\pi}\left[(s^+)^2 + (s^-)^2 + s^+s^-\right]\right).$$

The expression (2.1) was derived in Herman *et al.* (2007) using 3×3 matrices from the GOE. As suggested by our numerical simulations (see Fig. 2), (2.1) provides a good approximation for the covariance matrices that we study.

To derive the distribution $P(h)$ of H, we first consider the cumulative distribution

$$F(h) \triangleq P(H < h). \tag{B.1}$$

Given an eigenvalue λ, we can find all the pairs of gaps s^- and s^+, such that $\hat{h}$ in (2.2) is less than h. Let

$$\mathscr{S} \triangleq \{(s^-, s^+) : \hat{h}(s^-, s^+) < h\} \tag{B.2}$$

be this set. We then proceed to compute the measure of $\mathscr{S}$ using the joint probability density function defined above,

$$F(h) = \int_{\mathscr{S}} J(s^-, s^+)\,\mathrm{d}s^-\,\mathrm{d}s^+. \tag{B.3}$$

It turns out that we can describe analytically the set $\mathscr{S}$ (see Fig. A1). For a given value of h, $\hat{h}(s^-, s^+) < h$ implies that both $h^+ < h$ and $h^- < h$, where $h^{\pm}$ is given by (A.23) (with the subscript omitted).

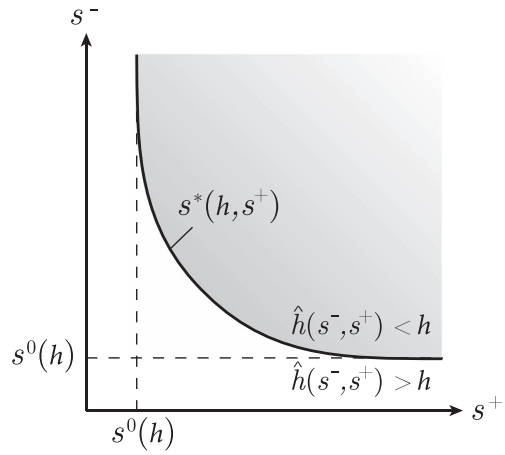

FIG. A1. The cumulative distribution $F(h)$ for the random variable H is shown as the integral of $J(s^-, s^+)$ over region $\mathscr{S}$ given by (B.2) (shaded region). This region corresponds to $s^+ \in (s^0(h), \infty)$ and $s^- \in \big(s^*(h, s^+), \infty\big)$, where $s^0(h)$ is found so that $h^+(s^+) < h$ for $s^+ > s^0(h)$ and $s^*(h, s^+)$ is found so that $\hat{h}(s^-, s^+) < h$ for $s^- > s^*(h, s^+)$.

Therefore the region of integration has the lower bounds $s^- > s^0(h)$ and $s^+ > s^0(h)$, where $s^0(h)$ is given by

$$s^0(h) = \frac{\lambda^2 p\rho(\lambda)}{2h} + \sqrt{\frac{\lambda^2}{h} + \left(\frac{\lambda^2 p\rho(\lambda)}{2h}\right)^2}$$
$$= \frac{\lambda^2 p\rho(\lambda)}{2h}\left(1 + \sqrt{1 + \frac{4h}{[\lambda p\rho(\lambda)]^2}}\ \right), \tag{B.4}$$

which follows directly from solving (A.23) for $s^\pm$ with $h^\pm = h$. We therefore integrate s^+ over the range $(s^0(h), \infty)$. For given values h and s^+, requiring that $\hat{h}(s^-, s^+) > h$ implies that $s^- > s^*(s^+, h)$, where $s^*(h, s^+)$ is found by substituting $h \mapsto \hat{h}(s^-, s^+)$ in (2.2) and solving for the positive root of s^-,

$$s^*(h, s^+) = \lambda^2 p\rho(\lambda)\frac{1 + \sqrt{1 + \frac{4}{[\lambda p\rho(\lambda)]^2}\left(h - \frac{\lambda^2}{(s^+)^2} - \frac{\lambda p\rho(\lambda)}{s^+}\right)}}{2\left(h - \frac{\lambda^2}{(s^+)^2} - \frac{\lambda p\rho(\lambda)}{s^+}\right)}. \tag{B.5}$$

We therefore integrate s^- over the range $(s^*(h, s^+), \infty)$,

$$F(h) = \int_{s^0(h)}^{\infty}\int_{s^*(h,s^+)}^{\infty} J(s^-, s^+)\,\mathrm{d}s^-\,\mathrm{d}s^+. \tag{B.6}$$

To obtain an estimate for the distribution of h, $f_H(h)$, we differentiate (B.6) with respect to h to obtain

$$\begin{aligned} f_H(h) &= \frac{\partial}{\partial h}\int_{s^0(h)}^{\infty}\int_{s^*(h,s^+)}^{\infty} J(s^-,s^+)\,\mathrm{d}s^-\,\mathrm{d}s^+ \\ &= -\frac{\partial s^0}{\partial h}(h)\int_{s^*(h,s^0(h))}^{\infty} J(s^-,s^0(h))\,\mathrm{d}s^- + \int_{s^0(h)}^{\infty}\frac{\partial}{\partial h}\left[\int_{s^*(h,s^+)}^{\infty} J(s^-,s^+)\,\mathrm{d}s^-\right]\mathrm{d}s^+ \\ &= -\int_{s^0(h)}^{\infty} J\left(s^*(h,s^+),s^+\right)\frac{\partial s^*(h,s^+)}{\partial h}\,\mathrm{d}s^+. \end{aligned} \tag{B.7}$$

We note that in the above derivation, the first term in the second line vanishes since $s^*(h,s^+)\to\infty$ in the limit $s^+\to s^0(h)$ and $J(s^-,s^+)$ is bounded.

C. Derivation of main result 3

With h distributed according to $f_H(h)$, given by (B.7), we derive in this section an asymptotic expression for $f_H(h)$ in the limit $h\to\infty$. Examining (2.2), we note that $\hat{h}(s^-,s^+)$ is large when s^- and/or s^+ are small, and thus in the large h limit one can consider only the contributions of the terms proportional to s_-^{-2} and s_+^{-2},

$$h \approx \frac{\lambda^2}{(s^-)^2}+\frac{\lambda^2}{(s^+)^2}. \tag{C.1}$$

In this case, we find

$$s^0(h) = \frac{\lambda}{\sqrt{h}}, \tag{C.2}$$

$$s^*(h,s^+) = \frac{\lambda s^+}{\left[(s^+)^2h-\lambda^2\right]^{1/2}}, \tag{C.3}$$

$$\frac{\partial}{\partial h}\left(s^*(h,s^+)\right) = \frac{-\lambda(s^+)^3}{2\left[(s^+)^2h-\lambda^2\right]^{3/2}} = \frac{-1}{2\lambda^2}[s^*(h,s^+)]^3. \tag{C.4}$$

Substituting these values into (B.7) and dropping the arguments for s^*, i.e. $s^*(h,s^+)\mapsto s^*$, we find

$$\begin{aligned} f_H(h) &= -\int_{\sqrt{\lambda^2/h}}^{\infty}\left(\frac{3^7\,[p\rho(\lambda)]^5}{32\pi^3}s^+s^*(s^*+s^+)\,\mathrm{e}^{-\frac{[3p\rho(\lambda)]^2}{4\pi}\left[(s^*)^2+(s^+)^2+s^*s^+\right]}\right)\left(\frac{-(s^*)^3}{2\lambda^2}\right)\mathrm{d}s^+ \\ &= \frac{3^7\,[p\rho(\lambda)]^5}{32\pi^3}\frac{1}{2\lambda^2}\int_{\sqrt{\lambda^2/h}}^{\infty}\left(\left(s^+(s^*)^5+(s^+)^2(s^*)^4\right)\mathrm{e}^{-\frac{[3p\rho(\lambda)]^2}{4\pi}\left[(s^*)^2+(s^+)^2+s^*s^+\right]}\right)\mathrm{d}s^+. \end{aligned}$$

The change of variables $u=(s^+)^2h-\lambda^2$ transforms this into

$$f_H(h) = \frac{3^7\,[p\rho(\lambda)]^5}{32\pi^3}\frac{\lambda^2}{4}h^{-7/2}I(h), \tag{C.5}$$

where we have defined

$$I(h) = \int_0^\infty \left(1 + \frac{\lambda^2}{u}\right)^{5/2} \left(u^{1/2} + \lambda\right) \mathrm{e}^{-\varphi(u)/h} \, \mathrm{d}u, \tag{C.6}$$

and

$$\varphi(u) = \frac{[3p\rho(\lambda)]^2}{4\pi} \left(u + \lambda^2\right) \left(1 + \frac{\lambda}{\sqrt{u}} + \frac{\lambda^2}{u}\right). \tag{C.7}$$

The distribution $f_H(h)$ in (C.5) depends on h through the power law $h^{-7/2}$ as well as $I(h)$. In Appendix D, we show that (C.6) has the large-h scaling $I(h) = \mathscr{O}\left(\frac{h^{3/2}}{p^3}\right)$. Combining this with (C.5), we find $f_H(h) = \mathscr{O}\left(\frac{p^2}{h^2}\right)$ for large h.

D. Large-h scaling of $I(h)$

We now study how $I(h)$ given by (C.6) scales in the limit of large h. Recall that the limit of large h corresponds to when an eigenvalue λ_i has a nearest-neighbouring eigenvalue that is very close (i.e. $|\lambda_i - \lambda_{i\pm j}| \ll 1$), which results in large values of h_i and subsequently the error of the empirical eigenvector (i.e. large $\|\boldsymbol{u}_i - \widetilde{\boldsymbol{u}}_i\| \approx h_i/n$).

Our strategy for evaluating (C.6) is to split the integral into three regions of integration, which are chosen based on studying the function $\varphi(u)$. Examining (C.7) for limiting values of u, we find that the function $\varphi(u)$ approaches $+\infty$, both as $u \to 0$ and as $u \to \infty$, and has the minimum

$$\min_{u\in[0\infty)} \varphi(u) = \varphi\left(\lambda^2\right) = \frac{27}{2\pi}[\lambda p\rho(\lambda)]^2, \tag{D.1}$$

which occurs at $u = \lambda^2$. For large h, there are two values of u such that $\varphi(u) = h$. We refer to these values as $u_1(h)$ and $u_2(h)$, with $u_1(h) < u_2(h)$. Considering the limits $u \to 0$ and $u \to \infty$, we find the asymptotic approximations

$$u_1(h) \to \frac{[3p\rho(\lambda)]^2}{4\pi} \lambda^4 h^{-1}, \tag{D.2}$$

$$u_2(h) \to \frac{4\pi}{[3p\rho(\lambda)]^2} h. \tag{D.3}$$

We will evaluate (C.6) by dividing the integration into three ranges, $I(h) = I_1(h) + I_2(h) + I_3(h)$, where we define

$$I_1(h) = \int_0^{u_1(h)} \left(1 + \frac{\lambda^2}{u}\right)^{5/2} \left(u^{1/2} + \lambda\right) \mathrm{e}^{-\varphi(u)/h} \, \mathrm{d}u, \tag{D.4}$$

$$I_2(h) = \int_{u_1(h)}^{u_2(h)} \left(1 + \frac{\lambda^2}{u}\right)^{5/2} \left(u^{1/2} + \lambda\right) \mathrm{e}^{-\varphi(u)/h} \, \mathrm{d}u, \tag{D.5}$$

$$I_3(h) = \int_{u_2(h)}^{\infty} \left(1 + \frac{\lambda^2}{u}\right)^{5/2} \left(u^{1/2} + \lambda\right) \mathrm{e}^{-\varphi(u)/h} \, \mathrm{d}u. \tag{D.6}$$

We now study the $h \to \infty$ scaling for integrals $I_1(h)$, $I_2(h)$ and $I_3(h)$. Beginning with (D.4), we first note that for the range $u \in (0, u_1(h)]$

$$\left(u + \lambda^2\right)^{5/2}\left(u^{1/2} + \lambda\right) \leqslant \left(u_1 + \lambda^2\right)^{5/2}\left(u_1^{1/2} + \lambda\right). \tag{D.7}$$

It follows that

$$\left(1 + \frac{\lambda^2}{u}\right)^{5/2}\left(u^{1/2} + \lambda\right) \leqslant Eu^{-5/2}, \tag{D.8}$$

where

$$E(\lambda) = \left(u_1(h) + \lambda^2\right)^{5/2}\left(u_1(h)^{1/2} + \lambda\right). \tag{D.9}$$

Note that $E(\lambda) \approx \lambda^6$ as $h \to \infty$, since $u_1(h) \to 0$. Similarly, since u is positive, one finds

$$\begin{aligned}\varphi(u) &= \frac{[3p\rho(\lambda)]^2}{4\pi}\left(u + \lambda^2\right)\left[1 + \left(\lambda^2/u\right) + \left(\lambda^2/u\right)^{1/2}\right] \\ &\geqslant \frac{[3p\rho(\lambda)]^2}{4\pi}\left(\lambda^2\right)\left(\lambda^2/u\right) \\ &= Fu^{-1},\end{aligned} \tag{D.10}$$

where we have defined

$$F = \frac{[3p\rho(\lambda)]^2}{4\pi}\lambda^4. \tag{D.11}$$

Using these two inequalities, we have

$$I_1(h) \leqslant E(\lambda)\int_0^{u_1(h)} u^{-5/2}\,\mathrm{e}^{-F/(hu)}\,\mathrm{d}u, \tag{D.12}$$

$$= E(\lambda)\left(\frac{h}{F}\right)^{3/2}\int_{F/(hu_1(h))}^{\infty} w^{1/2}\,\mathrm{e}^{-w}\,\mathrm{d}w, \tag{D.13}$$

which uses the change of variables $w = F/(hu(h))$. Using (D.2), the lower limit of integration converges as $F/(hu_1(h)) \to 1$ with $h \to \infty$. The integral in (D.13) therefore limits to a constant, implying that $I_1(h)$ is dominated by a term which scales like $h^{3/2}$.

To estimate $I_3(h)$, note for large h that (D.3) implies $u > \lambda^2$ for any $u > u_2(h)$. It follows that

$$\left(1 + \frac{\lambda^2}{u}\right)^{5/2}\left(u^{1/2} + \lambda\right) \leqslant 2^{5/2}\left(2u^{1/2}\right). \tag{D.14}$$

The integral $I_3(h)$ is thus dominated by

$$I_3(h) \leqslant 8\int_{u_2(h)}^{\infty} u^{1/2}\,\mathrm{e}^{-\varphi(u)/h}\,\mathrm{d}u, \tag{D.15}$$

$$\leqslant 8\int_{u_2(h)}^{\infty} u^{1/2}\exp\left(-\frac{[3p\rho(\lambda)]^2}{4\pi}\frac{u}{h}\right)\mathrm{d}u, \tag{D.16}$$

where the second inequality uses $u > 0$ and $\lambda^2/u > 0$ to bound

$$\varphi(u) = \frac{[3p\rho(\lambda)]^2}{4\pi}\left(u+\lambda^2\right)\left[1+\left(\lambda^2/u\right)+\left(\lambda^2/u\right)^{1/2}\right]$$
$$\geqslant \frac{[3p\rho(\lambda)]^2}{4\pi}u. \tag{D.17}$$

We define the change of variables $w = \frac{[3p\rho(\lambda)]^2}{4\pi}\frac{u}{h}$ to obtain

$$I_3(h) \leqslant 8\left(\frac{[3p\rho(\lambda)]^2}{4\pi}\right)^{3/2}\int_{\frac{[3p\rho(\lambda)]^2}{4\pi}u_2(h)/h}^{\infty} w^{1/2}\,\mathrm{e}^{-w}\,\mathrm{d}w. \tag{D.18}$$

From (D.3), the lower limit of integration converges as $\frac{[3p\rho(\lambda)]^2}{4\pi}u_2(h)/h \to 1$ and the integral in (D.18) converges to a constant as $h \to \infty$. Therefore, $I_3(h)$ is also bounded by a term scaling as $h^{3/2}$.

We will now show that $I_2(h)$ has scaling $\mathscr{O}(h^{3/2})$ (as opposed to the other terms, which we showed are bounded by terms that scale as $h^{3/2}$). Note that because of our definition of u_1 and u_2, and using that $\varphi(u)$ reaches a minimum at $u = \lambda^2 \in (u_1, u_2)$, we find the bounds

$$\varphi(\lambda^2)/h \leqslant \varphi(u)/h \leqslant 1 \tag{D.19}$$

for any $u \in (u_1, u_2)$. Substituting these into (D.6), we bound $I_2(h)$ as

$$Q(h)\,\mathrm{e}^{-1} \leqslant I_2(h) \leqslant Q(h)\,\mathrm{e}^{-\varphi(\lambda^2)/h}, \tag{D.20}$$

where we have defined

$$Q(h) \equiv \int_{u_1(h)}^{u_2(h)}\left(1+\frac{\lambda^2}{u}\right)^{5/2}\left(u^{1/2}+\lambda\right)\mathrm{d}u. \tag{D.21}$$

Using the asymptotic approximations for $u_1(h)$ and $u_2(h)$ given by (D.2) and (D.3), we integrate (D.21) using the software Mathematica (using the 'Series[$Q(h)$, {h, Infinity,1}]' command) to obtain its asymptotic behaviour,

$$Q(h) \approx \frac{2^4\pi^{3/2}}{3^4[p\rho(\lambda)]^3}h^{3/2}. \tag{D.22}$$

Furthermore, we combine $\varphi(\lambda^2)/h \to 0$ with (D.20) to obtain the asymptotic bound

$$Q(h)\,\mathrm{e}^{-1} \leqslant I_2(h) \leqslant Q(h). \tag{D.23}$$

We combine (D.23) with (D.13) and (D.18) to obtain the large-h scaling $I(h) = \mathscr{O}\left(\frac{h^{3/2}}{p^3}\right)$.