Published in final edited form as:

Stat Sin. 2021 April ; 31(2): 571–601. doi:10.5705/ss.202019.0052.

Asymptotics of eigenstructure of sample correlation matrices for high-dimensional spiked models

David Morales-Jimenez[1], **Iain M. Johnstone**[2], **Matthew R. McKay**[3], **Jeha Yang**[2]

[1]ECIT Institute, Queen's University Belfast, UK

[2]Department of Statistics, Stanford University, USA

[3]ECE Department, Hong Kong University of Science and Technology, Hong Kong

Abstract

Sample correlation matrices are widely used, but for high-dimensional data little is known about their spectral properties beyond "null models", which assume the data have independent coordinates. In the class of spiked models, we apply random matrix theory to derive asymptotic first-order and distributional results for both leading eigenvalues and eigenvectors of sample correlation matrices, assuming a high-dimensional regime in which the ratio p/n, of number of variables p to sample size n, converges to a positive constant. While the first-order spectral properties of sample correlation matrices match those of sample covariance matrices, their asymptotic distributions can differ significantly. Indeed, the correlation-based fluctuations of both sample eigenvalues and eigenvectors are often remarkably smaller than those of their sample covariance counterparts.

Keywords

Sample correlation; eigenstructure; spiked models

1. Introduction

Estimating a correlation matrix is a fundamental statistical task. It is widely applied in areas such as viral sequence analysis and vaccine design in biology (Dahirel et al., 2011, Quadeer et al., 2014, 2018), large portfolio design in finance (Plerou et al., 2002), signal detection in radio astronomy (Leshem and van der Veen, 2001), and collaborative filtering (Liu et al., 2014, Ruan et al., 2016), among many others. In classical statistical settings, with a limited number of variables p and a large sample size n, the sample correlation matrix performs well and its statistical properties are well understood; see, for example, Girshick (1939), Konishi (1979), Fang and Krishnaiah (1982), Schott (1991), Kollo and Neudecker (1993), and Boik (2003). Modern applications, however, often exhibit high dimensionality, with large p and,

David Morales-Jimenez: ECIT Institute, Queen's University Belfast, UK, d.morales@qub.ac.uk.

Supplementary Material

The online Supplementary Material provides proofs for the following: (*i*) the Gaussian particularizations of our main results (Corollaries 1 and 2); (*ii*) the instrumental tightness properties in Lemma 3; and (*iii*) the asymptotic properties of normalized bilinear forms in Lemma 1 and Proposition 1; see Sections S1, S2, and S3, respectively.

in many cases, limited n. In such cases, sample correlation matrices become inaccurate owing to an aggregation of statistical noise across the matrix coordinates that is visible in the eigen-spectrum (El Karoui, 2009). This is particularly important in principal component analysis (PCA), which often involves projecting data onto the leading eigenvectors of the sample correlation matrix or, equivalently, onto those of the sample covariance matrix after standardizing the data.

Despite the extensive use of sample correlation matrices, relatively little is known about theoretical properties of their eigen-spectra in high dimensions. In contrast, sample covariance matrices have been studied extensively, and a rich body of literature now exists (e.g., Yao et al. (2015)). Their asymptotic properties have typically been described in high-dimensional settings in which the number of samples and variables both grow large, often though not always at the same rate, based on the theory of random matrices. Specific first- and second-order results for the eigenvalues and eigenvectors of sample covariance matrices are reviewed in Bai and Silverstein (2009), Couillet and Debbah (2011), and Yao et al. (2015).

For the spectra of high-dimensional sample *correlation* matrices, current theoretical results focus on the simplest "null model" scenario, in which the data are assumed to be independent. In this null model, correlation matrices share many of the same asymptotic properties as covariance matrices from independent and identically distributed (i.i.d.) data, with zero mean and unit variance. Thus, the empirical eigenvalue distribution converges to the Marchenko–Pastur distribution, almost surely (Jiang, 2004b), and the largest and smallest eigenvalues converge to the edges of this distribution (Jiang, 2004b, Xiao and Zhou, 2010). Moreover, the rescaled largest and smallest eigenvalues asymptotically follow the Tracy–Widom law (Bao et al., 2012, Pillai and Yin, 2012). Central limit theorems (CLTs) for linear spectral statistics have also been derived (Gao et al., 2017). A separate line of work studies the maximum absolute off-diagonal entry of sample correlation matrices, referred to as "coherence" (Jiang, 2004a, Cai and Jiang, 2011, 2012), which has been proposed as a statistic for conducting independence tests; see also Cochran et al. (1995), Mestre and Vallet (2017), and the references therein. Hero and Rajaratnam (2011, 2012) use a related statistic to identify variables exhibiting strong correlations, an approach referred to as "correlation screening."

For non-trivial correlation models, however, asymptotic results for the spectra of sample correlation matrices are quite scarce. Notably, El Karoui (2009) shows that, for a fairly general class of covariance models with bounded spectral norm, to first order, the eigenvalues of sample correlation matrices asymptotically coincide with those of sample covariance matrices with unit-variance data, generalizing earlier results of Jiang (2004b) and Xiao and Zhou (2010). Under similar covariance assumptions, recent work also presents CLTs for linear spectral statistics of sample correlation matrices (Mestre and Vallet, 2017), extending the work of Gao et al. (2017). First order behavior again coincides with that of sample covariances. However, the asymptotic fluctuations are quite different for sample correlation matrices.

This study considers a particular class of correlation matrix models, the so-called "spiked models," in which a few large or small eigenvalues of the population covariance (or correlation) matrix are assumed to be well separated from the rest (Johnstone, 2001). Spiked covariance models are relevant in applications in which the primary covariance information lies in a relatively small number of eigenmodes. Such applications include collaborative signal detection in cognitive radio systems (Bianchi et al., 2009), fault detection in sensor networks (Couillet and Hachem, 2013), adaptive beamforming in array processing (Hachem et al., 2013, Vallet et al., 2015, Yang et al., 2018), and protein contact prediction in biology (Cocco et al., 2011, 2013). The spectral properties of spiked covariance models have been well studied, with precise analytical results established for the asymptotic first-order and distributional properties of both eigenvalues and eigenvectors; see, for example, Baik et al. (2005), Baik and Silverstein (2006), Paul (2007), Bai and Yao (2008), Benaych-Georges and Nadakuditi (2011), Couillet and Hachem (2013), Bloemendal et al. (2016). For reviews, see also Couillet and Debbah (2011, Chapter 9) and Yao et al. (2015, Chapter 11).

Less is known about the spectrum of sample correlation matrices under spiked models. Although the asymptotic first-order behavior is expected to coincide with that of the sample covariance, as a consequence of El Karoui (2009), a simple simulation reveals striking differences in the fluctuations of both sample eigenvalues and eigenvectors; see Figure 1.

Here, we present theoretical results to describe these observed phenomena. We obtain asymptotic first-order and distribution results for the eigenvalues and eigenvectors of sample correlation matrices under a spiked model. Paul (2007) proved theorems for sample covariance matrices in the special case of Gaussian data. In essence, we present analogs of these theorems for sample correlation matrices, and extend them to non-Gaussian data. To first order, the eigenvalues and eigenvectors coincide asymptotically with those of sample covariance matrices; however, their fluctuations can be very different. Indeed, for both the largest sample correlation eigenvalues (Theorem 1) and the projections of the corresponding eigenvectors (Theorem 2), the asymptotic variances admit a decomposition into three terms. The first term is just the asymptotic variance for sample covariance matrices generated from Gaussian data; the second adds corrections due to non-Gaussianity, and the third captures further corrections due to data normalization imposed by the sample correlation matrix. (This last amounts to normalizing the entries of the sample covariance matrix using the sample variances). Consistent with the example shown in Figure 1(a), in the CLT for the leading sample eigenvalues, the sample correlation eigenvalues often show lower fluctuations—despite the variance normalization—than those of the sample covariance eigenvalues. As seen in Figure 1(b), the (normalized) eigenvector projections are typically asymptotically correlated, even for Gaussian data, unlike the sample covariance setting of Paul (2007, Theorem 5).

Technical contributions

We build on and extend a set of random matrix tools for studying spiked covariance models. The companion manuscript (Johnstone and Yang, 2018) [JY], gives an exposition and parallel treatment for sample covariance matrices. Important adaptations are needed here to account for the data normalization imposed by sample correlation matrices. Among key

technical contributions of our work, basic to our main theorems, are asymptotic first-order and distributional properties for bilinear forms and matrix quadratic forms with normalized entries, Section 4. A novel regularization-based proof strategy is used to establish the inconsistency of eigenvector projections in the case of "subcritical" spiked eigenvalues, Theorem 3.

Model M

Let $x \in \mathbb{R}^{m+p}$ be a random vector with finite $(4+\delta)$th moment for some $\delta > 0$. Consider the partition

$$x = \begin{bmatrix} \xi \\ \eta \end{bmatrix}.$$

Assume that $\xi \in \mathbb{R}^m$ has mean zero and covariance Σ, and is independent of $\eta \in \mathbb{R}^p$, which has i.i.d components η_i with mean zero and unit variance. Let $\Sigma_D = \mathrm{diag}\left(\sigma_1^2, \ldots, \sigma_m^2\right)$ be the diagonal matrix containing the variances of ξ_i, and let $\Gamma = \Sigma_D^{-1/2} \Sigma \Sigma_D^{-1/2}$ be the correlation matrix of ξ with eigen-decomposition $\Gamma = PLP^T$, where $P = [p_1, \ldots, p_m]$ is the eigenvector matrix, and $L = \mathrm{diag}(\ell_1, \ldots, \ell_m)$ contains the spike correlation eigenvalues $\ell_1 \geq \ldots \geq \ell_m > 0$.

The correlation matrix of x is therefore $\Gamma_x = \mathrm{blkdiag}(\Gamma, I)$, with eigenvalues $\ell_1, \ldots, \ell_m, 1, \ldots, 1$, and corresponding eigenvectors $\mathfrak{p}_1, \ldots, \mathfrak{p}_m, e_{m+1}, \ldots, e_{m+p}$, where $\mathfrak{p}_i = \left[p_i^T 0_p^T\right]^T$ and e_j is the jth canonical vector (i.e., a vector of all zeros, except for a one in the jth coordinate).

Consider a sequence of i.i.d. copies of x, the first n of which fill the columns of the $(m+p) \times n$ data matrix $X = (x_{ij})$. We assume m is fixed, whereas p and n increase with

$$\gamma_n = p/n \to \gamma > 0 \quad \text{as} \quad p, n \to \infty.$$

Notation

Let $S = n^{-1} XX^T$ be the sample covariance matrix, and $S_D = \mathrm{diag}\left(\hat{\sigma}_1^2, \ldots, \hat{\sigma}_{m+p}^2\right)$ be the diagonal matrix containing the sample variances. Let $R = S_D^{-1/2} S S_D^{-1/2}$ be the sample correlation matrix, with corresponding νth sample eigenvalue and eigenvector satisfying

$$R\hat{\mathfrak{p}}_\nu = \hat{\ell}_\nu \hat{\mathfrak{p}}_\nu,$$

where, for later use, we partition $\hat{\mathfrak{p}}_\nu = \left[\hat{p}_\nu^T, \hat{v}_\nu^T\right]^T$. Here $\hat{p}_\nu$ is the subvector of $\hat{\mathfrak{p}}_\nu$ restricted to the first m coordinates.

For $\ell > 1 + \sqrt{\gamma}$, define

$$\rho(\ell, \gamma) = \ell + \gamma \frac{\ell}{\ell - 1}, \quad \dot{\rho}(\ell, \gamma) = \frac{\partial \rho(\ell, \gamma)}{\partial \ell} = 1 - \frac{\gamma}{(\ell - 1)^2}.$$

For an index ν, for which $\ell_\nu > 1+\sqrt{\gamma}$ is a simple eigenvalue, set

$$\rho_\nu = \rho(\ell_\nu, \gamma), \quad \rho_{\nu n} = \rho(\ell_\nu, \gamma_n), \quad \dot{\rho}_\nu = \dot{\rho}(\ell_\nu, \gamma), \quad \dot{\rho}_{\nu n} = \dot{\rho}(\ell_\nu, \gamma_n). \tag{1.1}$$

We refer to eigenvalues satisfying $\ell_\nu > 1+\sqrt{\gamma}$ as "supercritical," and those satisfying $\ell_\nu \le 1+\sqrt{\gamma}$ as "subcritical," with the quantity $1+\sqrt{\gamma}$ referred to as the "phase transition."

To describe and interpret the variance terms in the limiting distributions to follow, we need some definitions. Let $\bar{\xi}_i = \xi_i/\sigma_i$ and $\kappa_{ij} = \mathbb{E}\bar{\xi}_i\bar{\xi}_j$ denote the scaled components of ξ and their covariances; of course $\kappa_{ii} = 1$. The corresponding scaled fourth-order cumulants are

$$\kappa_{iji'j'} = \mathbb{E}\left[\bar{\xi}_i\bar{\xi}_j\bar{\xi}_{i'}\bar{\xi}_{j'}\right] - \kappa_{ij}\kappa_{i'j'} - \kappa_{ij'}\kappa_{ji'} - \kappa_{ii'}\kappa_{jj'}. \tag{1.2}$$

When ξ is Gaussian, $\kappa_{iji'j'} \equiv 0$.

The effect of variance scaling in the correlation matrix is described using additional quadratic functions of $(\bar{\xi}_i)$, defined by

$$\chi_{ij} = \bar{\xi}_i\bar{\xi}_j, \quad \psi_{ij} = \kappa_{ij}\left(\bar{\xi}_i^2 + \bar{\xi}_j^2\right)/2 \tag{1.3}$$

$$\check{\kappa}_{iji'j'} = \mathrm{Cov}\left(\psi_{ij}, \psi_{i'j'}\right) - \mathrm{Cov}\left(\psi_{ij}, \chi_{i'j'}\right) - \mathrm{Cov}\left(\chi_{ij}, \psi_{i'j'}\right). \tag{1.4}$$

Tensor notation

For convenience, it is useful to consider $\kappa_{iji'j'}$ and $\check{\kappa}_{iji'j'}$ as entries of four-dimensional tensor arrays κ and $\check{\kappa}$, respectively, and to define an additional array $\mathscr{P}^{\mu\mu'\nu\nu'}$ with entries $p_{\mu,i}p_{\mu',j}p_{\nu,i'}p_{\nu',j'}$. In addition, define $\mathscr{P}^\nu$ as $\mathscr{P}^{\nu\nu\nu\nu}$. Finally, for a second array A of the same dimensions,

$$\left[\mathscr{P}^\nu, A\right] = \sum_{i,j,i',j'} P^\nu_{iji'j'}A_{iji'j'}.$$

2. Main results

Our first main result, proved in Section 5, gives the asymptotic properties of the largest (spike) eigenvalues of the sample correlation matrix:

Theorem 1

Assume Model M, *and that* $\ell_\nu > 1+\sqrt{\gamma}$ *is a simple eigenvalue. As* $p/n \to \gamma > 0$,

$$\begin{aligned} &(i) \qquad \widehat{\ell}_\nu \xrightarrow{\text{a.s.}} \rho_\nu, \\ &(ii) \qquad \sqrt{n}(\widehat{\ell}_\nu - \rho_{\nu n}) \xrightarrow{\mathscr{D}} N\left(0, \tilde{\sigma}_\nu^2\right), \end{aligned} \tag{2.5}$$

where

$$\tilde{\sigma}_\nu^2 = 2\dot{\rho}_\nu \ell_\nu^2 + \dot{\rho}_\nu^2\left[\mathscr{P}^\nu, \kappa\right] + \dot{\rho}_\nu^2\left[\mathscr{P}^\nu, \check{\kappa}\right]. \tag{2.6}$$

Centering at $\rho_{\nu n}$ rather than at ρ_ν is important. If, for example, $\gamma_n = \gamma + an^{-1/2}$, then

$$\sqrt{n}(\hat{\ell}_\nu - \rho_\nu) \xrightarrow{\mathscr{D}} N(a\ell_\nu(\ell_\nu - 1)^{-1}, \tilde{\sigma}_\nu^2),$$

and we see a limiting shift. Furthermore, it may also be beneficial to consider $\tilde{\sigma}_{\nu n}^2$ instead of $\tilde{\sigma}_\nu^2$, obtained by replacing $\dot{\rho}_\nu$ with $\dot{\rho}_{\nu n}$ in (2.6), such that

$$\sqrt{n}(\hat{\ell}_\nu - \rho_{\nu n})/\tilde{\sigma}_{\nu n} \xrightarrow{\mathscr{D}} N(0,1).$$

The asymptotic first-order limit in (*i*), which follows as an easy consequence of El Karoui (2009), coincides with that of the νth largest eigenvalue of a sample covariance matrix computed from data with population covariance Γ (Paul, 2007). This implies that, when constructing R, normalizing by the sample variances has no effect on the leading eigenvalues, at least to first order.

However, key differences are seen when looking at the asymptotic distribution, given in (*ii*), and in the variance formula (2.6) in particular. This can be readily interpreted. The first term corresponds to the variance in the Gaussian-covariance case of Paul (2007), again for samples with covariance Γ. The second provides a correction of that result for non-Gaussian data, see the companion article [JY]. The third term describes the contribution specific to sample correlation matrices, representing the effect of normalizing the data by the sample variances. This term is often negative, and is evaluated explicitly for Gaussian data in Corollary 1 below, proved in the Supplementary Material, S1.1.

Corollary 1

For ξ Gaussian, the asymptotic variance in Theorem 1 simplifies to

$$\tilde{\sigma}_\nu^2 = 2\ell_\nu^2\dot{\rho}_\nu\left[1 - \dot{\rho}_\nu\left(2\ell_\nu \mathrm{tr}P_{D,\nu}^4 - \mathrm{tr}\left(P_{D,\nu}\Gamma P_{D,\nu}\right)^2\right)\right],$$

where $P_{D,\nu} = \mathrm{diag}(p_{\nu,1}, \ldots, p_{\nu,m})$.

Thus, computing the sample correlation results in the asymptotic variance being scaled by $1 - \dot{\rho}_\nu\Delta_\nu$, relative to the sample covariance, where

$$\Delta_\nu = 2\ell_\nu \mathrm{tr}P_{D,\nu}^4 - \mathrm{tr}\left(P_{D,\nu}\Gamma P_{D,\nu}\right)^2 = 2\ell_\nu \sum_i p_{\nu,i}^4 - \sum_{i,j}\left(p_{\nu,i}\kappa_{ij}p_{\nu,j}\right)^2$$

is often positive, implying that spiked eigenvalues of the sample correlation often exhibit a smaller variance than those of the sample covariance. Indeed, such variance reduction occurs iff

$$\sum_{i,j}\left(p_{\nu,i}\kappa_{ij}p_{\nu,j}\right)^2 < 2\ell_\nu\sum_i p_{\nu,i}^4 = \sum_{i,j} p_{\nu,i}\kappa_{ij}p_{\nu,j}(p_{\nu,i}^2 + p_{\nu,j}^2), \tag{2.7}$$

with the last identity following from the fact that $\ell_\nu p_{\nu,i} = \sum_j \kappa_{ij}p_{\nu,j}$. Condition (2.7), and variance reduction, holds in the following cases:

- **i.** both Γ and p_ν have nonnegative entries, or
- **ii.** $2\ell_\nu\sum_i p_{\nu,i}^4 > 1$, or
- **iii.** $2\ell_\nu > \ell_1^2$.

In case (i), the inequalities $0 \le p_{\nu,i}\kappa_{ij}p_{\nu,j} \le 2p_{\nu,i}p_{\nu,j} \le p_{\nu,i}^2 + p_{\nu,j}^2$ yield (2.7). Note that if Γ has nonnegative entries, then the Perron–Frobenius theorem establishes the existence of an eigenvector with nonnegative components for ℓ_1; furthermore, if Γ has positive entries, by the same theorem, ℓ_1 is simple and associated with an eigenvector with positive components. Case (ii) follows from $\sum_{i,j}\left(p_{\nu,i}\kappa_{ij}p_{\nu,j}\right)^2 \le \sum_{i,j}\left(p_{\nu,i}p_{\nu,j}\right)^2 = 1$, and holds if $\ell_\nu > m/2$, because $\sum_i p_{\nu,i}^4 \ge 1/m$. Case (iii) follows from the inequalities $2p_{\nu,i}^2p_{\nu,j}^2 \le p_{\nu,i}^4 + p_{\nu,j}^4$ and $\sum_j \kappa_{ij}^2 = \left(\Gamma^2\right)_{ii} \le \|\Gamma^2\| = \ell_1^2$. Note that this is rather special, in that it has nothing to do with eigenvectors, and a necessary condition for it to hold is $\ell_1 \le 2$.

Condition (2.7) can fail, however. For example, for even m and $r \in (0, 1)$, consider

$$\Gamma = \begin{pmatrix} 1 & -r \\ -r & 1 \end{pmatrix} \otimes 1_{m/2}1_{m/2}^T,$$

where $1_{m/2}$ is the $(m/2)$-dimensional vector of all ones, which corresponds to two negatively correlated groups of identical random vectors. This has simple supercritical eigenvalues $\ell_1 = (1 + r)m/2$ and $\ell_2 = (1 - r)m/2$ when $m > 2(1 + \sqrt{\gamma})/(1 - r)$, with $p_{\nu,i}^2 = m^{-1}$ for $\nu = 1, 2$. One finds that $\Delta_2 = (1 - 2r - r^2)/2 < 0$ for $r > \sqrt{2} - 1$, although $\Delta_1 > 0$ because $\ell_1 > m/2$, which implies case (ii).

We turn now to the eigenvectors. Again, fix an index ν for which $\ell_\nu > 1 + \sqrt{\gamma}$ is a simple eigenvalue of Γ, with corresponding eigenvector $\mathfrak{p}_\nu = \left[p_\nu^T\ 0_p^T\right]^T$. Recall that $\widehat{\mathfrak{p}}_\nu = \left[\hat{p}_\nu^T\ \hat{v}_\nu^T\right]^T$ is the νth sample eigenvector of R, and let $a_\nu = \hat{p}_\nu/\|\hat{p}_\nu\|$ be the corresponding normalized subvector of $\widehat{\mathfrak{p}}_\nu$, restricted to the first m coordinates. The next result establishes a limit for the eigenvector projection $\langle\widehat{\mathfrak{p}}_\nu, \mathfrak{p}_\nu\rangle$, and a CLT for the normalized cross-projections $P^Ta_\nu = \left[p_1^Ta_\nu, \ldots, p_m^Ta_\nu\right]^T$; see Sections 6.1 and 6.2.

Theorem 2

Assume Model M, *and that* $\ell_\nu > 1 + \sqrt{\gamma}$ *is a simple eigenvalue. Then, as* $p/n \to \gamma > 0$,

$$(i) \quad \langle \hat{\mathfrak{p}}_\nu, \mathfrak{p}_\nu \rangle^2 \xrightarrow{\text{a.s.}} \dot{\rho}_\nu \ell_\nu / \rho_\nu,$$
$$(ii) \quad \sqrt{n}\left(P^T a_\nu - e_\nu\right) \xrightarrow{\mathcal{D}} N(0, \Sigma_\nu),$$

where $\Sigma_\nu = \mathcal{D}_\nu \widetilde{\Sigma}_\nu \mathcal{D}_\nu$ *with*

$$\mathcal{D}_\nu = \sum_{k \neq \nu}^{m} (\ell_\nu - \ell_k)^{-1} e_k e_k^T \tag{2.8}$$

$$\widetilde{\Sigma}_{\nu, kl} = \dot{\rho}_\nu^{-1} \ell_k \ell_\nu \delta_{k,l} + \left[\mathcal{P}^{k\nu l\nu}, \kappa\right] + \left[\mathcal{P}^{k\nu l\nu}, \check{\kappa}\right], \tag{2.9}$$

where $\delta_{k,l} = 1$ *if* $k = l$, *and zero otherwise*.

The CLT result in (*ii*) can be rephrased in terms of the entries of a_ν, for which we readily obtain $\sqrt{n}(a_\nu - p_\nu) \xrightarrow{\mathcal{D}} N\left(0, P\Sigma_\nu P^T\right)$; note that Σ_ν has zeros in the νth row and the νth column.

As for the eigenvalues, Theorem 2 shows that the spiked eigenvectors of sample correlation matrices exhibit the same first-order behavior as those of the sample covariance (Paul, 2007). The difference again lies in the asymptotic fluctuations, captured by the covariance matrix Σ_ν. Note that this is decomposed as a product of $\mathcal{D}_\nu$ – a diagonal matrix—and the matrix $\widetilde{\Sigma}_\nu$, which involves the three terms in (2.9). These terms have similar interpretations as those discussed previously in (2.6). That is, the first term captures the asymptotic fluctuations for a Gaussian-covariance model (Paul, 2007), the second term captures the effect of non-Gaussianity in the covariance case [JY], and the third term captures information specific to the correlation case, representing fluctuations due to sample variance normalization. Note that only the first term is diagonal in general, suggesting that the eigenvector projections may be asymptotically correlated, as seen earlier in Figure 1(b), right panel. This holds also for Gaussian data, evaluated explicitly in Corollary 2 below; see Supplementary Material, S1.2, for the proof. We note an interesting contrast with the eigenvector projections for covariance matrices (Paul, 2007), described only by the leading term in (2.9).

Corollary 2

For ξ *Gaussian, the asymptotic covariance in Theorem 2 reduces to* $\Sigma_\nu = \mathcal{D}_\nu \widetilde{\Sigma}_\nu \mathcal{D}_\nu$,

$$\widetilde{\Sigma}_\nu = \frac{\ell_\nu}{\dot{\rho}_\nu} L + (\ell_\nu I + L)\left(\frac{1}{2}\mathcal{X} - \ell_\nu \mathcal{Y}\right)(\ell_\nu I + L) + \ell_\nu\left(\ell_\nu^2 \mathcal{Y} - L\mathcal{Y}L\right),$$

where $\mathcal{X} = P^T P_{D,\nu}(\Gamma \circ \Gamma) P_{D,\nu} P$, $\mathcal{Y} = P^T P_{D,\nu}^2 P$, *and* $\circ$ *denotes the Hadamard product*.

Thus, for Gaussian data, the entries of the asymptotic covariance matrix are given by (for $k, l \neq \nu$)

$$\Sigma_{\nu,kl} = (\ell_\nu - \ell_k)^{-1}(\ell_\nu - \ell_l)^{-1}\left[\frac{\ell_\nu}{\rho_\nu}\ell_k\delta_{k,l} + (\ell_\nu + \ell_k)(\ell_\nu + \ell_l)\frac{\mathcal{Z}_{kl}}{2} - \ell_\nu(\ell_\nu(\ell_k + \ell_l) + 2\ell_k\ell_l)\mathcal{Y}_{kl}\right].$$

Consider now the subcritical case in which ν is such that $1 < \ell_\nu \leq 1 + \sqrt{\gamma}$. Let $\mathfrak{p}_\nu$ denote the corresponding population eigenvector, and let $\widehat{\ell}_\nu$ and $\widehat{\mathfrak{p}}_\nu$ denote the corresponding sample eigenvalue and eigenvector, respectively. With proofs deferred to Sections 5.1 and 6.3, we have the following result:

Theorem 3

Assume Model M, *and that* $1 < \ell_\nu \leq 1 + \sqrt{\gamma}$ *is a simple eigenvalue. Then, as* $p/n \to \gamma > 0$,

$$(i) \quad \widehat{\ell}_\nu \xrightarrow{\text{a.s.}} (1 + \sqrt{\gamma})^2,$$
$$(ii) \quad \langle \widehat{\mathfrak{p}}_\nu, \mathfrak{p}_\nu \rangle^2 \xrightarrow{\text{a.s.}} 0.$$

Once again, the asymptotic first-order limits of the sample eigenvalue and its associated eigenvector are the same as those obtained for the sample covariance (Paul, 2007).

Recall that our high-dimensional results assume an asymptotic regime where $p/n \to \gamma > 0$, as opposed to the classical regime where p is fixed and $n \to \infty$. The case of fixed p corresponds to $\gamma = 0$ and the spectral properties of the sample correlation matrix are well understood; see, for example, Girshick (1939), Konishi (1979), Fang and Krishnaiah (1982), Schott (1991), Kollo and Neudecker (1993), and Boik (2003). When $\gamma = 0$, the function $\rho(\ell, \gamma)$ reduces to the identity. Indeed, for fixed p, there is no high-dimensional component η in Model M, and hence no biasing effect on $\rho(\ell, \gamma)$ that occurs when $\gamma > 0$. In particular, for fixed p there is no counterpart to our Theorem 3.

To summarize, in comparison to the high-dimensional ($p/n \to \gamma > 0$) sample covariance setting, our results for the spiked eigenvalues and eigenvectors of sample correlation matrices confirm that the first-order asymptotic behavior is indeed equivalent to that of sample covariance matrices, in agreement with previous results and observations (El Karoui, 2009, Mestre and Vallet, 2017). While the eigenvalue limits in Theorem 1 and Theorem 3 follow as a straightforward consequence of El Karoui (2009), the eigenvector results of Theorem 2-(*i*) and Theorem 3-(*ii*) do not. In contrast to the first-order equivalences, important differences arise in the fluctuations of both the eigenvalues and eigenvectors, as shown by the asymptotic distributions of Theorem 1-(*ii*) and Theorem 2-(*ii*).

We illustrate these differences with a simple example having covariance $\Gamma = (1 - r)I_m + r1_m1_m^T$, where $r \in [0, 1]$; that is, a model with unit variances and constant correlation r across all components. Moreover, ξ is assumed to be Gaussian for simplicity. In this setting, $L = \text{diag}(\ell_1, 1 - r, \ldots, 1 - r)$, where $\ell_1 = 1 + r(m - 1)$ is supercritical iff $r > \sqrt{\gamma}/(m - 1)$. Consider the largest sample eigenvalue $\widehat{\ell}_1$ in such a supercritical case. From Corollary 1, the asymptotic variances for the sample covariance and the sample correlation can be computed, yielding

$$\sigma_1^2 = 2\ell_1^2\dot{\rho}_1, \quad \tilde{\sigma}_1^2 = \sigma_1^2(1-\dot{\rho}_1\Delta),$$

respectively, with $\Delta = 2\ell_1 \mathrm{tr} P_D^4 - \mathrm{tr}(P_D\Gamma P_D)^2$, and where

$$P_D \triangleq P_{D,1} = m^{-1/2} I_m, \quad \dot{\rho}_1 = 1 - \frac{\gamma}{r^2(m-1)^2}.$$

Figure 2(a) plots these asymptotic variances versus r for various (γ,m). Indeed, the variance (fluctuation) for the sample correlation is consistently smaller than for the sample covariance. The difference is striking, becoming extremely large as $r \nearrow 1$. Similar trends are observed for various choices of m and γ, being more pronounced for higher m, while not much affected by varying γ. This may be understood from the fact that, after writing $\Delta = r(2 - r) + (1 - r)^2 m^{-1} = 1 - (1 - r)^2(1 - m^{-1})$,

$$\frac{\tilde{\sigma}_1^2}{\sigma_1^2} = 1 - \dot{\rho}_1\Delta \to \begin{cases} \dfrac{\gamma}{(m-1)^2} & \text{as } r \to 1, m \text{ fixed} \\ (1-r)^2 & \text{as } m \to \infty, r \text{fixed}. \end{cases}$$

Turn now to the fluctuations of the leading sample eigenvector, in the same setting as above. Note that, in Corollary 2, for this particular case, one can deduce from $P^T\Gamma P = L$ that

$$\mathscr{X} = m^{-1}\left(1 - r^2\right) I_m + r^2 e_1 e_1^T, \quad \mathscr{Y} = m^{-1} I_m.$$

Also from Corollary 2, the asymptotic variances for the normalized sample-to-population eigenvector projection $p_2^T a_1$, in the sample covariance and sample correlation cases, are computed as

$$\Sigma_{1,22}^{\mathrm{cov}} = \frac{\ell_1\ell_2}{(rm)^2\dot{\rho}_1}, \quad \Sigma_{1,22} = \Sigma_{1,22}^{\mathrm{cov}} - \frac{\zeta}{(rm)^2}\frac{\ell_1\ell_2(\ell_1+\ell_2)}{m},$$

respectively, where $\zeta = 1 - r + \frac{1}{2}(1+r)\left(1 + \frac{1-r}{rm}\right)^{-1}$, and we recall that $\ell_1 = 1 - r + rm$ and $\ell_2 = 1 - r$. These variances are numerically evaluated in Figure 2(b) for the same parameter choices as before and, again, as functions of r. Note, however, that for better visual appreciation, the range of r has been restricted to supercritical values sufficiently above the critical point $\sqrt{\gamma}/(m-1)$, because the variance explodes at that point. The comparative evaluation again shows smaller variances for the sample correlation. The variance reduction here is less visible in the graphs, because both $\Sigma_{1,22}$ and $\Sigma_{1,22}^{\mathrm{cov}}$ vanish as $r \to 1$. The ratio, however, behaves quite similarly to the variance ratio $\tilde{\sigma}_1^2/\sigma_1^2$:

$$\frac{\Sigma_{1,22}}{\Sigma_{1,22}^{\mathrm{cov}}} = 1 - \zeta_{\dot{\rho}1}\frac{(\ell_1+\ell_2)}{m} \to \begin{cases} \dfrac{\gamma}{(m-1)^2} & \text{as } r \to 1, m \text{ fixed} \\ (1-r)(1-r/2) & \text{as } m \to \infty, r \text{ fixed.} \end{cases}$$

We end the discussion of our main results with a few remarks about possible extensions. Our results assume that $\ell_\nu > 1$ is a simple eigenvalue, but extensions for small spikes with $\ell_\nu < 1$ and for spikes with multiplicities should be possible. Analogous results for eigenvalues have been obtained for sample covariance matrices for $\ell_\nu < 1$, including multiplicities greater than one (e.g., see Bai and Yao (2008)), giving reason to expect corresponding results for correlation matrices. Extensions of our results for eigenvalues and eigenvectors of sample correlation matrices for simple $\ell_\nu < 1$ should be fairly straightforward, though the cases $\gamma < 1$, $\gamma = 1$, and $\gamma > 1$ would need separate treatment. Extensions for spikes with multiplicities are also possible, but in this case the eigenvectors are not well defined and one would need to consider subspace projections, requiring non-trivial modifications of our technical arguments.

The remainder of the paper proceeds as follows. First, in Section 3, we introduce key quantities and identities used in the derivations. Section 4 presents necessary asymptotic properties for bilinear forms and matrix quadratic forms with normalized entries, with the corresponding proofs relegated to the Supplementary Material, Section S3. These properties provide a foundation for describing the asymptotic convergence and distribution of eigenvalues and eigenvectors of sample correlation matrices, derived in Sections 5 and 6 respectively.

As already noted, a parallel treatment for the simpler case of covariance matrices is given in a supplementary manuscript [JY]. This aims at a unified exposition of known spectral properties of spiked covariance matrices as a benchmark for the current work, along with additional citations to the literature.

3. Preliminaries

We begin with a block representation and some associated reductions for the sample correlation matrix R. These are well known in the covariance matrix setting. As with the partition of x in Model M, consider

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}, \quad X_1 \in \mathbb{R}^{m \times n}, \quad X_2 \in \mathbb{R}^{p \times n}.$$

Write $S_D = \mathrm{blkdiag}(S_{D1}, S_{D2})$, with S_{D1} containing the sample variances corresponding to ξ, and S_{D2} containing those corresponding to η. Define the "normalized" data matrices $\overline{X}_1 = S_{D1}^{-1/2} X_1$ and $\overline{X}_2 = S_{D2}^{-1/2} X_2$, such that

$$R = n^{-1}\begin{bmatrix} \overline{X}_1\overline{X}_1^T & \overline{X}_1\overline{X}_2^T \\ \overline{X}_2\overline{X}_1^T & \overline{X}_2\overline{X}_2^T \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix}; \quad \widehat{\mathfrak{p}}_\nu = \begin{bmatrix} \hat{p}_\nu \\ \hat{\upsilon}_\nu \end{bmatrix}.$$

This partitioning of the eigenvector equation $R\widehat{\mathfrak{p}}_v = \widehat{\ell}_v\widehat{\mathfrak{p}}_v$, along with $\widehat{\mathfrak{p}}_v = \left[\hat{p}_v^T, \hat{v}_v^T\right]^T$, yields

$$R_{11}\hat{p}_v + R_{12}\hat{v}_v = \widehat{\ell}_v\hat{p}_v$$

$$R_{21}\hat{p}_v + R_{22}\hat{v}_v = \widehat{\ell}_v\hat{v}_v .$$

From the second equation, $\hat{v}_v = \left(\widehat{\ell}_v I_p - R_{22}\right)^{-1} R_{21}\hat{p}_v$. Substituting this into the first equation yields

$$K\left(\widehat{\ell}_v\right)\hat{p}_v = \widehat{\ell}_v\hat{p}_v, \quad \text{with} \quad K(t) = R_{11} + R_{12}\left(tI_p - R_{22}\right)^{-1}R_{21} .$$

Thus, $\widehat{\ell}_v$ is an eigenvalue of $K\left(\widehat{\ell}_v\right)$, with associated eigenvector $\hat{p}_v$; this is central to our derivations. Note that $K\left(\widehat{\ell}_v\right)$ is well defined if $\widehat{\ell}_v$ is well separated from the eigenvalues of R_{22}; Section 5.1 shows that this occurs with probability one for all large n when ℓ_v is supercritical. Furthermore, the normalization condition, $\hat{p}_v^T\hat{p}_v + \hat{v}_v^T\hat{v}_v = 1$ yields

$$\hat{p}_v^T(I_m + Q_v)\hat{p}_v = 1, \quad Q_v = R_{12}\left(\widehat{\ell}_v I_p - R_{22}\right)^{-2}R_{21} .$$

Phrased in terms of the signal-space normalized eigenvector $a_v = \hat{p}_v/\|\hat{p}_v\|$, we have

$$K\left(\widehat{\ell}_v\right)a_v = \widehat{\ell}_v a_v, \quad a_v^T(I_m + Q_v)a_v = \|\hat{p}_v\|^{-2} . \tag{3.10}$$

Note also that the sample-to-population inner product can be rewritten as

$$\left\langle \widehat{\mathfrak{p}}_v, \mathfrak{p}_v \right\rangle = \langle \hat{p}_v, p_v \rangle = \|\hat{p}_v\|\langle a_v, p_v \rangle . \tag{3.11}$$

In the derivation of our CLT results, we use an eigenvector perturbation formula with quadratic error bound given in [JY, Lemma 13], itself a modification of the arguments in Paul (2007). This yields the key expansion

$$a_v - p_v = -\mathscr{R}_{vn}D_v p_v + r_v, \tag{3.12}$$

where

$$\mathscr{R}_{vn} = \frac{\ell_v}{\rho_{vn}}\sum_{k\neq v}^{m}(\ell_k - \ell_v)^{-1}p_k p_k^T, \quad D_v = K\left(\widehat{\ell}_v\right) - (\rho_{vn}/\ell_v)\Gamma, \quad \|r_v\| = O\left(\|D_v\|^2\right).$$

The derivations of our eigenvalue and eigenvector results, presented in Sections 5 and 6 respectively, take (3.10), (3.11) and (3.12) as points of departure, and rely on asymptotic

properties of the key objects $K(\widehat{\ell}_\nu)$ and Q_ν. In particular, $K(t)$ can be expressed as the random matrix quadratic form

$$K(t) = n^{-1}\overline{X}_1 B_n(t)\overline{X}_1^T, \tag{3.13}$$

where, using the Woodbury identity,

$$\begin{aligned} B_n(t) &= I_n + n^{-1}\overline{X}_2^T\left(tI_p - R_{22}\right)^{-1}\overline{X}_2 \\ &= t\left(tI_n - n^{-1}\overline{X}_2^T\overline{X}_2\right)^{-1}. \end{aligned}$$

Thus, our key objects are random quadratic forms involving the normalized data matrices $\overline{X}_1$ and $\overline{X}_2$. The asymptotic properties of these forms are foundational to our results, and are presented next.

4. Quadratic forms with normalized entries

In this section, we establish the first-order (deterministic) convergence and a CLT for matrix quadratic forms of the type $n^{-1}\overline{X}_1 B_n \overline{X}_1^T$, where B_n is a matrix with bounded spectral norm. While being essential to our purposes, some of the technical results may be of independent interest; thus, we first present the general results, and then apply these in the context of Model M.

4.1 First-order convergence

To establish the first-order convergence, we first require some results on bilinear forms involving correlated random vectors of unit length. A main technical result (see Supplementary Material, S3.1) is the following:

Lemma 1—Let B be an n × n nonrandom symmetric matrix, and let $x, y \in \mathbb{R}^n$ be random vectors of i.i.d. entries with mean zero, variance one, $\mathbb{E}|x_i|^l$, $\mathbb{E}|y_i|^l \le \nu_l$, and $\mathbb{E}[x_i y_i] = \rho$. Let $\bar{x} = \sqrt{n}x/\|x\|$ and $\bar{y} = \sqrt{n}y/\|y\|$. Then, for any s ≥ 1,

$$\mathbb{E}\left|n^{-1}\bar{x}^T B\bar{y} - \rho n^{-1}\mathrm{tr}B\right|^s \le \mathscr{C}_s\left[n^{-s}\left(\nu_{2s}\mathrm{tr}B^s + \left(\nu_4\mathrm{tr}B^2\right)^{s/2}\right) + \|B\|^s\left(n^{-s/2}\nu_4^{s/2} + n^{-s+1}\nu_{2s}\right)\right],$$

where $\mathscr{C}_s$ is a constant depending only on s.

This is a generalization of Gao et al. (2017, Lemma 5), which established a corresponding bound for normalized quadratic forms. Lemma 1 leads to the following first-order convergence result:

Corollary 3—*Let x,y be random vectors of i.i.d. entries with mean zero, variance one,* $\mathbb{E}|x_i|^{4+\delta}$, $\mathbb{E}|y_i|^{4+\delta} < \infty$ *for some* $\delta > 0$, *and* $\mathbb{E}[x_i y_i] = \rho$. *Define* $\bar{x} = \sqrt{n}x/\|x\|$ *and* $\bar{y} = \sqrt{n}y/\|y\|$, *and let* B_n *be a sequence of* $n \times n$ *symmetric matrices, with* $\|B_n\|$ *bounded. Then,*

$$n^{-1}\bar{x}^T B_n \bar{y} - n^{-1}\rho \mathrm{tr} B_n \xrightarrow{\text{a.s.}} 0.$$

Proof. Because the $(4 + \delta)$th moment and $\|B_n\|$ are bounded, from Lemma 1,

$$\mathbb{E}\left|n^{-1}\bar{x}^T B_n \bar{y} - n^{-1}\rho \mathrm{tr} B_n\right|^{2+\delta/2} \leq O\left(n^{-(1+\delta/4)}\right).$$

The convergence then follows from Markov's inequality and the Borel–Cantelli lemma. □

We now apply this to our Model M with *random* matrices $B_n(\bar{X}_2)$, independent of $\bar{X}_1$:

Lemma 2—Assume Model M, and suppose that $B_n = B_n(\bar{X}_2)$ is a sequence of random symmetric matrices, for which $\|B_n\|$ is $O_{a.s.}(1)$. Then,

$$n^{-1}\bar{X}_1 B_n(\bar{X}_2)\bar{X}_1^T - n^{-1}\mathrm{tr} B_n(\bar{X}_2)\Gamma \xrightarrow{\text{a.s.}} 0.$$

Proof. This follows from Fubini's theorem. Specifically, one may use the arguments in the proof of [JY, Lemma 5], applying Corollary 3, and noting that $\bar{X}_1$ is independent of $B_n(\bar{X}_2)$. □

4.2 Central Limit Theorem

To establish our main matrix quadratic-form CLT result, we first derive a CLT for scalar bilinear forms involving normalized random vectors. To this end, we must introduce some further notation. Consider zero-mean random vectors $(x, y) \in \mathbb{R}^M \times \mathbb{R}^M$, with

$$\mathrm{Cov}\begin{pmatrix} x \\ y \end{pmatrix} = C = \begin{pmatrix} C^{xx} & C^{xy} \\ C^{yx} & C^{yy} \end{pmatrix},$$

where $C^{xy}_{ll'} = \mathbb{E}[x_l y_{l'}]$. Assume $C^{xx}_{ll} = C^{yy}_{ll} = 1$; that is, all components of the x and y vectors have unit variance and $\rho_l = C^{xy}_{ll} = \mathbb{E}[x_l y_l]$. We first introduce notation for some quadratic functions of x_l, y_l. Let $z, w \in \mathbb{R}^M$, with

$$z_l = x_l y_l, \quad w_l = \rho_l\left(x_l^2 + y_l^2\right)/2, \quad C^{zz} = \mathrm{Cov}(z), \quad C^{wz} = \mathrm{Cov}(z, w), etc.$$

Let $X = (x_{li})_{M\times n}$ and $Y = (y_{li})_{M\times n}$ be data matrices based on n i.i.d. observations of (x, y), and define the "normalized" data matrices $\bar{X} = \hat{\Sigma}_x^{-1/2} X$ and $\bar{Y} = \hat{\Sigma}_y^{-1/2} Y$, where $\hat{\Sigma}_x = \mathrm{diag}\left(\hat{\sigma}^2_{x1}, \ldots, \hat{\sigma}^2_{xM}\right)$, $\hat{\Sigma}_y = \mathrm{diag}\left(\hat{\sigma}^2_{y1}, \ldots, \hat{\sigma}^2_{yM}\right)$, and $\hat{\sigma}^2_{xl} = n^{-1}\sum_{i=1}^n x^2_{li}, \hat{\sigma}^2_{yl} = n^{-1}\sum_{i=1}^n y^2_{li}$. Then, we use the following notation for the rows $\bar{x}_l^T$ and $\bar{y}_l^T$ of the normalized data matrices

$$\bar{X} = (\bar{x}_{li})_{M \times n} = \begin{bmatrix} \bar{x}_{1\cdot}^T \\ \vdots \\ \bar{x}_{M\cdot}^T \end{bmatrix}, \quad \bar{Y} = (\bar{y}_{li})_{M \times n} = \begin{bmatrix} \bar{y}_1^T \\ \vdots \\ \bar{y}_M^T \end{bmatrix}.$$

With this setup, we have the following result, proved in the Supplementary Material, S3.2:

Proposition 1—Let $B_n = (b_{n,ij})$ be random symmetric n × n matrices, independent of X, Y, such that for some finite β, $\|B_n\| \le \beta$ for all n, and

$$n^{-1} \sum_{i=1}^{n} b_{n,ii}^2 \xrightarrow{p} \omega, \quad n^{-1} \mathrm{tr} B_n^2 \xrightarrow{p} \theta, \quad \left(n^{-1} \mathrm{tr} B_n\right)^2 \xrightarrow{p} \phi,$$

all finite. In addition, define $Z_n \in \mathbb{R}^M$, with components

$$Z_{n,l} = n^{-1/2}\left[\bar{x}_{l\cdot}^T B_n \bar{y}_{l\cdot} - \rho_l \mathrm{tr} B_n\right].$$

Then, $Z_n \xrightarrow{\mathcal{D}} N_M(0, D)$, with

$$D = (\theta - \omega)J + \omega K_1 + \phi K_2 = \theta J + \omega K + \phi K_2, \tag{4.14}$$

where $K = K_1 - J$ and J, K_1, K_2 are matrices defined by

$$\begin{aligned} J &= C^{xy} \circ C^{yx} + C^{xx} \circ C^{yy} \\ K_1 &= C^{zz} \\ K_2 &= C^{ww} - C^{wz} - C^{zw}. \end{aligned} \tag{4.15}$$

The entries of K are fourth-order cumulants of x and y:

$$K_{ll'} = \mathbb{E}(x_l y_l x_{l'} y_{l'}) - \mathbb{E}(x_l y_l)\mathbb{E}(x_{l'} y_{l'}) - \mathbb{E}(x_l y_{l'})\mathbb{E}(y_l x_{l'}) - \mathbb{E}(x_l x_{l'})\mathbb{E}(y_l y_{l'}). \tag{4.16}$$

Hence, K vanishes if x, y are Gaussian.

The corresponding result with unnormalized vectors is established in [JY Theorem 10]. The terms $\theta J + \omega K$ appear in that case, and the additional term ϕK_2 reflects the normalization in $\bar{x}_{l\cdot}$ and $\bar{y}_{l\cdot}$. As in [JY], the proof is based on the martingale CLT, rather than the moment method used in Bai and Yao (2008), which stated a similar result for quadratic forms involving unnormalized random vectors.

While potentially of independent interest, Proposition 1 is important for our purposes through its application to Model M.

Proposition 2—Assume Model M, and consider B_n as in Proposition 1. Then,

$$W_n = n^{-1/2}\left[\overline{X}_1 B_n \overline{X}_1^T - (\mathrm{tr}B_n)\Gamma\right] \xrightarrow{\mathscr{D}} W,$$

where W is a symmetric m × m Gaussian matrix with entries W_{ij}, mean zero, and covariances given by

$$\mathrm{Cov}\left[W_{ij}, W_{i'j'}\right] = \theta\left(\kappa_{ij'}\kappa_{ji'} + \kappa_{ii'}\kappa_{jj'}\right) + \omega\kappa_{iji'j'} + \phi\check{\kappa}_{iji'j'}, \tag{4.17}$$

for $i \le j$ and $i' \le j'$.

Proof. The result follows from Proposition 1 by turning the matrix quadratic form $\overline{X}_1 B_n \overline{X}_1^T$ into a vector of bilinear forms; see, for example, [JY, Proposition 6] and Bai and Yao (2008, Proposition 3.1). Specifically, use an index l for the $M = m(m+1)/2$ pairs (i, j), with $1 \le i \le j \le m$. Build the random vectors (x, y) for Proposition 1 as follows: if $l = (i, j)$, then set $x_l = \xi_i/\sigma_i$ and $y_l = \xi_j/\sigma_j$. In the resulting covariance matrix C for (x, y), if also $l' = (i', j')$,

$$C_{ll'}^{xy} = \mathbb{E}\left[\xi_i \xi_{j'}\right]/\left(\sigma_i \sigma_{j'}\right) = \kappa_{ij'}, \quad C_{ll'}^{yx} = \kappa_{ji'}, \quad C_{ll'}^{xx} = \kappa_{ii'}, \quad C_{ll'}^{yy} = \kappa_{jj'}$$

and, in particular, $\rho_l = C_{ll}^{xy} = \kappa_{ij}$ and $\rho_{l'} = \kappa_{i'j'}$, whereas $C_{ll}^{xx} = C_{ll}^{yy} = 1$. Component $W_{n,ij}$ corresponds to component Z_l in Proposition 1. Thus, we conclude that $W_n \xrightarrow{\mathscr{D}} W$, where W is a Gaussian matrix with zero mean and $\mathrm{Cov}\left(W_{ij}, W_{i',j'}\right) = D_{ll'}$, given by Proposition 1. It remains to interpret the quantities in (4.14) in terms of Model M. Substituting $x_l = \bar{\xi}_i$ and $y_l = \bar{\xi}_j$ into (4.16) and chasing definitions, we obtain $J_{ll'} = \kappa_{ij'}\kappa_{ji'} + \kappa_{ii'}\kappa_{jj'}$ and $K_{ll'} = \kappa_{iji'j'}$. Observing that $z_l = x_l y_l = \chi_{ij}$ and $w_l = \rho_l\left(x_l^2 + y_l^2\right)/2 = \psi_{ij}$, we similarly find that $K_{2,ll'} = \check{\kappa}_{iji'j'}$. □

5. Proofs of the eigenvalue results

In this section, we derive the main eigenvalue results, presented in Theorem 1 and Theorem 3-(*i*).

5.1 Preliminaries

Convergence properties of the eigenvalues of R_{22}—It is well known that the empirical spectral density (ESD) of S_{22} converges weakly a.s. to the Marchenko–Pastur (MP) law F_γ, and that the extreme non-trivial eigenvalues converge to the edges of the support of F_γ. For the sample correlation case, Jiang (2004b) shows that the same is true for R_{22}. That is, the empirical distribution of the eigenvalues $\mu_1 \ge \ldots \ge \mu_p$ of the "noise" correlation matrix $R_{22} = n^{-1}\overline{X}_2\overline{X}_2^T$ converges weakly a.s. to the MP law F_γ, supported on $\left[a_\gamma, b_\gamma\right] = \left[(1-\sqrt{\gamma})^2, (1+\sqrt{\gamma})^2\right]$, if $\gamma \le 1$, and on $\{0\} \cup [a_\gamma, b_\gamma]$ otherwise. Furthermore, the ESD of the $n \times n$ companion matrix $C_n = n^{-1}\overline{X}_2^T\overline{X}_2$, denoted by F_n, converges weakly a.s. to

the "companion MP law" $\mathrm{F}_\gamma = (1-\gamma)\mathbf{1}_{[0,\infty)} + \gamma F_\gamma$, where $\mathbf{1}_A$ denotes the indicator function on set A.

In addition, Jiang (2004b) shows that

$$\mu_1 \xrightarrow{\text{a.s.}} b_\gamma \quad \text{and} \quad \mu_{p\wedge n} \xrightarrow{\text{a.s.}} a_\gamma . \tag{5.18}$$

Based on these results, if $f_n \to f$ uniformly as continuous functions on the closure $\mathcal{I}$ of a bounded neighborhood of the support of F_γ, then:

$$\int f_n(x)\mathrm{F}_n(dx) \xrightarrow{\text{a.s.}} \int f(x)\mathrm{F}_\gamma(dx) . \tag{5.19}$$

If supp(F_n) is not contained in $\mathcal{I}$, then the left side integral may not be defined. However, such an event occurs for at most finitely many n with probability one.

Almost sure limit of $\widehat{\ell}_\nu$—The statements in Theorem 1-(i) and Theorem 3-(i) follow easily from known results. Specifically, denote the νth eigenvalue of the sample covariance S by $\hat{\lambda}_\nu$. The almost sure limits

$$\hat{\lambda}_\nu \xrightarrow{\text{a.s.}} \begin{cases} \rho_\nu, & \ell_\nu > 1+\sqrt{\gamma} \\ (1+\sqrt{\gamma})^2, & 1 < \ell_\nu \le 1+\sqrt{\gamma} \end{cases} \tag{5.20}$$

were established in Baik and Silverstein (2006). From the proof of El Karoui (2009, Lemma 1),

$$\max_{i=1,\ldots,m} \left|\hat{\lambda}_i - \widehat{\ell}_i\right| \xrightarrow{\text{a.s.}} 0.$$

Therefore, the same almost sure limits as (5.20) hold for $\widehat{\ell}_\nu$.

High-probability events $J_{n\epsilon}$, $J_{n\epsilon 1}$—When necessary, we may confine attention to the event $J_{n\epsilon} = \left\{\widehat{\ell}_\nu > \min(\rho_\nu, \rho_{\nu n}) - \epsilon, \mu_1 \le b_\gamma + \epsilon\right\}$ or $J_{n\epsilon 1} = \left\{\mu_1 \le b_\gamma + \epsilon\right\}$, with $\epsilon > 0$ chosen such that $\rho_\nu - b_\gamma \ge 3\epsilon$, because from (2.5) (proven above) and (5.18), these events occur with probability one for all large n.

Asymptotic expansion of $K\left(\widehat{\ell}_\nu\right)$—We establish an asymptotic stochastic expansion for the quadratic form $K\left(\widehat{\ell}_\nu\right)$. Specifically, using the decomposition

$$K\left(\widehat{\ell}_\nu\right) = K(\rho_{\nu n}) + \left[K\left(\widehat{\ell}_\nu\right) - K(\rho_{\nu n})\right], \tag{5.21}$$

we show that

$$K(\rho_{\nu n}) \xrightarrow{\text{a.s.}} -\rho_\nu \mathrm{m}(\rho_\nu;\gamma)\Gamma = (\rho_\nu/\ell_\nu)\Gamma \tag{5.22}$$

and

$$K\left(\widehat{\ell}_v\right) - K(\rho_{vn}) = -\left(\widehat{\ell}_v - \rho_{vn}\right)[c(\rho_v)\Gamma + o_{\mathrm{a.s.}}(1)], \tag{5.23}$$

where, for $t \notin \mathrm{supp}(\mathrm{F}_\gamma)$,

$$\mathrm{m}(t;\gamma) = \int (x-t)^{-1}\mathrm{F}_\gamma(dx), \quad c(t) = \int x(t-x)^{-2}\mathrm{F}_\gamma(dx).$$

Here, m is the Stieltjes transform of the companion distribution F_γ.

In establishing (5.22), start by taking sufficiently large n such that $|\rho_{vn} - \rho_v| \le \epsilon$, with ϵ defined as above. For such n, on $J_{n\epsilon 1}$, we have

$$\|B_n(\rho_{vn})\| \le \frac{\rho_v + \epsilon}{\epsilon}.$$

Because $J_{n\epsilon 1}$ holds with probability one for all large n, $\|B_n(\rho_{vn})\| = O_{\mathrm{a.s.}}(1)$ and, therefore, it follows from Lemma 2 that

$$K(\rho_{vn}) - n^{-1}\mathrm{tr}B_n(\rho_{vn})\Gamma \xrightarrow{\mathrm{a.s.}} 0.$$

In addition, (5.19) yields

$$n^{-1}\mathrm{tr}B_n(\rho_{vn}) = \int \rho_{vn}(\rho_{vn} - x)^{-1}F_n(dx) \xrightarrow{\mathrm{a.s.}} \int \rho_v(\rho_v - x)^{-1}F_\gamma(dx) = -\rho_v m(\rho_v;\gamma).$$

Explicit evaluation gives $\mathrm{m}(\rho_v; \gamma) = -1/\ell_v$, [JY, Appendix A], and (5.22) follows.

To establish (5.23), we start by recalling that $C_n = n^{-1}\overline{X}_2^T\overline{X}_2$, and introduce the resolvent notation $Z(t) = (tI_n - C_n)^{-1}$, such that $B_n(t) = tZ(t)$ and $K(t) = n^{-1}\overline{X}_1 t Z(t)\overline{X}_1^T$. From the resolvent identity, that is, $A^{-1} - B^{-1} = A^{-1}(B - A)B^{-1}$ for square invertible A and B, and noting that $tZ(t) = C_nZ(t) + I$ from the Woodbury identity, we have, for $t_1, t_2 > b_\gamma$,

$$t_1Z(t_1) - t_2Z(t_2) = -(t_1 - t_2)C_nZ(t_1)Z(t_2)$$

and, therefore,

$$K\left(\widehat{\ell}_v\right) - K(\rho_{vn}) = -\left(\widehat{\ell}_v - \rho_{vn}\right)n^{-1}\overline{X}_1C_nZ\left(\widehat{\ell}_v\right)Z(\rho_{vn})\overline{X}_1^T.$$

Moreover, again by the resolvent identity, $Z\left(\widehat{\ell}_v\right) = Z(\rho_{vn}) - \left(\widehat{\ell}_v - \rho_{vn}\right)Z\left(\widehat{\ell}_v\right)Z(\rho_{vn})$, which yields

$$K\left(\widehat{\ell}_\nu\right) - K(\rho_{\nu n}) = -\left(\widehat{\ell}_\nu - \rho_{\nu n}\right) n^{-1} \overline{X}_1 B_{n1}(\rho_{\nu n}, \rho_{\nu n}) \overline{X}_1^T + \left(\widehat{\ell}_\nu - \rho_{\nu n}\right)^2 n^{-1} \overline{X}_1 B_{n2}\left(\widehat{\ell}_\nu, \rho_{\nu n}\right) \overline{X}_1^T, \tag{5.24}$$

with $B_{nl}(t_1, t_2)$ defined as

$$B_{nr}(t_1, t_2) = C_n Z(t_1) Z^r(t_2). \tag{5.25}$$

We now characterize the first-order behavior of the two matrix quadratic forms in (5.24). For the first, we simply mirror the arguments of the proof of (5.22) to obtain

$$n^{-1} \overline{X}_1 B_{n1}(\rho_{\nu n}, \rho_{\nu n}) \overline{X}_1^T \xrightarrow{\text{a.s.}} c(\rho_\nu) \Gamma.$$

For the second, we again apply similar reasoning, operating on the event $J_{n\epsilon}$. Specifically, it is easy to establish that on $J_{n\epsilon}$, and for n sufficiently large that $|\rho_{\nu n} - \rho_\nu| \le \epsilon$, $\left\| B_{n2}\left(\widehat{\ell}_\nu, \rho_{\nu n}\right) \right\|$ is bounded. Hence, $\left\| B_{n2}\left(\widehat{\ell}_\nu, \rho_{\nu n}\right) \right\| = O_{\text{a.s.}}(1)$, and it follows from Lemma 2 and (5.19) that

$$n^{-1} \overline{X}_1 B_{n2}\left(\widehat{\ell}_\nu, \rho_{\nu n}\right) \overline{X}_1^T = O_{\text{a.s.}}(1).$$

The expansion in (5.23) is obtained by combining the latter two equations with (5.24).

CLT of $K(\rho_{\nu n})$—We now specialize Proposition 2 for the matrix quadratic form $K(\rho_{\nu n})$.

Proposition 3—Assume Model M, and define $\rho_{\nu n}$ by (1.1) and $K(\rho_{\nu n})$ by (3.13). Then,

$$W_n(\rho_{\nu n}) = \sqrt{n}\left[K(\rho_{\nu n}) - n^{-1} \text{tr} B_n(\rho_{\nu n}) \Gamma\right] \xrightarrow{\mathcal{D}} W^\nu,$$

which is a symmetric Gaussian random matrix with entries W^ν_{ij}, mean zero, and covariances given by

$$\text{Cov}\left[W^\nu_{ij}, W^\nu_{i'j'}\right] = \frac{\rho_\nu^2}{\ell_\nu^2 \dot{\rho}_\nu}\left(\kappa_{ij'}\kappa_{ji'} + \kappa_{ii'}\kappa_{jj'}\right) + \frac{\rho_\nu^2}{\ell_\nu^2}\left(\kappa_{iji'j'} + \breve{\kappa}_{iji'j'}\right), \tag{5.26}$$

where ρ_ν and $\dot{\rho}_\nu$ are defined in (1.1), and the terms in parentheses are defined in (1.2) and (1.4).

Proof. Recall that $J_{n\epsilon 1} = \{\mu_1 \le b_\gamma + \epsilon\}$, and consider sufficiently large n such that $\rho_{\nu n} > \rho_\nu - \epsilon$. Then, we may apply Proposition 2 with $B_n = B_n(\rho_{\nu n}) 1_{J_{n\epsilon 1}}$, which is independent of $\overline{X}_1$, and for which $\|B_n\|$ is bounded. Specifically, the result follows by applying Proposition 2 to $W_n(\rho_{\nu n}) 1_{J_{n\epsilon 1}}$, along with the fact that $1_{J_{n\epsilon 1}} \xrightarrow{\text{a.s.}} 1$, and particularizing ω, θ, and ϕ in (4.17). These quantities, denoted respectively by ω_ν, θ_ν, and ϕ_ν, can be computed as in [JY, Appendix A], yielding

$$\omega_\nu = \phi_\nu = \frac{(\ell_\nu - 1 + \gamma)^2}{(\ell_\nu - 1)^2} = \frac{\rho_\nu^2}{\ell_\nu^2}, \quad \theta_\nu = \frac{(\ell_\nu - 1 + \gamma)^2}{(\ell_\nu - 1)^2 - \gamma} = \frac{\omega_\nu}{\dot{\rho}_\nu}.$$

Tightness properties—Lastly, we establish some tightness properties essential to the derivation of our second-order results.

We first establish a refinement of (5.22). Define $K_0(\rho;\ \gamma) := -\rho \mathrm{m}(\rho;\ \gamma)\Gamma$, such that (5.22) is rewritten as $K(\rho_{\nu n}) \xrightarrow{\text{a.s.}} K_0(\rho_\nu; \gamma)$. Set $g_\rho(x) = \rho(\rho - x)^{-1}$, and write

$$\mathrm{tr} B_n(\rho) = \sum_{i=1}^{n} \rho(\rho - \mu_i)^{-1} = \sum_{i=1}^{n} g_\rho(\mu_i).$$

In addition, introducing

$$G_n(g) := \sum_{i=1}^{n} g(\mu_i) - n \int g(x) F_{\gamma_n}(dx),$$

we have

$$\begin{aligned} K(\rho) - K_0(\rho; \gamma_n) &= K(\rho) - n^{-1} \mathrm{tr} B_n(\rho)\Gamma \\ &+ \rho n^{-1} \left[\sum_{i=1}^{n} (\rho - \mu_i)^{-1} - n \int (\rho - x)^{-1} F_{\gamma_n}(dx) \right] \Gamma \\ &= n^{-1/2} W_n(\rho) + n^{-1} G_n(g_\rho)\Gamma. \end{aligned} \tag{5.27}$$

Lemma 3—Assume that Model M holds, and that $\ell_\nu > 1 + \sqrt{\gamma}$ is simple. For some $b > \rho_1$, let I denote the interval $[b_\gamma + 3\epsilon, b]$. Then,

$$\left\{ G_n(g_\rho), \rho \in I \right\} \textit{ is uniformly tight}, \tag{5.28}$$

$$\left\{ n^{1/2} [K(\rho) - K_0(\rho; \gamma_n)], \rho \in I \right\} \textit{ is uniformly tight}, \tag{5.29}$$

$$\widehat{\ell}_\nu - \rho_{\nu n} = O_p\left(n^{-1/2}\right), \tag{5.30}$$

$$a_\nu - p_\nu = O_p\left(n^{-1/2}\right). \tag{5.31}$$

Proof. The proofs of (5.28)–(5.30) appear in the Supplementary Material, S2. We show (5.31) using the expansion $a_\nu - p_\nu = -\mathscr{R}_{\nu n} D_\nu p_\nu + r_\nu$, given in (3.12), from which we recall $\|r_\nu\| = O(\|D_\nu\|^2)$ and note that $\|\mathscr{R}_{\nu n}\| \leq C$ and $D_\nu = K\left(\widehat{\ell}_\nu\right) - K_0(\rho_{\nu n}; \gamma_n)$. We then have $a_\nu - p_\nu = O_p(\|D_\nu\| + \|D_\nu\|^2)$. Furthermore, from

$$\|D_\nu\| \le \left\|K\left(\widehat{\ell}_\nu\right) - K(\rho_{\nu n})\right\| + \|K(\rho_{\nu n}) - K_0(\rho_{\nu n};\gamma_n)\|,$$

the first term is $O_p(n^{-1/2})$ by (5.23) and (5.30), as is the second term by (5.29). Hence,

$$\|D_\nu\| = O_p\left(n^{-1/2}\right), \tag{5.32}$$

and the proof is completed. □

5.2 Eigenvalue uctuations (Theorem 1-(*ii*))

The proof of Theorem 1-(*ii*) relies on the key expansion

$$\sqrt{n}\left(\widehat{\ell}_\nu - \rho_{\nu n}\right)\left[1 + c(\rho_\nu)\ell_\nu + o_p(1)\right] = p_\nu^T W_n(\rho_{\nu n}) p_\nu + o_p(1), \tag{5.33}$$

which is obtained by combining the vector equations $K\left(\widehat{\ell}_\nu\right)a_\nu = \widehat{\ell}_\nu a_\nu$ and $K_0(\rho_{\nu n};\gamma_n)p_\nu = \rho_{\nu n}p_\nu$ with expansions (5.24) for $K\left(\widehat{\ell}_\nu\right) - K(\rho_{\nu n})$ and (5.27) for $K(\rho_{\nu n}) - K_0(\rho_{\nu n};\gamma_n)$. Specifically, we first use $\left[K\left(\widehat{\ell}_\nu\right) - \widehat{\ell}_\nu I_m\right]a_\nu = 0$ to obtain

$$p_\nu^T\left[K\left(\widehat{\ell}_\nu\right) - \widehat{\ell}_\nu I_m\right]p_\nu = (a_\nu - p_\nu)^T\left[K\left(\widehat{\ell}_\nu\right) - \widehat{\ell}_\nu I_m\right](a_\nu - p_\nu) = O_p\left(n^{-1}\right), \tag{5.34}$$

because $\left\|K\left(\widehat{\ell}_\nu\right) - \widehat{\ell}_\nu I_m\right\| = O_p(1)$ from (5.21)–(5.23) and (2.5), and $a_\nu - p_\nu = O_p(n^{-1/2})$ from Lemma 3. In addition, because $[K_0(\rho_{\nu n};\gamma_n) - \rho_{\nu n}I_m]p_\nu = 0$, it follows that

$$\begin{aligned} p_\nu^T\left[K\left(\widehat{\ell}_\nu\right) - \widehat{\ell}_\nu I_m\right]p_\nu &= p_\nu^T\left[K\left(\widehat{\ell}_\nu\right) - K_0(\rho_{\nu n};\gamma_n) - \left(\widehat{\ell}_\nu - \rho_{\nu n}\right)I_m\right]p_\nu \\ &= p_\nu^T\left[K\left(\widehat{\ell}_\nu\right) - K(\rho_{\nu n}) - \left(\widehat{\ell}_\nu - \rho_{\nu n}\right)I_m\right]p_\nu \\ &+ p_\nu^T[K(\rho_{\nu n}) - K_0(\rho_{\nu n};\gamma_n)]p_\nu \\ &= -\left(\widehat{\ell}_\nu - \rho_{\nu n}\right)\left[1 + c(\rho_\nu)\ell_\nu + o_p(1)\right] \\ &+ n^{-1/2}p_\nu^T W_n(\rho_{\nu n})p_\nu + o_p\left(n^{-1/2}\right), \end{aligned} \tag{5.35}$$

where the last equality follows from (5.23), (5.27), and (5.28). Combining (5.34) and (5.35) yields (5.33).

The asymptotic normality of $\sqrt{n}\left(\widehat{\ell}_\nu - \rho_{\nu n}\right)$ now follows from Proposition 3, with asymptotic variance

$$\tilde{\sigma}_\nu^2 = [1 + c(\rho_\nu)\ell_\nu]^{-2}\mathrm{Var}\left[p_\nu^T W^\nu p_\nu\right] = (\dot{\rho}_\nu\ell_\nu/\rho_\nu)^2 \sum_{i,j,i',j'} \mathscr{P}_{iji'j'}^\nu \mathrm{Cov}\left[W_{ij}^\nu, W_{i'j'}^\nu\right],$$

where W^ν is the $m \times m$ symmetric Gaussian random matrix defined in Proposition 3, with covariance $\mathrm{Cov}\left[W_{ij}^\nu, W_{i'j'}^\nu\right]$ given by (5.26). Using this in the developed expression for the variance above leads to

$$\tilde{\sigma}_\nu^2 = \dot{\rho}_\nu \sum_{i,j,i',j'} \mathscr{P}_{iji'j'}^\nu\left(\kappa_{ij'}\kappa_{ji'} + \kappa_{ii'}\kappa_{jj'}\right) + \dot{\rho}_\nu^2\left[\mathscr{P}^\nu, \kappa + \check{\kappa}\right]. \tag{5.36}$$

By symmetry and the eigen equation $(\Gamma p_\nu)_i = \sum_j \kappa_{ij} p_{\nu,j} = \ell_\nu p_{\nu,i}$, we have

$$\sum_{i,j,i',j'} \mathscr{P}^{\nu}_{iji'j'} \kappa_{ii'} \kappa_{jj'} = \sum_{i,j,i',j'} \mathscr{P}^{\nu}_{iji'j'} \kappa_{ij'} \kappa_{ji'} = \sum_{i,j} p_{\nu,i} p_{\nu,j} (\Gamma p_\nu)_i (\Gamma p_\nu)_j = \ell_\nu^2 \sum_{i,j} \left(p_{\nu,i} p_{\nu,j}\right)^2 = \ell_\nu^2 .$$

Therefore, the first sum in (5.36) reduces to $2\dot{\rho}_\nu \ell_\nu^2$, yielding formula (2.6) of Theorem 1.

6. Proofs of the eigenvector results

We now derive the main eigenvector results, presented in Theorem 2 and Theorem 3-(*ii*).

6.1 Eigenvector inconsistency (Theorem 2-(*i*))

The convergence result of Theorem 2-(*i*) follows from two facts: $a_\nu \xrightarrow{\text{a.s.}} p_\nu$ and $Q_\nu \xrightarrow{\text{a.s.}} c(\rho_\nu)\Gamma$, which are shown below. Once these facts are established, from (3.10),

$$\|\hat{p}_\nu\|^{-2} \xrightarrow{\text{a.s.}} p_\nu^T (I_m + c(\rho_\nu)\Gamma) p_\nu = 1 + c(\rho_\nu)\ell_\nu = \frac{\rho_\nu}{\ell_\nu \dot{\rho}_\nu},$$

which leads to

$$\text{a.s. lim} \langle \hat{\mathfrak{p}}_\nu, \mathfrak{p}_\nu \rangle^2 = \text{a.s.lim} \langle \hat{p}_\nu, p_\nu \rangle^2 = \text{a.s.lim} \|\hat{p}_\nu\|^2 = \frac{\ell_\nu \dot{\rho}_\nu}{\rho_\nu}.$$

Proof of $a_\nu \xrightarrow{\text{a.s.}} p_\nu$**—**This is a direct consequence of (3.12) and

$$D_\nu = K(\rho_{\nu n}) - (\rho_{\nu n}/\ell_\nu)\Gamma + K\left(\hat{\ell}_\nu\right) - K(\rho_{\nu n}) \xrightarrow{\text{a.s.}} 0,$$

which follows from (5.22), (5.23), and the fact that $\hat{\ell}_\nu - \rho_{\nu n} \xrightarrow{\text{a.s.}} 0$, given in (2.5).

Proof of $Q_\nu \xrightarrow{\text{a.s.}} c(\rho_\nu)\Gamma$**—**With $\breve{Z}(t) = \left(tI_p - R_{22}\right)^{-1}$, we have

$$Q_\nu = R_{12} \breve{Z}^2(\rho_\nu) R_{21} + R_{12} \left[\breve{Z}^2\left(\hat{\ell}_\nu\right) - \breve{Z}^2(\rho_\nu) \right] R_{21} \triangleq Q_{\nu 1} + Q_{\nu 2}.$$

Rewrite $Q_{\nu 1} = n^{-1} \overline{X}_1 \breve{B}_{n1} \overline{X}_1^T$, with $\breve{B}_{n1} = n^{-1} \overline{X}_2^T \breve{Z}^2(\rho_\nu) \overline{X}_2$. On the high-probability event $J_{n\epsilon 1} = \{\mu_1 \le b_\gamma + \epsilon\}$, with $\epsilon > 0$ such that $\rho_\nu - b_\gamma \ge 2\epsilon$, it is easily established that $\|\breve{B}_{n1}\|$ is bounded and, consequently, that $\|\breve{B}_{n1}\| = O_{\text{a.s.}}(1)$. Hence, Lemma 2 can be applied to $Q_{\nu 1}$. Moreover, from (5.19) and noting that

$$n^{-1} \text{tr} \breve{B}_{n1} = n^{-1} \text{tr} B_{n1}(\rho_\nu, \rho_\nu),$$

with B_{n1} defined in (5.25), we have

$$n^{-1}\mathrm{tr}\check{B}_{n1} \xrightarrow{\text{a.s.}} \int x(\rho_v - x)^{-2}\mathrm{F}_\gamma(dx) = c(\rho_v).$$

This and Lemma 2 imply that $Q_{v1} \xrightarrow{\text{a.s.}} c(\rho_v)\Gamma$.

It remains to show $Q_{v2} \xrightarrow{\text{a.s.}} 0$. Using a variant of the resolvent identity, that is, $A^{-2} - B^{-2} = -A^{-2}(A^2 - B^2)B^{-2}$ for square invertible A and B, we rewrite

$$Q_{v2} = -2\left(\hat{\ell}_v - \rho_v\right)n^{-1}\overline{X}_1\check{B}_{n2}\overline{X}_1^T,$$

with $\check{B}_{n2} = n^{-1}\overline{X}_2^T\check{Z}^2\left(\hat{\ell}_v\right)\left[\frac{1}{2}\left(\hat{\ell}_v + \rho_v\right)I - R_{22}\right]\check{Z}^2(\rho_v)\overline{X}_2$. Working on the high-probability event $J_{n\epsilon}$, it can be verified that $\left\|\check{B}_{n2}\right\| = O_{\text{a.s.}}(1)$. Thus, Lemma 2 together with (5.19) imply that $n^{-1}\overline{X}_1\check{B}_{n2}\overline{X}_1^T = O_{\text{a.s.}}(1)$. Because $\hat{\ell}_v \xrightarrow{\text{a.s.}} \rho_v$, we conclude that $Q_{v2} \xrightarrow{\text{a.s.}} 0$.

6.2 Eigenvector fluctuations (Theorem 2-(*ii*))

Again, we use the key expansion (3.12). Because $\|r_v\| = O(\|D_v\|^2) = O_p(n^{-1})$ from (5.32), we have

$$\sqrt{n}(a_v - p_v) = -\mathscr{R}_{vn}\sqrt{n}D_v p_v + o_p(1).$$

Furthermore, using a similar decomposition to the derivation of (5.35),

$$\begin{aligned}\sqrt{n}D_v &= \sqrt{n}\left[K\left(\hat{\ell}_v\right) - K(\rho_{vn})\right] + \sqrt{n}[K(\rho_{vn}) - K_0(\rho_{vn}, \gamma_n)] \\ &= W_n(\rho_{vn}) - \sqrt{n}\left(\hat{\ell}_v - \rho_{vn}\right)c(\rho_v)\Gamma + o_p(1),\end{aligned}$$

where we use (5.23) and (5.27), along with (5.28) and (5.30) of Lemma 3. Hence, noting that $\mathscr{R}_{vn}\Gamma p_v = \ell_v\mathscr{R}_{vn}p_v = 0$ from the definition of $\mathscr{R}_{vn}$ in (3.12), we have

$$\sqrt{n}(a_v - p_v) = -\mathscr{R}_{vn}W_n(\rho_{vn})p_v + o_p(1),$$

or equivalently,

$$\sqrt{n}\left(P^T a_v - e_v\right) = -\widetilde{\mathscr{R}}_{vn}\widetilde{W}_n(\rho_{vn})e_v + o_p(1),$$

where

$$\widetilde{\mathscr{R}}_{vn} = \frac{\ell_v}{\rho_{vn}}\sum_{k\neq v}^{m}(\ell_k - \ell_v)^{-1}e_k e_k^T, \quad \widetilde{W}_n(\rho_{vn}) = P^T W_n(\rho_{vn})P.$$

The CLT for $P^T a_v$ now follows from Proposition 3. In particular,

$$\sqrt{n}\left(P^T a_\nu - e_\nu\right) \xrightarrow{\mathcal{D}} \widetilde{\mathcal{R}}_\nu w_\nu \sim N(0, \Sigma_\nu),$$

where $\widetilde{\mathcal{R}}_\nu = (\ell_\nu/\rho_\nu)\mathcal{D}_\nu$, recall (2.8), and $w_\nu = P^T W^\nu p_\nu$, with W^ν defined in Proposition 3. The covariance matrix $\Sigma_\nu = \widetilde{\mathcal{R}}_\nu \mathbb{E}\left[w_\nu w_\nu^T\right]\widetilde{\mathcal{R}}_\nu = \mathcal{D}_\nu \widetilde{\Sigma}_\nu \mathcal{D}_\nu$, with $\widetilde{\Sigma}_\nu = (\ell_\nu/\rho_\nu)^2 \mathbb{E}\left[w_\nu w_\nu^T\right]$. The kth component of w_ν is given by $w_\nu(k) = p_k^T W^\nu p_\nu = \sum_{i,j} p_{k,i} W^\nu_{ij} p_{\nu,j}$ and, therefore,

$$\widetilde{\Sigma}_{\nu,kl} = \sum_{i,j,i',j'} p_{k,i} p_{\nu,j} p_{l,i'} p_{\nu,j'} (\ell_\nu/\rho_\nu)^2 \mathrm{Cov}\left[W^\nu_{ij}, W^\nu_{i'j'}\right]. \tag{6.37}$$

Theorem 2-(*ii*) follows after substituting (5.26) for $\mathrm{Cov}\left[W^\nu_{ij}, W^\nu_{i'j'}\right]$ and noting that, when $k, l \neq \nu$,

$$\sum_{i,j,i',j'} p_{k,i} p_{\nu,j} p_{l,i'} p_{\nu,j'} \left(\kappa_{ii'}\kappa_{jj'} + \kappa_{ij'}\kappa_{ji'}\right) = p_k^T \Gamma p_l \cdot p_\nu^T \Gamma p_\nu + p_k^T \Gamma p_\nu \cdot p_\nu^T \Gamma p_l = \delta_{kl} \ell_k \ell_\nu.$$

6.3 Eigenvector inconsistency in the subcritical case (Theorem 3-(*ii*))

From (3.10) and (3.11), it suffices to show that $a_\nu^T Q_\nu a_\nu \xrightarrow{\text{a.s.}} \infty$ in order for Theorem 3-(*ii*) to hold. We establish this by showing that $\lambda_{\min}(Q_\nu) \xrightarrow{\text{a.s.}} \infty$. The approach uses a regularized version of Q_ν,

$$Q_{\nu\epsilon}(t) = R_{12}\left[\left(tI_p - R_{22}\right)^2 + \epsilon^2 I_p\right]^{-1} R_{21},$$

for $\epsilon > 0$. Observe that $Q_\nu \succ Q_{\nu\epsilon}\left(\widehat{\ell}_\nu\right)$, such that

$$\liminf \lambda_{\min}(Q_\nu) \geq \liminf \lambda_{\min}\left(Q_{\nu\epsilon}\left(\widehat{\ell}_\nu\right)\right) = \liminf \lambda_{\min}\left(Q_{\nu\epsilon}(b_\gamma) + \Delta_{\nu\epsilon}\right),$$

where $\Delta_{\nu\epsilon} := Q_{\nu\epsilon}\left(\widehat{\ell}_\nu\right) - Q_{\nu\epsilon}(b_\gamma)$ (Recall that $\widehat{\ell}_\nu \xrightarrow{\text{a.s.}} b_\gamma$). We show that $\Delta_{\nu\epsilon} \xrightarrow{\text{a.s.}} 0$, and

$$Q_{\nu\epsilon}(b_\gamma) \xrightarrow{\text{a.s.}} \int x\left[\left(b_\gamma - x\right)^2 + \epsilon^2\right]^{-1} \mathrm{F}_\gamma(dx) \cdot \Gamma = c_\gamma(\epsilon)\Gamma, \tag{6.38}$$

say. Because $\lambda_{\min}(\cdot)$ is a continuous function on $m \times m$ matrices, we conclude that

$$\liminf \lambda_{\min}(Q_\nu) \geq c_\gamma(\epsilon)\lambda_{\min}(\Gamma), \tag{6.39}$$

and because $c_\gamma(\epsilon) \geq c(b_\gamma + \epsilon)$ and $c(b_\gamma + \epsilon) \nearrow \infty$ as $\epsilon \searrow 0$, by [JY, Appendix A], we obtain $\lambda_{\min}(Q_\nu) \xrightarrow{\text{a.s.}} \infty$. We write $Q_{\nu\epsilon}(t) = n^{-1}\overline{X}_1 \breve{B}_{n\epsilon}(t)\overline{X}_1$, with

$$\begin{aligned}\breve{B}_{n\epsilon}(t) &= n^{-1}\overline{X}_2^T\left[\left(tI_p - n^{-1}\overline{X}_2\overline{X}_2^T\right)^2 + \epsilon^2 I_p\right]^{-1}\overline{X}_2 \\ &= H\mathrm{diag}\{f_\epsilon(\mu_i, t)\}H^T,\end{aligned}$$

if we write the singular-value decomposition of $n^{-1/2}\overline{X}_2 = V\mathscr{M}^{1/2}H^T$, with $\mathscr{M} = \mathrm{diag}(\mu_i)_{i=1}^p$ and define $f_\epsilon(\mu, t) = \mu\left[(t-\mu)^2 + \epsilon^2\right]^{-1}$. Evidently, $\left\|\breve{B}_{n\epsilon}(t)\right\| \le \epsilon^{-2}\mu_1$ is bounded almost surely. Thus, Lemma 2 may be applied to $Q_{\nu\epsilon}(b_\gamma)$, and because

$$n^{-1}\mathrm{tr}\breve{B}_{n\epsilon}(b_\gamma) \xrightarrow{\text{a.s.}} \int f_\epsilon(x, b_\gamma)\mathrm{F}_\gamma(dx) = c_\gamma(\epsilon)$$

from (5.19), our claim (6.38) follows.

Now consider $\Delta_{\nu\epsilon}$. Fix $a \in \mathbb{R}^m$ such that $\|a\|_2 = 1$, and set $b = n^{-1/2}H^T\overline{X}_1^T a$. We have

$$a^T\Delta_{\nu\epsilon}a = \sum_{i=1}^p b_i^2\left[f_\epsilon\left(\mu_i, \widehat{\ell}_\nu\right) - f_\epsilon\left(\mu_i, b_\gamma\right)\right].$$

Because $|\partial f_\epsilon(\mu, t)/\partial t| = |2\mu(t-\mu)|/\left[(t-\mu)^2 + \epsilon^2\right]^2 \le \mu/\epsilon^3$, for $\mu, \epsilon > 0$, by the arithmeticmean–geometric-mean inequality, we have

$$\left|a^T\Delta_{\nu\epsilon}a\right| \le \mu_1\epsilon^{-3}\left|\widehat{\ell}_\nu - b_\gamma\right| \cdot \|b\|_2^2 = \mu_1\epsilon^{-3}\left|\widehat{\ell}_\nu - b_\gamma\right|a^T R_{11}a \le \mu_1\epsilon^{-3}\left|\widehat{\ell}_\nu - b_\gamma\right|\widehat{\ell}_1 \xrightarrow{\text{a.s.}} 0,$$

from Cauchy's interlacing inequality for eigenvalues of symmetric matrices, Theorem 1-(*i*) and Theorem 3-(*i*). Therefore, $\Delta_{\nu\epsilon} \xrightarrow{\text{a.s.}} 0$, and the proof of (6.39) and, hence, of Theorem 3-(*ii*) is complete.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported, in part, by NIH R01 EB001988 (IMJ, JY), the Hong Kong RGC General Research Fund 16202918 (MRM, DMJ), and a Samsung Scholarship (JY).

References

Bai Z and Yao J-F (2008). Central limit theorems for eigenvalues in a spiked population model. Annales de l'Institut Henri Poincaré, Probabilités et Statistiques 44(3), 447–474.

Bai ZD and Silverstein J (2009). Spectral Analysis of Large Dimensional Random Matrices (2nd ed.). New York: Springer.

Baik J, Ben Arous G, and Péché S (2005). Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. Annals of Probability 33(5), 1643–1697.

Baik J and Silverstein JW (2006). Eigenvalues of large sample covariance matrices of spiked population models. Journal of Multivariate Analysis 97(6), 1382–1408.

Bao Z, Pan G, and Zhou W (2012). Tracy-Widom law for the extreme eigenvalues of sample correlation matrices. Electronic Journal of Probability 17, 1–32.

Benaych-Georges F and Nadakuditi RR (2011). The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. Advances in Mathematics 227(1), 494–521.

Bianchi P, Najim J, Maida M, and Debbah M (2009). Performance analysis of some eigen-based hypothesis tests for collaborative sensing. In 2009 IEEE/SP 15th Workshop on Statistical Signal Processing, pp. 5–8.

Bloemendal A, Knowles A, Yau H-T, and Yin J (2016). On the principal components of sample covariance matrices. Probability Theory and Related Fields 164(1), 459–552.

Boik RJ (2003). Principal component models for correlation matrices. Biometrika 90(3), 679–701.

Cai TT and Jiang T (2011). Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices. Annals of Statistics 39(3), 1496–1525.

Cai TT and Jiang T (2012). Phase transition in limiting distributions of coherence of high-dimensional random matrices. Journal of Multivariate Analysis 107, 24–39.

Cocco S, Monasson R, and Sessak V (2011). High-dimensional inference with the generalized Hopfield model: Principal component analysis and corrections. Physical Review E 83(5), 051123.

Cocco S, Monasson R, and Weigt M (2013). From principal component to direct coupling analysis of co-evolution in proteins: Low-eigenvalue modes are needed for structure prediction. PLoS Computational Biology 9(8), 1–17.

Cochran D, Gish H, and Sinno D (1995). A geometric approach to multiple-channel signal detection. IEEE Transactions on Signal Processing 43(9), 2049–2057.

Couillet R and Debbah M (2011). Random Matrix Methods for Wireless Communications. Cambridge University Press.

Couillet R and Hachem W (2013). Fluctuations of spiked random matrix models and failure diagnosis in sensor networks. IEEE Transactions on Information Theory 59(1), 509–525.

Dahirel V, Shekhar K, Pereyra F, Miura T, Artyomov M, Talsania S, Allen TM, Altfeld M, Carrington MN, Irvine DJ, Walker BD, and Chakraborty AK (2011). Coordinate linkage of HIV evolution reveals regions of immunological vulnerability. Proceedings of the National Academy of Sciences 108(28), 11530–11535.

El Karoui N (2009). Concentration of measure and spectra of random matrices: Applications to correlation matrices, elliptical distributions and beyond. Annals of Applied Probability 19(6), 2362–2405.

Fang C and Krishnaiah P (1982). Asymptotic distributions of functions of the eigenvalues of some random matrices for nonnormal populations. Journal of Multivariate Analysis 12(1), 39–63.

Gao J, Han X, Pan G, and Yang Y (2017). High dimensional correlation matrices: The central limit theorem and its applications. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 79(3), 677–693.

Girshick MA (1939). On the sampling theory of roots of determinantal equations. Annals of Mathematical Statistics 10(3), 203–224.

Hachem W, Loubaton P, Mestre X, Najim J, and Vallet P (2013). A subspace estimator for fixed rank perturbations of large random matrices. Journal of Multivariate Analysis 114, 427–447.

Hero A and Rajaratnam B (2011). Large-scale correlation screening. Journal of the American Statistical Association 106(496), 1540–1552.

Hero A and Rajaratnam B (2012). Hub discovery in partial correlation graphs. IEEE Transactions on Information Theory 58(9), 6064–6078.

Jiang T (2004a). The asymptotic distributions of the largest entries of sample correlation matrices. Annals of Applied Probability 14(2), 865–880.

Jiang T (2004b). The limiting distributions of eigenvalues of sample correlation matrices. Sankhyā: The Indian Journal of Statistics (2003–2007) 66(1), 35–48.

Johnstone IM (2001). On the distribution of the largest eigenvalue in principal components analysis. Annals of Statistics 29(2), 295–327.

Johnstone IM and Yang J (2018). Notes on asymptotics of sample eigenstructure for spiked models with non-Gaussian data. arXiv:1810.10427.

Kollo T and Neudecker H (1993). Asymptotics of eigenvalues and unit-length eigenvectors of sample variance and correlation matrices. Journal of Multivariate Analysis 47(2), 283–300.

Konishi S (1979). Asymptotic expansions for the distributions of statistics based on the sample correlation matrix in principal component analysis. Hiroshima Mathematical Journal 9(3), 647–700.

Leshem A and van der Veen A-J (2001). Multichannel detection of Gaussian signals with uncalibrated receivers. IEEE Signal Processing Letters 8(4), 120–122.

Liu H, Hu Z, Mian A, Tian H, and Zhu X (2014). A new user similarity model to improve the accuracy of collaborative filtering. Knowledge-Based Systems 56, 156–166.

Mestre X and Vallet P (2017). Correlation tests and linear spectral statistics of the sample correlation matrix. IEEE Transactions on Information Theory 63(7), 4585–4618.

Paul D (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. Statistica Sinica 17, 1617–1642.

Pillai NS and Yin J (2012). Edge universality of correlation matrices. Annals of Statistics 40(3), 1737–1763.

Plerou V, Gopikrishnan P, Rosenow B, Amaral L, Guhr T, and Stanley H (2002). A random matrix approach to cross-correlations in financial data. Physical Review E 65, 066126.

Quadeer AA, Louie RHY, Shekhar K, Chakraborty AK, Hsing I-M, and McKay MR (2014). Statistical linkage analysis of substitutions in patient-derived sequences of genotype 1a Hepatitis C virus nonstructural protein 3 exposes targets for immunogen design. Journal of Virology 88(13), 7628–7644. [PubMed: 24760894]

Quadeer AA, Morales-Jimenez D, and McKay MR (2018). Co-evolution networks of HIV/HCV are modular with direct association to structure and function. PLoS Computational Biology 14(9), 1–29.

Ruan D, Meng T, and Gao K (2016). A hybrid recommendation technique optimized by dimension reduction. In 2016 8th International Conference on Modelling, Identification and Control (ICMIC), pp. 429–433.

Schott JR (1991). A test for a specific principal component of a correlation matrix. Journal of the American Statistical Association 86(415), 747–751.

Vallet P, Mestre X, and Loubaton P (2015). Performance analysis of an improved MUSIC DoA estimator. IEEE Transactions on Signal Processing 63(23), 6407–6422.

Xiao H and Zhou W (2010). Almost sure limit of the smallest eigenvalue of some sample correlation matrices. Journal of Theoretical Probability 23(1), 1–20.

Yang L, McKay MR, and Couillet R (2018). High-dimensional MVDR beamforming: Optimized solutions based on spiked random matrix models. IEEE Transactions on Signal Processing 66(7), 1933–1947.

Yao J, Zheng S, and Bai Z (2015). Large Sample Covariance Matrices and High-Dimensional Data Analysis. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

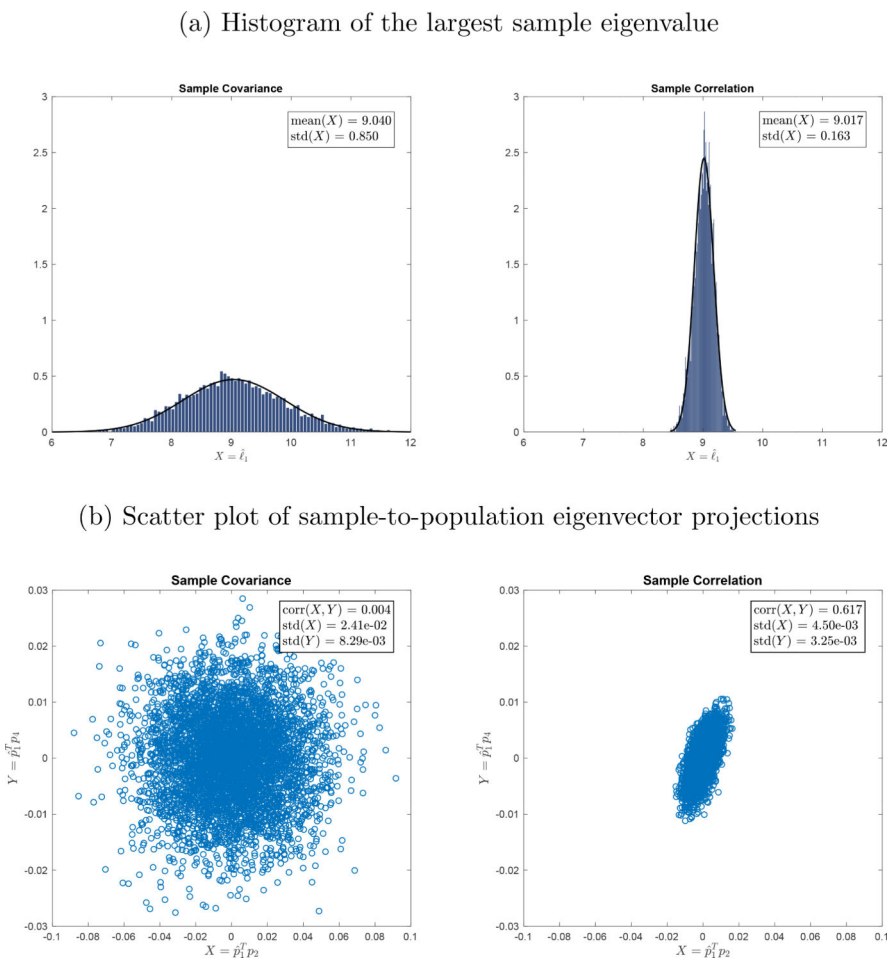

Figure 1:

A simple simulation shows remarkable distributional differences between sample covariance and sample correlation. From $n = 200$ i.i.d. Gaussian samples, $x_i \in \mathbb{R}^{100}$, with covariance $\Sigma = \text{blkdiag}(\Sigma_s, I_{90})$, where $(\Sigma_s)_{i,j=1}^{10} = \left(r^{|i-j|}\right)_{i,j=1}^{10}$, for $r = 0.95$, we compute the sample covariance and sample correlation, and show: (a) the empirical density (normalized histogram) of the largest sample eigenvalue, along with a Gaussian distribution with its estimated mean and standard deviation (solid line), and (b) a scatter plot of the leading sample eigenvector, projected onto the second (x-axis) and fourth (y-axis) population eigenvectors. A striking variance reduction is observed in the sample correlation for both (a) and (b). A similar variance reduction is observed for different choices of population eigenvectors in (b); the selected choice (being the second and fourth eigenvectors) facilitates the illustration of an additional correlation effect in the sample-to-population eigenvector projections.

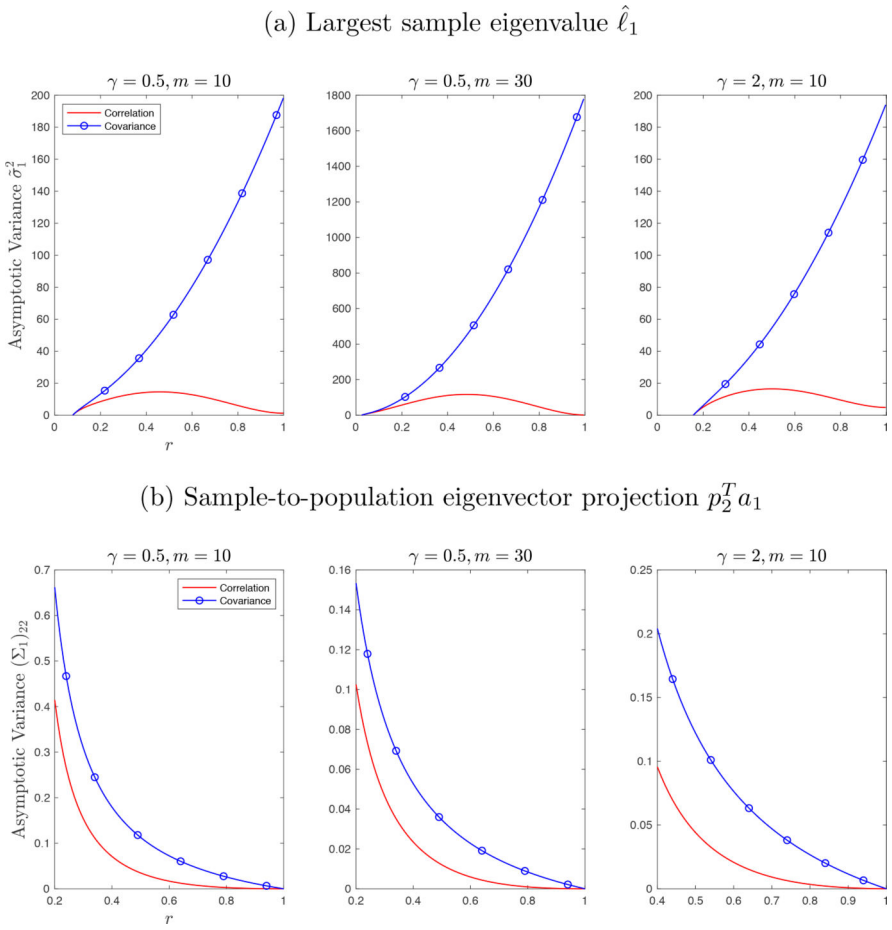

Figure 2:

Differences in the fluctuations of sample eigenvalues and eigenvectors for an example Gaussian model with $\Gamma = (1 - r)I_m + r1_m1_m^T$. Asympotic variances are shown for (a) the largest sample eigenvalue $\widehat{\ell}_1$, and (b) the normalized sample-to-population eigenvector projection $p_2^T a_1$.