Contents lists available at ScienceDirect

# Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

# Spiked singular values and vectors under extreme aspect ratios

# Michael J. Feldman

Department of Statistics, Sequoia Hall, Stanford University, Stanford CA 94305, United States of America

#### ARTICLE INFO

Article history: Received 10 January 2022 Received in revised form 28 March 2023 Accepted 29 March 2023 Available online 3 April 2023

AMS 2020 subject classifications: 60B20

*Keywords:* Random matrix Singular value Singular vector Spiked covariance Spiked model

#### ABSTRACT

The behavior of the leading singular values and vectors of noisy low-rank matrices is fundamental to many statistical and scientific problems. Theoretical understanding currently derives from asymptotic analysis under one of two regimes: *classical*, with a fixed number of rows, large number of columns or vice versa; and *proportional*, with large numbers of rows and columns, proportional to one another. This paper is concerned with the *disproportional* regime, where the matrix is either "tall and narrow" or "short and wide": we study sequences of matrices of size  $n \times m_n$  with aspect ratio  $n/m_n \rightarrow 0$  or  $n/m_n \rightarrow \infty$  as  $n \rightarrow \infty$ . This regime has important "big data" applications.

Theory derived here shows that the displacement of the empirical singular values and vectors from their noise-free counterparts and the associated phase transitions well-known under proportional growth asymptotics—still occur in the disproportionate setting. They must be quantified, however, on a novel scale of measurement that adjusts with the changing aspect ratio as the matrix size increases. In this setting, the top singular vectors corresponding to the longer of the two matrix dimensions are asymptotically uncorrelated with the noise-free signal.

© 2023 Elsevier Inc. All rights reserved.

#### 1. Introduction

The low-rank signal-plus-noise model is a simple statistical model of data with latent low-rank structure. Data  $\widetilde{X}_n$  is the sum of a low-rank matrix and white noise:

$$\frac{1}{\sqrt{m_n}}\widetilde{X}_n = \sum_{i=1}^r \theta_i u_i v_i^\top + \frac{1}{\sqrt{m_n}} X_n \tag{1}$$

where  $\theta_i \in \mathbb{R}$  are signal strengths,  $u_i \in \mathbb{R}^n$  and  $v_i \in \mathbb{R}^{m_n}$  are the left and right signal vectors, and the noise matrix  $X_n \in \mathbb{R}^{n \times m_n}$  contains independent and identically distributed (i.i.d.) entries with mean zero and variance one. The noise matrix is normalized so that rows are asymptotically unit norm.

### 1.1. Proportionate growth asymptotic

Recent work studies this model in the high-dimensional setting where n and  $m_n$  are large, in particular, where n and  $m_n$  are of comparable magnitude. Such analyses derive limiting behavior of sequences  $\widetilde{X}_n$  as

$$n, m_n \to \infty \quad \frac{n}{m_n} \to \beta > 0.$$

https://doi.org/10.1016/j.jmva.2023.105187 0047-259X/© 2023 Elsevier Inc. All rights reserved.





E-mail address: feldman6@stanford.edu.

M.J. Feldman

Here, the parameter  $\beta > 0$  is the limiting aspect ratio of the data; thus for  $\beta = 1$ , the matrices are effectively square for all large *n*. Baik, Ben Arous, and Péché [6], Baik and Silverstein [7], and Paul [16] study eigenvalues of the spiked covariance model, which is closely related to model (1). Benaych-Georges and Nadakuditi [9] derive asymptotic properties of the singular value decomposition of  $\tilde{X}_n$  (details of model (1) are provided in Section 1.3). Two phenomena arise not present in classical fixed-*n* asymptotics:

*Leading eigenvalue displacement.* Let  $\tilde{\lambda}_1 \geq \cdots \geq \tilde{\lambda}_n$  denote the eigenvalues of  $\tilde{S}_n = \frac{1}{m_n} \tilde{X}_n \tilde{X}_n^\top$ , the sample covariance matrix. We assume signal strengths  $\theta_1 > \cdots > \theta_r$  are constant and distinct. The leading eigenvalues of  $\tilde{S}_n$  are inconsistent estimators of the eigenvalues of  $\tilde{ES}_n$ : for fixed  $i \geq 1$ ,

$$\tilde{\lambda}_i \xrightarrow{a.s.} \begin{cases} \frac{(1+\theta_i^2)(\beta+\theta_i^2)}{\theta_i^2}, & i \le r, \ \theta_i > \beta^{1/2} \\ (1+\sqrt{\beta})^2, & \text{otherwise} \,. \end{cases}$$

The singular values of  $\frac{1}{\sqrt{m_n}}\widetilde{X}_n$  are of course determined by the eigenvalues of  $\widetilde{S}_n$ , and vice versa. We will occasionally switch between the two without comment.

Leading singular vector inconsistency. Let  $\tilde{u}_1, \ldots, \tilde{u}_n$  and  $\tilde{v}_1, \ldots, \tilde{v}_m$  denote the left and right singular vectors of  $\tilde{X}_n$ . The leading singular vectors  $\tilde{u}_1, \ldots, \tilde{u}_r$  and  $\tilde{v}_1, \ldots, \tilde{v}_r$  are inconsistent estimators of the left and right signal vectors. For  $1 \le i \le n$  and  $1 \le j \le r$ ,

$$|\langle \tilde{u}_i, u_j \rangle|^2 \xrightarrow{a.s.} 1 - \frac{\beta(1+\theta_i^2)}{\theta_i^2(\theta_i^2+\beta)}, \quad |\langle \tilde{v}_i, v_j \rangle|^2 \xrightarrow{a.s.} 1 - \frac{\beta+\theta_i^2}{\theta_i^2(\theta_i^2+1)}, \quad i=j, \ \theta_i > \beta^{1/4}$$

and  $|\langle \tilde{u}_i, u_j \rangle|, |\langle \tilde{v}_i, v_j \rangle| \xrightarrow{a.s.} 0$  otherwise.

#### 1.2. Disproportionate growth asymptotic

Contrary to the proportionate growth asymptotic, the dimensions of many large data matrices are not comparable. For example, Novembre et al. [15] demonstrate a genetic dataset with 3,000 rows (people) and 250,000 columns (genetic measurements) has low-rank structure. Here the aspect ratio is  $\beta \approx 3/250$ .

This paper considers the signal-plus-noise model under disproportional growth:

$$n, m_n \to \infty \quad \beta_n = n/m_n \to 0$$
.

Transposing  $\widetilde{X}_n$  if necessary, our results also apply to  $\beta_n \to \infty$ . Substitution of  $\beta = 0$  into the above proportional-limit formulas heuristically suggests  $\lambda_i \xrightarrow{a.s.} 1 + \theta_i^2$  and  $|\langle \widetilde{u}_i, u_j \rangle|^2 \xrightarrow{a.s.} 1$  (right signal vectors are partially recovered). Indeed, these formulas are corollaries of Theorems 2.9 and 2.10 of [9]. Thus, under  $\beta_n \to 0$  and constant signal strengths, leading eigenvalues of  $\widetilde{S}_n$  consistently estimate those of  $\widetilde{ES}_n$ , and left singular vectors  $\widetilde{u}_i$ , corresponding to the shorter matrix dimension, consistently estimate left signal vectors  $u_i$ . In particular, leading eigenvalue displacement and (left) singular vector inconsistency no longer occur, and seemingly no phase transition exists.

Our main contribution is the discovery of a vanishingly small phase transition located at  $\beta_n^{1/4}$ . Signal strengths  $\theta_i = \theta_{i,n}$ , hitherto fixed, now vanish. For signal strengths above this "microscopic" threshold, left singular vectors partially recover their signal counterparts—enabling signal estimation at signal-to-noise ratios previously thought insufficiently strong. Right singular vectors, corresponding to the longer matrix dimension, are asymptotically uncorrelated with the signal.

Right singular vectors, corresponding to the longer matrix dimension, are asymptotically uncorrelated with the signal. We introduce a particular calibration  $\theta_n = \tau \cdot \beta_n^{1/4}$  between signal strength  $\theta_n$  and  $\beta_n$  depending on a new (constant) formal parameter  $\tau$ . Under this calibration, we rigorously establish a new set of formulas for the limiting displacement of singular values and inconsistency of singular vectors, in explicit terms of the parameter  $\tau$ . These results may be heuristically derived by substitution of  $\theta_n = \tau \cdot \beta_n^{1/4}$  into proportional formulas and taking limits. Many important applications previously made under proportional growth may in parallel fashion now be rigorously made under disproportional growth may in parallel fashion now be rigorously made under disproportional growth. For example, in the high-dimensional Gaussian mixture model with many more parameters *m* than samples *n*, our theory gives a phase transition for recovery of length-*n* signal vectors (encoding cluster membership). The optimal singular value hard-thresholding level for low-rank matrix recovery as  $\beta_n \to 0$  may be calculated, analogous to proportional results of Donoho and Gavish [12]. Furthermore, there are new potential uses. Spiked tensors closely relate to the disproportionate asymptotic: Montanari and Richard propose a tensor unfolding algorithm [14] that produces noisy low-rank matrices of size  $n \times n^k$ ,  $k \ge 2$  (equivalently,  $\beta_n = n^{-(k-1)}$ ). Precise asymptotic analysis of tensor unfolding is made possible by our results.

The related spiked covariance model is studied in the disproportionate asymptotic by Bloemendal, Knowles, Yau, and Yin [10]. An advantage of this work is that  $\beta_n$  is permitted to vanish or diverge at any rate, while [10] requires the existence of a constant k > 0 such that  $n^{1/k} \le m_n \le n^k$ .

#### 1.3. Assumptions and notation

We make the following assumptions:

(A1. Noise) The entries of  $X_n = (x_{ij} : 1 \le i \le n, 1 \le j \le m_n)$  are i.i.d. with  $Ex_{11} = 0$ ,  $Ex_{11}^2 = 1$ , and  $Ex_{11}^4 < \infty$ . (A2. Signal vectors) The signal rank r is fixed. Let  $u_i$  and  $v_i$  be the *i*th columns of  $U_n \in \mathbb{R}^{n \times r}$  and  $V_n \in \mathbb{R}^{m_n \times r}$ , respectively. The entries of  $\sqrt{n}U_n$  and  $\sqrt{m_n}V_n$  are i.i.d. with mean zero, variance one, and finite eighth moment. (A3. Signal Strength) The signal strengths  $\theta_i = \theta_{i,n}$  obey

$$\theta_i = \tau_i \beta_n^{1/4} (1 + \varepsilon_{i,n})$$

for distinct constants  $\tau_1 > \cdots > \tau_r \ge 0$  and (nonrandom) sequences  $\varepsilon_{i,n} \to 0$ .

Alternatively, A1 and A2 may be replaced with

(B1. Noise) The entries of  $X_n = (x_{ij} : 1 \le i \le n, 1 \le j \le m_n)$  are independent standard Gaussians. (B2. Signal vectors) The signal rank r is fixed. Let  $u_i$  and  $v_i$  be the *i*th columns of  $U_n \in \mathbb{R}^{n \times r}$  and  $V_n \in \mathbb{R}^{m_n \times r}$ , respectively.  $U_n$  and  $V_n$  have orthonormal columns:  $U_n^\top U_n = V_n^\top V_n = I_r$ .

Henceforth, we suppress the subscript *n* of  $m_n$ . Unless explicitly stated otherwise, all results and lemmas in this paper hold simultaneously under (A1, A2, A3) and (B1, B2, A3). These assumptions are similar to those of [9]. The primary difference is [9] permits anisotropic noise  $X_n$ , provided two key conditions are satisfied: (1) the empirical spectral distribution (ESD) of  $S_n = m^{-1}X_nX_n^{\top}$  converges almost surely weakly to a deterministic, compactly supported measure  $\mu$  and (2) the maximum eigenvalue of  $S_n$  converges almost surely to the supremum of the support of  $\mu$ . This paper considers isotropic noise for the following reasons: firstly, the extension from proportionate growth to disproportionate growth and vanishing signals is most clear in the fundamental, isotropic case. Secondly, the anisotropic extension requires disproportionate analogs of (1) and (2), such as (3) convergence of the ESD of  $\beta_n^{-1/2}(m^{-1}\Sigma_n^{1/2}X_nX_n^{\top}\Sigma_n^{1/2} - \Sigma_n)$  to a deterministic, compactly supported measure  $\mu$  and (4) convergence of the maximum eigenvalue of this matrix to the supremum of the support of  $\mu$ . While (3) is established by Wang and Paul [17], (4) is proven only in the isotropic case, by Chen and Pan [11] (see Lemma 14). Assumption A2 is a relaxation of assumption 2.4 of [9], which requires sub-Gaussian moments of the entries of  $\sqrt{n}U_n$  and  $\sqrt{m_n}V_n$ . The assumption of distinct signal strengths (A3) is for simplicity.

In this paper, "almost surely eventually" means with probability one, a sequence of events indexed by *n* occurs for all sufficiently large *n*. The notation  $a_n \leq b_n$  means eventually  $a_n \leq Cb_n$  for some universal constant *C*;  $a_n \asymp b_n$  means  $b_n \leq a_n \leq b_n$ ; and  $a_n = O(b_n)$  means  $|a_n| \leq b_n$ . For a parameter  $\ell$ ,  $a_n \leq_\ell b_n$  means  $a_n \leq C(\ell)b_n$ .

#### 1.4. Results

As stated above, we study vanishing signal strengths  $\theta_i = \tau_i \beta_n^{1/4} (1 + \varepsilon_{i,n})$ . The formal parameter  $\tau$  describes the signal strength on a refined scale of analysis, with a phase transition occurring at  $\tau = 1$ . While for  $\tau < 1$ , both left and right singular vectors are asymptotically uncorrelated with the underlying signal vectors, for  $\tau > 1$ , left singular vectors correlate with left signal vectors. Define the following useful index:

$$i_0 = \max\{1 \le i \le r : \tau_i > 1\}$$

**Theorem 1.** Let  $\widetilde{X}_n$  denote a sequence of signal-plus-noise models satisfying the above assumptions. As  $n \to \infty$  and  $\beta_n \to 0$ , for any fixed  $i \ge 1$ ,

$$\frac{\tilde{\lambda}_i - 1}{\sqrt{\beta_n}} \xrightarrow{a.s.} \begin{cases} \tau_i^2 + \frac{1}{\tau_i^2} & i \le i_0\\ 2 & i > i_0 \end{cases}$$
(2)

For  $1 \leq i, j \leq i_0$ ,

$$|\langle \tilde{u}_i, u_j \rangle|^2 \xrightarrow{a.s.} \delta_{ij} \cdot (1 - \tau_i^{-4}), \qquad |\langle \tilde{v}_i, v_j \rangle|^2 \xrightarrow{a.s.} 0.$$
(3)

Theorem 1 is a consequence of the following stronger result: for  $i \leq i_0$ , define

$$\bar{\lambda}_{i} = \frac{(1+\theta_{i}^{2})(\beta_{n}+\theta_{i}^{2})}{\theta_{i}^{2}} = 1 + (\tau_{i}^{2}+\tau_{i}^{-2})\sqrt{\beta_{n}} + \beta_{n}.$$
(4)

**Theorem 2.** Adopt the setting of Theorem 1. For  $1 \le i, j \le i_0$ ,  $i \ne j$ , and  $\ell < 1/4$ , almost surely,

$$|\tilde{\lambda}_i - \bar{\lambda}_i| \lesssim (n^{-\ell} + |\varepsilon_{i,n}|) \sqrt{\beta_n} \,, \tag{5}$$

$$|\langle \tilde{u}_i, u_i \rangle|^2 = 1 - \tau_i^{-4} + O(n^{-\ell/2} + |\varepsilon_{i,n}| + \sqrt{\beta_n}), \qquad \qquad |\langle \tilde{u}_i, u_j \rangle|^2 \lesssim n^{-2\ell}, \tag{6}$$

$$|\langle \tilde{v}_i, v_i \rangle|^2 \lesssim \sqrt{\beta_n} \,, \qquad \qquad |\langle \tilde{v}_i, v_j \rangle|^2 \lesssim n^{-2\ell} \sqrt{\beta_n} \,. \tag{7}$$

#### 2. Preliminaries

## 2.1. Overview

This section reviews the framework developed in Benaych-Georges and Nadakuditi's work [8,9] addressing the  $\beta_n \rightarrow \beta > 0$ , fixed signal setting, followed by a discussion of adaptations required to study the  $\beta_n \rightarrow 0$ , vanishing signal setting. As before, let  $S_n = \frac{1}{m} X_n X_n^{\top}$  be the sample covariance matrix of noise. Let  $F_n$  denote the ESD of  $S_n$  and  $\overline{F}_n$  the Marchenko-

Pastur law with parameter  $\beta_n$ .  $s_n(z)$  and  $\bar{s}_n(z)$  will denote the corresponding Stieltjes transforms, defined as follows:

$$s_n(z) = \int \frac{1}{\lambda - z} dF_n(\lambda) = \frac{1}{n} \operatorname{tr}(S_n - zI_n)^{-1}, \qquad \bar{s}_n(z) = \int \frac{1}{\lambda - z} d\bar{F}_n(\lambda) = \frac{1 - \beta_n - z + \sqrt{(1 + \beta_n - z)^2 - 4\beta_n}}{2\beta_n z}, \quad (8)$$

defined for z outside the support of  $F_n$ ,  $\overline{F}_n$ . The square root is the principal branch.

Let  $\Theta_n = \text{diag}(\theta_1, \dots, \theta_r)$ . Recall that  $U_n$  and  $V_n$ , defined in Section 1.3, contain as columns the left and right signal vectors. The following  $2r \times 2r$  matrices will be the central objects of study, providing insight into the spectral decomposition of  $\widetilde{S}_n$ :

$$\widetilde{M}_{n}(z) = \begin{bmatrix} \sqrt{z}U_{n}^{\top}(S_{n} - zI_{n})^{-1}U_{n} & \frac{1}{\sqrt{m}}U_{n}^{\top}(S_{n} - zI_{n})^{-1}X_{n}V_{n} + \Theta_{n}^{-1} \\ \frac{1}{\sqrt{m}}V_{n}^{\top}X_{n}^{\top}(S_{n} - zI_{n})^{-1}U_{n} + \Theta_{n}^{-1} & \sqrt{z}V_{n}^{\top}(\frac{1}{m}X_{n}^{\top}X_{n} - zI_{m})^{-1}V_{n} \end{bmatrix},$$
(9)

$$M_n(z) = \begin{bmatrix} \sqrt{z}s_n(z)I_r & \Theta_n^{-1} \\ \Theta_n^{-1} & \left(\beta_n\sqrt{z}s_n(z) - \frac{1-\beta_n}{\sqrt{z}}\right)I_r \end{bmatrix},$$
(10)

$$\overline{M}_{n}(z) = \begin{bmatrix} \sqrt{z}s_{n}(z)I_{r} & \Theta_{n}^{-1} \\ \Theta_{n}^{-1} & \left(\beta_{n}\sqrt{z}\overline{s}_{n}(z) - \frac{1-\beta_{n}}{\sqrt{z}}\right)I_{r} \end{bmatrix}$$
(11)

Pivotal to their utility is this essential observation:

**Lemma 1** (Lemma 4.1 of [9]). The eigenvalues of  $\widetilde{S}_n$  that are not eigenvalues of  $S_n$  are the real values z such  $\widetilde{M}_n(z)$  is non-invertible.

Lemma 1 focuses the study of the leading eigenvalues of  $\tilde{S}_n$ —those resulting from the signal component of  $\tilde{X}_n$ —away from the *increasing*-dimensional sequence  $\tilde{S}_n$  to the *fixed*-dimensional sequence  $\tilde{M}_n$  and its associates,  $M_n$  and  $\overline{M}_n$ . The roots of det  $\tilde{M}_n(z)$  will be shown to localize near the roots of det  $\overline{M}_n(z)$ . Using the closed form (8) of the Stieltjes transform  $\bar{s}_n$ , det  $\overline{M}_n$  has explicitly known roots. As [9] shows, these roots are precisely  $\bar{\lambda}_1, \ldots, \bar{\lambda}_{i_0}$ , introduced in Section 1.4 (assuming  $\varepsilon_{i,n} = 0$ ):

Lemma 2 (Section 3.1 of [9]). Define

$$\overline{D}_n(z) = \sqrt{z}\overline{s}_n(z) \left(\beta_n \sqrt{z}\overline{s}_n(z) - \frac{1 - \beta_n}{\sqrt{z}}\right) = -\frac{1 + \beta_n - z + \sqrt{(1 + \beta_n - z)^2 - 4\beta_n}}{2\beta_n}.$$
(12)

As the blocks of  $\overline{M}_n(z)$  commute,

$$\det \overline{M}_n(z) = \det(\overline{D}_n(z)I_r - \Theta_n^{-2}) = \prod_{i=1}^r (\overline{D}_n(z) - \theta_i^{-2}).$$
(13)

On the real axis,  $\overline{D}_n(z)$  maps  $((1 + \sqrt{\beta_n})^2, \infty)$  in a one-to-one fashion onto  $(0, \beta_n^{-1/2})$ . The compositional inverse  $\overline{D}_n^{(-1)}(t)$  exists on the real interval  $(0, \beta_n^{-1/2})$  and is given in closed form by

$$\overline{D}_n^{(-1)}(t) = \frac{(t+1)(\beta_n t+1)}{t}, \qquad 0 < t < \beta_n^{-1/2}.$$
(14)

Thus, assuming  $\varepsilon_{i,n} = 0$ ,  $\bar{\lambda}_i = \overline{D}_n^{(-1)}(\theta_i^{-2})$  is a root of det  $\overline{M}_n(z)$  provided that  $\theta_i^{-2} < \beta_n^{-1/2}$ , or  $i \le i_0$ .

The singular vectors of  $\widetilde{X}_n$  are related to  $U_n$  and  $V_n$  via the next lemma.

**Lemma 3** (Lemma 5.1 of [9]). Let  $\tilde{\sigma} = \sqrt{\tilde{\lambda}}$  be a singular value of  $\frac{1}{\sqrt{m}}\tilde{X}_n$  that is not a singular value of  $\frac{1}{\sqrt{m}}X_n$  and let  $\tilde{u}, \tilde{v}$  be the corresponding singular vectors. Then the column vector

$$\begin{bmatrix} \Theta_n V_n^\top \tilde{v} \\ \Theta_n U_n^\top \tilde{u} \end{bmatrix}$$

belongs to the kernel of  $\widetilde{M}_n(\tilde{\lambda})$ . Moreover, for  $P_n = \sum_{i=1}^r \theta_i u_i v_i^\top$ ,

$$1 = \tilde{\lambda} \tilde{v}^{\top} P_n^{\top} (S_n - \tilde{\lambda} I_n)^{-2} P_n \tilde{v} + \frac{1}{m} \tilde{u}^{\top} P_n X_n^{\top} (S_n - \tilde{\lambda} I_n)^{-2} X_n P_n^{\top} \tilde{u} + \frac{\tilde{\sigma}}{\sqrt{m}} \tilde{u}^{\top} P_n X_n^{\top} (S_n - \tilde{\lambda} I_n)^{-2} P_n \tilde{v} + \frac{\tilde{\sigma}}{\sqrt{m}} \tilde{v}^{\top} P_n^{\top} (S_n - \tilde{\lambda} I_n)^{-2} X_n P_n^{\top} \tilde{u}.$$

$$(15)$$

The results quoted above lie at the heart of the strategy of [8,9] for proportionate growth and fixed signals. That strategy is able to employ convenient convergence arguments. Indeed, as  $\beta_n \to \beta > 0$ , almost surely, the ESD of noise eigenvalues  $F_n$  converges weakly to  $\overline{F}_{\beta}$ , the Marchenko–Pastur law with parameter  $\beta$ . Moreover, the top noise eigenvalue  $\lambda_1 \xrightarrow{a.s.} (1 + \sqrt{\beta})^2$ , the upper edge of the bulk (the support of  $\overline{F}_{\beta}$ ). Let [a, b] be a compact interval outside the bulk:  $(1 + \sqrt{\beta})^2 < a$ . By the Arzela–Ascoli theorem, the Stieltjes transform of the noise eigenvalues  $s_n(z)$  converges uniformly on [a, b] to  $\overline{s}_{\beta}(z)$ , the Stieltjes transform of  $\overline{F}_{\beta}$ . The implications of this key observation include the uniform convergence of  $M_n(z)$  to a non-random matrix  $\overline{M}_{\beta}(z)$  (the fixed- $\beta$  analog of  $\overline{M}_n(z)$ ). From the uniform convergence of  $M_n(z)$ , convergence of the roots of det  $M_n(z)$  to those of det  $\overline{M}_{\beta}(z)$  is established.

Now, consider the disproportionate asymptotic. Leading eigenvalue displacement and singular vector inconsistency no longer occur under fixed signals. Rather, the regime of interest is one where signal strengths vanish as  $\beta_n^{1/4}$ , mandating study of  $M_n(z)$  and  $\overline{M}_n(z)$  very near one. The convergences which were previously so helpful under  $\beta_n \rightarrow \beta > 0$  and fixed signals are not useful as  $\beta_n \rightarrow 0$ . The limits of  $F_n$  and  $\overline{M}_n(z)$  are degenerate, as all noise eigenvalues converge to one. The upper edge of the support of  $F_n$  lies approximately at  $(1 + \sqrt{\beta_n})^2$ . Theorem 1 tells us the leading eigenvalues of  $\widetilde{S}_n$  may still emerge from the bulk, but only by a multiple of  $\sqrt{\beta_n}$ . Phenomena of interest are no longer made visible by taking limits, but must be captured by detailed finite-n analysis. This will require bounds on the convergence rate of  $s_n(z)$  to  $\overline{s}_n(z)$ , developed in Section 5, as well as careful tracking of the precise size of remainders, where previously it sufficed to know that such remainders tended to zero. The proofs of Theorems 1 and 2 are refinements of the proofs of Theorems 2.9 and 2.10 of [9] and Lemma 6.1 of [8].

#### 2.2. Preliminary lemmas

This section contains lemmas used in the proof of Theorem 2.

**Lemma 4.** Let  $Y_n$  be the matrix obtained by entrywise truncation and normalization of  $X_n$  as in Lemma 15 and  $\widetilde{Y}_n = \sqrt{m} \sum_{i=1}^r \theta_i u_i v_i^\top + Y_n$ . For  $1 \le i \le n$ , almost surely,

$$| ilde{\lambda}_i - \lambda_i (m^{-1} \widetilde{Y}_n \widetilde{Y}_n^ op)| \lesssim rac{\sqrt{eta_n}}{n}$$

Moreover, if  $|\lambda_i(m^{-1}\widetilde{Y}_n\widetilde{Y}_n^{\top}) - \lambda_{i+1}(m^{-1}\widetilde{Y}_n\widetilde{Y}_n^{\top})| \asymp \sqrt{\beta_n}$  for  $1 \le i, j \le i_0$ ,

$$|\langle \tilde{u}_{i}, u_{j} \rangle|^{2} = |\langle \tilde{u}_{Y,i}, u_{j} \rangle|^{2} + O_{a.s.}(n^{-1}) \qquad |\langle \tilde{v}_{i}, v_{j} \rangle|^{2} = |\langle \tilde{v}_{Y,i}, v_{j} \rangle|^{2} + O_{a.s.}(n^{-1})$$

where  $\tilde{u}_{Y,1}, \ldots \tilde{u}_{Y,i_0}$  and  $\tilde{v}_{Y,1}, \ldots, \tilde{v}_{Y,i_0}$  are the leading left and right singular vectors of  $\tilde{Y}_n$ , respectively.

#### **Proof.** This follows from Lemma 16.

By Lemma 4, without loss of generality, we henceforth assume the entries of  $X_n$  are truncated and normalized in accordance with Lemma 15. Comparing Lemma 4 and Theorem 2, the errors induced by truncation and normalization are negligible; it therefore suffices to prove Theorem 2 in this modified setting. By Lemma 15,  $\bar{\lambda}_1, \ldots, \bar{\lambda}_{i_0}$  emerge from the ESD of  $S_n$  with high probability, a fact we shall use repeatedly: for any  $\eta$ ,  $\ell > 0$ ,

$$\Pr(\lambda_1 \ge 1 + (2 + \eta/2)\sqrt{\beta_n}) = o(n^{-\ell}).$$
(16)

The following lemma bounds  $s_n(z)$  and  $\bar{s}_n(z)$  on a region containing  $\bar{\lambda}_1, \ldots, \bar{\lambda}_{i_0}$ . The first point is a standard result. See, for example, Lemma 8.17 of [4]. The second point relies on (16).

**Lemma 5.** Let  $\eta > 0$  and  $\mathcal{Z}_{\eta,n} = \{z : \Re(z) \ge 1 + (2 + \eta)\sqrt{\beta_n}\}$ . Then,

$$\sup_{z\in \mathcal{Z}_{\eta,n}} |\bar{s}_n(z)| \lesssim \beta_n^{-1/2}, \qquad \sup_{z\in \mathcal{Z}_{\eta,n}} \left|\frac{d}{dz}\bar{s}_n(z)\right| \lesssim \beta_n^{-1}.$$

Furthermore, almost surely,

$$\sup_{z\in\mathcal{Z}_{\eta,n}}|s_n(z)|\lesssim\beta_n^{-1/2}\,,\qquad \sup_{z\in\mathcal{Z}_{\eta,n}}\left|\frac{d}{dz}s_n(z)\right|\lesssim\beta_n^{-1}.$$

The implied coefficients depend on  $\eta$  only.

**Proof.** As the upper edge of the support of the Marchenko–Pastur law lies at  $(1 + \sqrt{\beta_n})^2$ ,

$$\inf_{z\in\mathcal{Z}_{\eta,n}}\inf_{\lambda\in\mathrm{supp}\overline{F}_n}|\lambda-z|\gtrsim\sqrt{\beta_n}\,,\qquad\inf_{z\in\mathcal{Z}_{\eta,n}}\inf_{\lambda\in\mathrm{supp}\overline{F}_n}|\lambda-z|^2\gtrsim\beta_n\,.\tag{17}$$

The first claims of the lemma now follow from the integral representations

$$\bar{s}_n(z) = \int \frac{1}{\lambda - z} d\bar{F}_n(\lambda), \qquad \frac{d}{dz} \bar{s}_n(z) = -\int \frac{1}{(\lambda - z)^2} d\bar{F}_n(\lambda).$$

By (16), almost surely eventually,

$$\inf_{z \in \mathcal{Z}_{\eta,n}} \min_{1 \le \alpha \le n} |\lambda_{\alpha} - z| \ge \frac{\eta \sqrt{\beta_n}}{2}, \qquad \inf_{z \in \mathcal{Z}_{\eta,n}} \min_{1 \le \alpha \le n} |\lambda_{\alpha} - z|^2 \ge \frac{\eta^2 \beta_n}{4}.$$
(18)

Together with

$$s_n(z) = \frac{1}{n} \sum_{\alpha=1}^n \frac{1}{\lambda_\alpha - z}, \qquad \frac{d}{dz} s_n(z) = -\frac{1}{n} \sum_{\alpha=1}^n \frac{1}{(\lambda_\alpha - z)^2},$$

we obtain the second group of claims.  $\Box$ 

The follow lemmas bound the deviation of  $\widetilde{M}_n(z)$  (defined in (9)) and its entrywise derivative  $\frac{d}{dz}\widetilde{M}_n(z)$  from  $M_n(z)$  (defined in (10)) and  $\frac{d}{dz}\widetilde{M}_n(z)$ , respectively. Under assumption A2,  $M_n(z)$  and  $\frac{d}{dz}M_n(z)$  are precisely the entrywise conditional expectations of  $\widetilde{M}_n(z)$  and  $\frac{d}{dz}\widetilde{M}_n(z)$  given  $X_n$ .

**Lemma 6.** Let  $a_n \ge 1 + (2 + \eta)\sqrt{\beta_n}$  be a bounded sequence. For any  $\ell < 1/4$  and  $1 \le i, j \le r$ , almost surely,

$$\sup_{|z-a_n|\leq n^{-1/4}\sqrt{\beta_n}} \left| (\widetilde{M}_n(z) - M_n(z))_{i,j} \right| \lesssim n^{-\ell} \beta_n^{-1/2} ,$$
(19)

$$\sup_{|z-a_n| \le n^{-1/4} \sqrt{\beta_n}} \left| (\widetilde{M}_n(z) - M_n(z))_{i,j+r} \right| \le n^{-\ell} ,$$

$$(20)$$

$$\sup_{|z-a_n| \le n^{-1/4} \sqrt{\beta_n}} \left| (\widetilde{M}_n(z) - M_n(z))_{i+r,j+r} \right| \lesssim n^{-\ell} \sqrt{\beta_n} \,. \tag{21}$$

**Proof.** The argument is presented under assumptions (A1, A2, A3) for the diagonal of  $\widetilde{M}_n(z) - M_n(z)$  only; the off-diagonal argument or argument assuming (B1, B2, A3) is similar and omitted. Consider z satisfying  $|z - a_n| \le n^{-1/4}\sqrt{\beta_n}$  and  $z' = a_n + in^{-1/4}\sqrt{\beta_n}$ . We will decompose  $\widetilde{M}_n(z) - M_n(z)$  as

$$\widetilde{M}_n(z) - M_n(z) = \left(\widetilde{M}_n(z) - \widetilde{M}_n(z')\right) + \left(\widetilde{M}_n(z') - M_n(z')\right) + \left(M_n(z') - M_n(z)\right)$$

and bound (the diagonal entries of) each term. *Relation* (19): *the upper left block.* For  $1 \le \alpha \le n$ , by the almost-sure eventual spacing bound (18) and  $|z - z'| \le 2n^{-1/4}\sqrt{\beta_n}$ ,

$$\left|\frac{1}{\lambda_{\alpha}-z}-\frac{1}{\lambda_{\alpha}-z'}\right| = \frac{|z-z'|}{|\lambda_{\alpha}-z||\lambda_{\alpha}-z'|} \le \frac{8}{\eta^2 n^{1/4}\sqrt{\beta_n}}.$$
(22)

Hence, almost surely,

$$\begin{aligned} \left| \left( \frac{1}{\sqrt{z}} \widetilde{M}_n(z) - \frac{1}{\sqrt{z'}} \widetilde{M}_n(z') \right)_{i,i} \right| &= \left| u_i^\top (S_n - zI_n)^{-1} u_i - u_i (S_n - z'I_n)^{-1} u_i \right| \le \|u_i\|_2^2 \left\| (S_n - zI_n)^{-1} - (S_n - z'I_n)^{-1} \right\|_2 \\ &= \|u_i\|_2^2 \left| \frac{1}{\lambda_1 - z} - \frac{1}{\lambda_1 - z'} \right| \lesssim \frac{1}{n^{1/4} \sqrt{\beta_n}} \end{aligned}$$

$$(23)$$

$$\left| \left( \frac{1}{\sqrt{z}} M_n(z) - \frac{1}{\sqrt{z'}} M_n(z') \right)_{i,i} \right| = |s_n(z) - s_n(z')| \le \frac{1}{n} \sum_{\alpha=1}^n \left| \frac{1}{\lambda_\alpha - z} - \frac{1}{\lambda_\alpha - z'} \right| \le \frac{1}{n^{1/4} \sqrt{\beta_n}} \,. \tag{24}$$

Next, consider  $\widetilde{M}_n(z') - M_n(z')$ . Applying Lemma 17 conditional on  $X_n$ ,

$$\begin{split} \mathsf{E}|u_{i}^{\top}(S_{n}-z'I_{n})^{-1}u_{i}-s_{n}(z')|^{4} &\lesssim \frac{1}{n^{4}}\mathsf{E}|\mathsf{tr}(S_{n}-z'I_{n})^{-1}(S_{n}-\bar{z}'I_{n})^{-1}|^{2} + \frac{1}{n^{4}}\mathsf{E}\mathsf{tr}\big((S_{n}-z'I_{n})^{-1}(S_{n}-\bar{z}'I_{n})^{-1}\big)^{2} \\ &\leq \frac{2}{n^{4}}\mathsf{E}\Big(\sum_{\alpha=1}^{n}\frac{1}{|\lambda_{\alpha}-z'|^{2}}\Big)^{2}. \end{split}$$

By (16),  $|\lambda_{\alpha} - z'| \ge \eta \sqrt{\beta_n}/2$  with probability at least  $1 - o(n^{-1})$ , while with probability one,  $|\lambda_{\alpha} - z'| \ge \Im(z')$ :

$$\mathbb{E}|u_i^{\top}(S_n-z'I_n)^{-1}u_i-s_n(z')|^4 \lesssim \frac{1}{n^2}\left(\frac{1}{\beta_n^2}+\frac{1}{n\Im(z')^4}\right).$$

This yields the tail probability bound

$$\Pr(|u_i^{\top}(S_n - z'I_n)^{-1}u_i - s_n(z')| \ge n^{-\ell}\beta_n^{-1/2}) \le n^{4\ell}\beta_n^2 \mathsf{E}|u_i^{\top}(S_n - z'I_n)^{-1}u_i - s_n(z')|^4 \lesssim \frac{1}{n^{2-4\ell}}$$

which is summable. Thus, by the Borel-Cantelli lemma,

$$\left|\frac{1}{\sqrt{z'}} \left(\widetilde{M}_n(z') - M_n(z')\right)_{i,i}\right| \lesssim \frac{1}{n^\ell \sqrt{\beta_n}} \tag{25}$$

almost surely. (19) follows from (23)-(25),  $|\sqrt{z'/z}| \le 2$ , and  $|\sqrt{z'} - \sqrt{z}| \ge n^{-1/4}\sqrt{\beta_n}$ . *Relation* (21): *the lower right block.* Let  $S_n = \frac{1}{m}X_n^{\top}X_n$  denote the "companion" matrix to  $S_n$ , so-called because the eigenvalues of  $\check{S}_n$  are the *n* (possibly repeated) eigenvalues of  $S_n$  joined by a zero eigenvalue with multiplicity m - n. Let W be an orthogonal matrix such that  $W^{\dagger}\check{S}_{n}W$  is diagonal. By Lemma 21, we have an almost-sure bound on the magnitude of the component of  $v_i$  orthogonal to the nullspace of  $\tilde{S}_n$ :

$$\|(W^{\top}v_{i})_{1:n}\|_{2}^{2} \lesssim \beta_{n}\log(n).$$
(26)

Consider first  $\widetilde{M}_n(z) - \widetilde{M}_n(z')$ .

$$\begin{split} \left| \left( \frac{1}{\sqrt{z}} \widetilde{M}_{n}(z) - \frac{1}{\sqrt{z'}} \widetilde{M}_{n}(z') \right)_{i+r,i+r} \right| &= \left| v_{i}^{\top} \left( (\breve{S}_{n} - zI_{m})^{-1} - (\breve{S}_{n} - z'I_{m})^{-1} \right) v_{i} \right| \\ &\leq \sum_{\alpha=1}^{n} (W^{\top} v_{i})_{\alpha}^{2} \left| \frac{1}{\lambda_{\alpha} - z} - \frac{1}{\lambda_{\alpha} - z'} \right| + \sum_{\alpha=n+1}^{m} (W^{\top} v_{i})_{\alpha}^{2} \left| \frac{1}{z} - \frac{1}{z'} \right| \\ &\leq \left\| (W^{\top} v_{i})_{1:n} \right\|_{2}^{2} \left| \frac{1}{\lambda_{1} - z} - \frac{1}{\lambda_{1} - z'} \right| + \left\| (W^{\top} v_{i})_{n+1:m} \right\|_{2}^{2} \left| \frac{1}{z} - \frac{1}{z'} \right|. \end{split}$$

The first term above is bounded using (22) and (26) and the second using  $|1/z - 1/z'| \le 2n^{-1/4}\sqrt{\beta_n}$ : almost surely,

$$\left|\frac{1}{\sqrt{z}} \left(\widetilde{M}_n(z) - \widetilde{M}_n(z')\right)_{i+r,i+r}\right| \lesssim \frac{\sqrt{\beta_n}}{n^\ell} \,. \tag{27}$$

Next, consider  $M_n(z) - M_n(z')$ . By (24),

$$\left| \left( \frac{1}{\sqrt{z}} M_n(z) - \frac{1}{\sqrt{z'}} M_n(z') \right)_{i+r,i+r} \right| = \left| \beta_n s_n(z) - \frac{1 - \beta_n}{z} - \beta_n s_n(z') + \frac{1 - \beta_n}{z'} \right| \\ \leq \beta_n |s_n(z) - s_n(z')| + (1 - \beta_n) \left| \frac{1}{z} - \frac{1}{z'} \right| \lesssim \frac{\sqrt{\beta_n}}{n^{1/4}} \,.$$
(28)

 $\widetilde{M}_n(z') - M_n(z')$  is bounded similarly to (25): applying Lemma 17 conditional on  $X_n$  and (16),

$$\begin{split} \mathsf{E} \Big| v_i^\top (\breve{S}_n - z'I_m)^{-1} v_i - \beta_n s_n(z') + \frac{1 - \beta_n}{z'} \Big|^4 &\lesssim \frac{1}{m^4} \mathsf{E} \Big| \mathsf{tr}(\breve{S}_n - z'I_m)^{-1} (\breve{S}_n - \bar{z}'I_m)^{-1} \Big|^2 = \frac{1}{m^4} \mathsf{E} \Big( \sum_{\alpha=1}^n \frac{1}{|\lambda_\alpha - z'|^2} + \frac{m - n}{|z'|^2} \Big)^2 \\ &\lesssim \frac{1}{m^4} \Big( \frac{n}{\beta_n} + m \Big)^2 \,, \end{split}$$

Thus,

$$\Pr\left(\left|\frac{1}{\sqrt{z'}}(\widetilde{M}_n(z') - M_n(z'))_{i+r,i+r}\right| \ge n^{-\ell}\sqrt{\beta_n}\right) \lesssim \frac{n^{4\ell}}{m^2\beta_n^2} \le \frac{1}{n^{2-4\ell}}.$$
(29)

Summability of the right-hand side, together with (27) and (28), yields (21). The analysis of the off-diagonal entries of  $M_n(z) - M_n(z)$  uses Lemma 18 rather than 17. Under assumptions B1, B2, the proof uses that the left and right singular vectors of  $X_n$  are independent and each Haar-distributed.  $\Box$ 

**Lemma 7.** Let  $a_n \ge 1 + (2 + \eta)\sqrt{\beta_n}$  be a bounded sequence. For any  $\ell < 1/4$  and  $1 \le i, j \le r$ , almost surely,

$$\sup_{|z-a_n| \le n^{-1/4} \sqrt{\beta_n}} \left| \frac{d}{dz} (\widetilde{M}_n(z) - M_n(z))_{i,j} \right| \lesssim n^{-\ell} \beta_n^{-1} ,$$
(30)

$$\sup_{|z-a_n|\leq n^{-1/4}\sqrt{\beta_n}}\left|\frac{d}{dz}(\widetilde{M}_n(z)-M_n(z))_{i,j+r}\right|\lesssim n^{-\ell}\beta_n^{-1/2}\,,\tag{31}$$

$$\sup_{|z-a_n| \le n^{-1/4} \sqrt{\beta_n}} \left| \frac{d}{dz} (\widetilde{M}_n(z) - M_n(z))_{i+r,j+r} \right| \lesssim n^{-\ell} .$$
(32)

Furthermore, for  $i \neq j$ , almost surely,

$$\sup_{|z-a_n| \le n^{-1/4} \sqrt{\beta_n}} \frac{1}{m} \left| v_i^\top X_n^\top (S_n - zI_n)^{-2} X_n v_i - \operatorname{tr}(S_n - zI_n)^{-2} S_n \right| \le n^{-\ell} ,$$
(33)

$$\sup_{|z-a_n| \le n^{-1/4} \sqrt{\beta_n}} \frac{1}{m} |v_i^\top X_n^\top (S_n - zI_n)^{-2} X_n v_j| \lesssim n^{-\ell} \,. \tag{34}$$

**Proof.** The proof is similar to that of Lemma 6 and is omitted.  $\Box$ 

#### 3. Proof of singular value results: (2) and (5)

The following lemmas, which we shall use to bound  $|\det \widetilde{M}_n(z) - \det \overline{M}_n(z)|$  and  $\left|\frac{d}{dz}(\det \widetilde{M}_n(z) - \det \overline{M}_n(z))\right|$  in the vicinities of  $\overline{\lambda}_1, \ldots, \overline{\lambda}_{i_0}$ , are critical to the proof of the theorem.

**Lemma 8.** Let  $a_n \ge 1 + (2 + \eta)\sqrt{\beta_n}$  be a bounded sequence. For any  $\ell < 1/4$ , almost surely,

$$\sup_{|z-a_n| \le n^{-1/4}\sqrt{\beta_n}} |\det \widetilde{M}_n(z) - \det \overline{M}_n(z)| \lesssim n^{-\ell} \beta_n^{-r/2} \,. \tag{35}$$

**Proof.** Partition  $\widetilde{M}_n(z)$  into four blocks of size  $r \times r$ :

$$\widetilde{M}_n(z) = \begin{bmatrix} \widetilde{M}_{11} & \widetilde{M}_{12} \\ \widetilde{M}_{21} & \widetilde{M}_{22} \end{bmatrix}$$

Let  $\underline{\widetilde{M}}_{n}(z)$  denote the matrix containing the above four submatrices individually rescaled:

$$\underline{\widetilde{M}}_{n}(z) = \begin{bmatrix} \sqrt{\beta_{n}} \widetilde{M}_{11} & \beta_{n}^{1/4} \widetilde{M}_{12} \\ \beta_{n}^{1/4} \widetilde{M}_{21} & \widetilde{M}_{22} \end{bmatrix}.$$
(36)

Let  $\underline{M}_n(z)$  and  $\overline{\underline{M}}_n(z)$  denote  $M_n(z)$  and  $\overline{\underline{M}}_n(z)$  similarly partitioned and rescaled. This rescaling simplifies comparison of determinants as all blocks are of similar size. Indeed, as consequences of Lemma 5, Lemma 6, and Theorem 3 (stated in Section 5), we have the following almost-sure results:

$$\begin{aligned} \sup_{|z-a_n| \le n^{-1/4} \sqrt{\beta_n}} \|\underline{\underline{M}}_n(z)\|_{\infty} \lesssim 1 & \sup_{|z-a_n| \le n^{-1/4} \sqrt{\beta_n}} \|\underline{\underline{M}}_n(z) - \underline{\underline{M}}_n(z)\|_{\infty} \lesssim n^{-\ell} \\ \sup_{|z-a_n| \le n^{-1/4} \sqrt{\beta_n}} \|\underline{\underline{M}}_n(z) - \overline{\underline{M}}_n(z)\|_{\infty} \lesssim n^{-\ell} \end{aligned}$$
(37)

where  $\|\cdot\|_{\infty}$  is the maximum-magnitude entry of the input. Applying Lemma 19, we obtain

$$\sup_{|z-a_n|\leq n^{-1/4}\sqrt{\beta_n}} |\det \widetilde{\underline{M}}_n(z) - \det \overline{\underline{M}}_n(z)| \lesssim n^{-\ell},$$

which is equivalent to (35) as det  $\widetilde{M}_n(z) = \beta_n^{-r/2} \det \underline{\widetilde{M}}_n(z)$ .  $\Box$ 

**Lemma 9.** Let  $a_n \ge 1 + (2 + \eta)\sqrt{\beta_n}$  be a bounded sequence. For any  $\ell < 1/8$ , almost surely,

$$\sup_{|z-a_n|\leq n^{-1/4}\sqrt{\beta_n}}\left|\frac{d}{dz}(\det\widetilde{M}_n(z)-\det\overline{M}_n(z))\right| \lesssim n^{-\ell}\beta_n^{-(r+1)/2} \qquad \sup_{|z-a_n|\leq n^{-1/4}\sqrt{\beta_n}}\left|\frac{d}{dz}\det\widetilde{M}_n(z)\right| \lesssim \beta_n^{-(r+1)/2}$$

Proof. The proof is similar to that of Lemma 8. By Lemma 5, Lemma 7, and Corollary 1,

$$\sup_{\substack{|z-a_n| \le n^{-1/4}\sqrt{\beta_n}}} \left\| \frac{d}{dz} \underline{\overline{M}}_n(z) \right\|_{\infty} \lesssim \beta_n^{-1/2} \qquad \sup_{|z-a_n| \le n^{-1/4}\sqrt{\beta_n}} \left\| \frac{d}{dz} (\underline{\widetilde{M}}_n(z) - \underline{M}_n(z)) \right\|_{\infty} \lesssim n^{-\ell} \beta_n^{-1/2} \qquad (38)$$

The claim follows from Lemma 20, (37), and (38):

$$\left|\frac{d}{dz}(\det \underline{\widetilde{M}}_n(z) - \det \underline{\overline{M}}_n(z))\right| \lesssim \left\|\frac{d}{dz}(\underline{\widetilde{M}}_n(z) - \underline{\overline{M}}_n(z))\right\|_{\infty} + \left\|\frac{d}{dz}\underline{\overline{M}}_n(z)\right\|_{\infty} \|\underline{\widetilde{M}}_n(z) - \underline{M}_n(z)\|_{\infty} \lesssim n^{-\ell}\beta_n^{-1/2}. \quad \Box$$

**Proof of (2) and (5).** We initially assume  $\varepsilon_{i,n} = 0$ ,  $1 \le i \le r$ . Fix  $\ell < 1/4$  and let  $\gamma_{i,n}$  denote a circular contour with center  $\overline{\lambda}_i$  and radius  $n^{-\ell}\sqrt{\beta_n}$ . We divide the proof into four claims:

- (i)  $\gamma_{i,n}$  eventually encircles a single root of det  $\overline{M}_n(z)$ .
- (ii) Almost surely eventually, there exists a unique root  $\tilde{\lambda}_{(i)}$  of det  $\widetilde{M}_n(z)$  encircled by  $\gamma_{i,n}$ .
- (iii)  $\tilde{\lambda}_{(1)}, \ldots, \tilde{\lambda}_{(i_0)}$  are real and therefore eigenvalues of  $\tilde{S}_n$  by Lemma 1.
- (iv) Almost surely eventually,  $\tilde{S}_n$  has no eigenvalues other than  $\tilde{\lambda}_{(1)}, \ldots, \tilde{\lambda}_{(i_0)}$  larger than  $1 + (2 + \eta)\sqrt{\beta_n}$ , where  $\eta > 0$  is arbitrary.

Notice that  $\gamma_{i,n}$  eventually encircles a single real root of det  $\overline{M}_n(z)$ : as  $\tau_1, \ldots, \tau_{i_0}$  are distinct,  $|\overline{\lambda}_i - \overline{\lambda}_j| \gtrsim \sqrt{\beta_n}$ ,  $i \neq j$ . We now argue  $\gamma_{i,n}$  eventually encircles no complex roots of det  $\overline{M}_n(z)$ . For  $\xi \in \mathbb{C}$  such that  $|\xi| \leq n^{-\ell} \sqrt{\beta_n}$ ,

$$\overline{D}_{n}(\bar{\lambda}_{i}+\xi)-\theta_{i}^{-2} = \frac{\xi+\sqrt{(\bar{\lambda}_{i}-\beta_{n}-1)^{2}-4\beta_{n}(1-\sqrt{1+y})}}{2\beta_{n}}$$
$$y \coloneqq \frac{2(\bar{\lambda}_{i}-\beta_{n}-1)\xi+\xi^{2}}{(\bar{\lambda}_{i}-\beta_{n}-1)^{2}-4\beta_{n}} = \frac{2(\tau_{i}^{2}+\tau_{i}^{-2})\beta_{n}^{-1/2}\xi+\beta_{n}^{-1}\xi^{2}}{(\tau_{i}^{2}+\tau_{i}^{-2})^{2}-4}.$$

Using  $|1 - \sqrt{1 + z} + z/2| \le |z|^2/8$  for  $z \in \{\Re(z) \ge 0, |z| \le 1\}$ ,  $|\xi| \le n^{-\ell}\sqrt{\beta_n}$ , and  $|y| \le n^{-\ell}$ ,

$$|\overline{D}_n(\overline{\lambda}_i + \xi) - \theta_i^{-2}| \asymp |\xi| \beta_n^{-1}.$$
(39)

For  $j \neq i$ , almost surely,

$$|\overline{D}_n(\bar{\lambda}_i + \xi) - \theta_j^{-2}| = |\overline{D}_n(\bar{\lambda}_i + \xi) - \theta_i^{-2} + (\theta_i^{-2} - \theta_j^{-2})| \asymp \beta_n^{-1/2}.$$
(40)

Recalling the formula det  $\overline{M}_n(z) = \prod_{j=1}^r (\overline{D}_n(z) - \theta_j^{-2})$ , (39) and (40) imply

$$|\det \overline{M}_n(\bar{\lambda}_i + \xi)| \asymp |\xi| \beta_n^{-(r+1)/2} \,. \tag{41}$$

Consequently,  $\gamma_{i,n}$  eventually encircles no root of det  $\overline{M}_n(z)$  other than  $\overline{\lambda}_i$  (and no roots occur on  $\gamma_{i,n}$ ). Otherwise, there exists a sequence of perturbations  $\xi_n \neq 0$  such that det  $\overline{M}_n(\overline{\lambda}_i + \xi_n) = 0$ , contradicting (41). This proves Claim (i). As a consequence, by the winding number theorem,

$$\frac{1}{2\pi i} \int_{\gamma_{i,n}} \frac{\frac{d}{dz} \det \overline{M}_n(z)}{\det \overline{M}_n(z)} dz = 1.$$
(42)

To prove Claim (ii), it suffices to show that almost surely eventually,

$$\left|\frac{1}{2\pi i}\int_{\gamma_{i,n}}\frac{\frac{d}{dz}\det\widetilde{M}_n(z)}{\det\widetilde{M}_n(z)}dz-\frac{1}{2\pi i}\int_{\gamma_{i,n}}\frac{\frac{d}{dz}\det\overline{M}_n(z)}{\det\overline{M}_n(z)}dz\right|<1,$$

as the above integrals are integer-valued. Writing  $\frac{a}{c} - \frac{b}{d} = \frac{1}{d} \left( \frac{d-c}{c} a + a - b \right)$ , an upper bound on the left-hand side is

$$n^{-\ell}\sqrt{\beta_n}\sup_{z\in\gamma_{l,n}}\left|\frac{\frac{d}{dz}\det\widetilde{M}_n(z)}{\det\widetilde{M}_n(z)} - \frac{\frac{d}{dz}\det\overline{M}_n(z)}{\det\overline{M}_n(z)}\right| = n^{-\ell}\sqrt{\beta_n}\sup_{z\in\gamma_{l,n}}\left|\frac{1}{\det\overline{M}_n}\right|\left|\frac{\det\overline{M}_n - \det\widetilde{M}_n}{\det\widetilde{M}_n} \cdot \frac{d}{dz}\det\widetilde{M}_n + \frac{d}{dz}\left(\det\widetilde{M}_n - \det\overline{M}_n\right)\right|.$$
(43)

Below, the almost-sure bounds we have developed on the terms of (43) are summarized. Let  $\ell' \in (\ell, 1/4)$  and  $\ell'' \in (0, 1/8)$ . By Lemmas 8 and 9 (applied to  $a_n = \overline{\lambda}_i$  and  $\eta \in (0, \tau_{i_0}^2 + \tau_{i_0}^{-2} - 2)$ ),

$$\sup_{z \in \gamma_{l,n}} |\det \widetilde{M}_n(z) - \det \overline{M}_n(z)| \lesssim n^{-\ell'} \beta_n^{-r/2},$$
(44)

$$\sup_{z\in\gamma_{i,n}} \left|\frac{d}{dz} (\det \widetilde{M}_n(z) - \det \overline{M}_n(z))\right| \lesssim n^{-\ell''} \beta_n^{-(r+1)/2}$$
(45)

$$\sup_{z\in\gamma_{i,n}}\left|\frac{d}{dz}\det\widetilde{M}_n(z)\right|\lesssim\beta_n^{-(r+1)/2}\,.$$
(46)

As (41) depends on  $\xi$  only through  $|\xi|$ ,

$$|\det \overline{M}_n(z)| \asymp n^{-\ell} \beta_n^{-r/2}$$
 (47)

Finally, (44), (47), and  $\ell < \ell'$  imply,

$$|\det \widetilde{M}_n(z)| \asymp n^{-\ell} \beta_n^{-r/2} \,. \tag{48}$$

M.J. Feldman

Thus, from (43)–(48), almost surely,

$$n^{-\ell}\sqrt{\beta_n} \sup_{z \in \gamma_{l,n}} \left| \frac{\frac{d}{dz} \det \widetilde{M}_n(z)}{\det \widetilde{M}_n(z)} - \frac{\frac{d}{dz} \det \overline{M}_n(z)}{\det \overline{M}_n(z)} \right| \lesssim \frac{n^{-\ell}\sqrt{\beta_n}}{n^{-\ell}\beta_n^{-r/2}} \left( \frac{n^{-\ell'}\beta_n^{-r/2}}{n^{-\ell}\beta_n^{-r/2}} \cdot \beta_n^{-(r+1)/2} + n^{-\ell''}\beta_n^{-(r+1)/2} \right) \\ \lesssim n^{\ell-\ell'} + n^{-\ell''} \longrightarrow 0.$$
(49)

This completes the proof of Claim (ii): almost surely eventually, there exist roots  $\tilde{\lambda}_{(1)}, \ldots, \tilde{\lambda}_{(i_0)}$  respectively encircled by  $\gamma_{1,n}, \ldots, \gamma_{i_0,n}$ . Claim (iii)—that  $\tilde{\lambda}_{(1)}, \ldots, \tilde{\lambda}_{(i_0)}$  are eigenvalues of  $\tilde{S}_n$ —follows from Lemma 1 and the fact that  $\tilde{M}_n(z)$  is invertible for  $z \notin \mathbb{R}$  (Lemma A.1 of [9]).

It remains to prove Claim (iv):  $\tilde{\lambda}_1 = \tilde{\lambda}_{(1)}, \ldots, \tilde{\lambda}_{i_0} = \tilde{\lambda}_{(i_0)}$ . Suppose det  $\widetilde{M}_n(z)$  has a sequence of roots  $\tilde{\lambda} \in [1 + (2 + \eta)\sqrt{\beta_n}, C]$ , for some  $\eta > 0, C > 1$ , such that infinitely often,  $|\tilde{\lambda} - \bar{\lambda}_i| > n^{-\ell}\sqrt{\beta_n}$  for all  $i \le i_0$ . Let  $\gamma_n$  denote a sequence of circular contours of radius  $n^{-\ell}\sqrt{\beta_n}$  centered at  $\tilde{\lambda}$ . Then, eventually,

$$\frac{1}{2\pi i} \int_{\gamma_n} \frac{\frac{d}{dz} \det \widetilde{M}_n(z)}{\det \widetilde{M}_n(z)} dz = 1, \qquad \frac{1}{2\pi i} \int_{\gamma_n} \frac{\frac{d}{dz} \det \overline{M}_n(z)}{\det \overline{M}_n(z)} dz = 0.$$
(50)

Note that on  $\gamma_n$ , (44)–(47) hold while (48) becomes  $\beta_n^{-r/2}$ . Thus, similarly to (49), the left integral in (50) converges to the right integral, a contradiction. Furthermore, eventually  $\tilde{S}_n$  has no eigenvalues greater than *C*:

$$\frac{1}{\sqrt{m}}\|\widetilde{X}_n\|_2 \leq \sum_{i=1}^r \theta_i \|u_i\|_2 \|v_i\|_2 + \frac{1}{\sqrt{m}} \|X_n\|_2 \leq C,$$

the last inequality holding almost surely eventually. This establishes Claim (iv). We conclude that almost surely eventually, the leading eigenvalues of  $\tilde{S}_n$  satisfy

$$|\lambda_i - \lambda_i| \leq n^{-\ell} \sqrt{eta_n}$$
 ,  $i \leq i_0$ 

which is (5) in the case  $\varepsilon_{i,n} = 0$ .

For general signal strengths  $\theta_i = \tau \beta_n^{1/4} (1 + \varepsilon_{i,n})$ , the above argument may be used to show that  $\tilde{\lambda}_i$  localizes about  $\overline{D}_n^{(-1)}(\theta_i^{-2})$  and that Claim (iv) holds. Recalling that  $\overline{D}_n^{(-1)}(t) = (t+1)(\beta_n t+1)/t$  (Lemma 2), for  $i \le i_0$ ,

$$\begin{split} \left| \overline{D}_{n}^{(-1)}(\theta_{i}^{-2}) - \overline{\lambda}_{i} \right| &= \left| \overline{D}_{n}^{(-1)}(\theta_{i}^{-2}) - \overline{D}_{n}^{(-1)} \left( (\tau_{i}\beta_{n}^{1/4})^{-2} \right) \right| = (\tau_{i}^{2} + \tau_{i}^{-2}(1 + \varepsilon_{i,n})^{-2}) |\varepsilon_{i,n}| (2 + \varepsilon_{i,n}) \sqrt{\beta_{n}} \\ &\lesssim |\varepsilon_{i,n}| \sqrt{\beta_{n}} \,. \end{split}$$
(51)

Thus, almost surely eventually,

$$|\tilde{\lambda}_i - \bar{\lambda}_i| \le \left|\tilde{\lambda}_i - \overline{D}_n^{(-1)}(\theta_i^{-2})\right| + \left|\overline{D}_n^{(-1)}(\theta_i^{-2}) - \bar{\lambda}_i\right| \le (n^{-\ell} + |\varepsilon_{i,n}|)\sqrt{\beta_n}$$

establishing (5). (2) is a consequence of (5) and Claim (iv).

#### 4. Proof of singular vector results: (3), (6), and (7)

We prove (6) and (7), from which (3) immediately follows. For notational lightness, this section assumes  $\varepsilon_{i,n} = 0$ . All bounds are understood to hold only almost surely.

**Lemma 10.** Let  $i \leq i_0$ ,  $j \leq r$ , and  $\ell < 1/4$ . Let  $\tilde{u}_i$ ,  $\tilde{v}_i$  be the singular vectors corresponding to  $\tilde{\sigma}_i = \sqrt{\tilde{\lambda}_i}$ . Almost surely,

$$|\bar{\sigma}_i \bar{s}_n(\bar{\lambda}_i) \theta_j \langle \tilde{v}_i, v_j \rangle + \langle \tilde{u}_i, u_j \rangle| \lesssim n^{-\ell} (\beta_n^{-1/4} \| V_n^\top \tilde{v}_i \|_1 + \beta_n^{1/4})$$
(52)

and for  $i \neq j$ ,

$$|\langle \tilde{v}_i, v_j \rangle| \lesssim n^{-\ell} (|\langle \tilde{v}_i, v_i \rangle| + \sqrt{\beta_n}). \quad \Box$$
(53)

Proof. Recall Lemma 3: the column vector

$$\begin{bmatrix} \Theta_n V_n^\top \tilde{v}_i \\ \Theta_n U_n^\top \tilde{u}_i \end{bmatrix}$$
(54)

belongs to the kernel of  $\widetilde{M}_n(\tilde{\lambda}_i)$ . As (54) is orthogonal to the *j*th row of  $\widetilde{M}_n(\tilde{\lambda}_i)$ ,

$$\sum_{k=1}^{\prime} \tilde{\sigma}_i u_j^{\top} (S_n - \tilde{\lambda}_i I_n)^{-1} u_k \theta_k \langle \tilde{v}_i, v_k \rangle + \sum_{k=1}^{\prime} \left( \theta_k^{-1} \mathbf{1}_{\{j=k\}} + \frac{1}{\sqrt{m}} u_j^{\top} (S_n - \tilde{\lambda}_i I_n)^{-1} X_n v_k \right) \theta_k \langle \tilde{u}_i, u_k \rangle = 0.$$

Isolating the *j*th term, by Lemma 6,

$$\begin{split} |\tilde{\sigma}_{i}u_{j}^{\top}(S_{n}-\tilde{\lambda}_{i}I_{n})^{-1}u_{j}\theta_{j}\langle\tilde{v}_{i},v_{j}\rangle+\langle\tilde{u}_{i},u_{j}\rangle| &\leq \tilde{\sigma}_{i}\sum_{k\neq j}\theta_{k}|u_{j}^{\top}(S_{n}-\tilde{\lambda}_{i}I_{n})^{-1}u_{k}||\langle\tilde{v}_{i},v_{k}\rangle|+\frac{1}{\sqrt{m}}\sum_{k=1}^{r}\theta_{k}|u_{j}^{\top}(S_{n}-\tilde{\lambda}_{i}I_{n})^{-1}X_{n}v_{k}|\\ &\lesssim n^{-\ell}\beta_{n}^{-1/4}\sum_{k\neq j}|\langle\tilde{v}_{i},v_{k}\rangle|+n^{-\ell}\beta_{n}^{1/4} \end{split}$$
(55)

$$\begin{split} |\tilde{\sigma}_{i}s_{n}(\tilde{\lambda}_{i})\theta_{j}\langle\tilde{v}_{i},v_{j}\rangle + \langle\tilde{u}_{i},u_{j}\rangle| &\leq |\tilde{\sigma}_{i}u_{j}^{\top}(S_{n}-\tilde{\lambda}_{i}I_{n})^{-1}u_{j}\theta_{j}\langle\tilde{v}_{i},v_{j}\rangle + \langle\tilde{u}_{i},u_{j}\rangle| + \tilde{\sigma}_{i}\theta_{j}|u_{j}^{\top}(S_{n}-\tilde{\lambda}_{i}I_{n})^{-1}u_{j}-s_{n}(\tilde{\lambda}_{i})||\langle\tilde{v}_{i},v_{j}\rangle| \\ &\leq n^{-\ell}\beta_{n}^{-1/4}\sum_{j=1}^{r}|\langle\tilde{v}_{i},v_{k}\rangle| + n^{-\ell}\beta_{n}^{1/4}. \end{split}$$
(56)

$$\gtrsim n \quad p_n \quad \sum_{k=1}^{n} |\langle v_i, v_k \rangle| + n \quad p_n' \quad .$$

Note that eventually,  $|s_n(\tilde{\lambda}_i) - s_n(\bar{\lambda}_i)| \le \frac{|\tilde{\lambda}_i - \bar{\lambda}_i|}{|\tilde{\lambda}_i - \lambda_1||\tilde{\lambda}_i - \lambda_1|}$ . By an application of (5) and Theorem 3,

$$\begin{aligned} |\tilde{\sigma}_{i}s_{n}(\tilde{\lambda}_{i}) - \bar{\sigma}_{i}\bar{s}_{n}(\bar{\lambda}_{i})| &\leq |\tilde{\sigma}_{i} - \bar{\sigma}_{i}||s_{n}(\tilde{\lambda}_{i})| + |\bar{\sigma}_{i}||s_{n}(\tilde{\lambda}_{i}) - s_{n}(\bar{\lambda}_{i})| + |\bar{\sigma}_{i}||s_{n}(\bar{\lambda}_{i}) - \bar{s}_{n}(\bar{\lambda}_{i})| \\ &\leq n^{-\ell} + n^{-\ell}\beta_{n}^{-1/2} + n^{-\ell}\beta_{n}^{-1/2} . \end{aligned}$$

$$(57)$$

(52) follows from (56) and (57). By a similar argument for row j + r of  $\widetilde{M}_n(\tilde{\lambda}_i)$ ,

$$\left|\langle \tilde{v}_i, v_j \rangle + \bar{\sigma}_i \left( \beta_n \bar{s}_n(\bar{\lambda}_i) - \frac{1 - \beta_n}{\bar{\lambda}_i} \right) \theta_j \langle \tilde{u}_i, u_j \rangle \right| \lesssim n^{-\ell} (\beta_n^{1/4} \| V_n^\top \tilde{v}_i \|_1 + \beta_n^{3/4}).$$
(58)

Substitution of (52) into (58) yields

$$\left|\langle \tilde{v}_i, v_j \rangle - \bar{\lambda}_i \bar{s}_n(\bar{\lambda}_i) \left( \beta_n \bar{s}_n(\bar{\lambda}_i) - \frac{1 - \beta_n}{\bar{\lambda}_i} \right) \theta_j^2 \langle \tilde{v}_i, v_j \rangle \right| \lesssim n^{-\ell} (\|V_n^\top \tilde{v}_i\|_1 + \sqrt{\beta_n})$$

Recall that  $\overline{D}_n(z) = \beta_n z \overline{s}_n(z)^2 - (1 - \beta_n) \overline{s}_n(z)$  and  $\overline{\lambda}_i$  is defined by  $\overline{D}_n(\overline{\lambda}_i) = \theta_i^{-2}$ . Thus, for  $j \neq i$ ,

 $|\langle \tilde{v}_i, v_j \rangle| \lesssim n^{-\ell} (\|V_n^\top \tilde{v}_i\|_1 + \sqrt{\beta_n}) .$ 

Rewriting this as  $(1 - Cn^{-\ell})(\|V_n^{\top} \tilde{v}_i\|_1 - |\langle \tilde{v}_i, v_i \rangle|) \lesssim n^{-\ell}(|\langle \tilde{v}_i, v_i \rangle| + \sqrt{\beta_n})$ , we recover (53):

$$|\langle \tilde{v}_i, v_j \rangle| \lesssim n^{-\ell} (|\langle \tilde{v}_i, v_i \rangle| + \sqrt{\beta_n}).$$

**Proof of (6) and (7).** Let  $i \le i_0$  and  $\ell < 1/8$ . By (15),  $a_n + b_n + 2c_n = 1$  with

$$a_{n} = \tilde{\lambda}_{i} \tilde{v}_{i}^{\top} P_{n}^{\top} (S_{n} - \tilde{\lambda}_{i})^{-2} P_{n} \tilde{v}_{i} = \tilde{\lambda}_{i} \sum_{j,k=1}^{r} \theta_{j} \theta_{k} \langle \tilde{v}_{i}, v_{j} \rangle \langle \tilde{v}_{i}, v_{k} \rangle u_{j}^{\top} (S_{n} - \tilde{\lambda}_{i} I_{n})^{-2} u_{k} ,$$

$$b_{n} = \frac{1}{m} \tilde{u}_{i}^{\top} P_{n} X_{n}^{\top} (S_{n} - \tilde{\lambda} I_{n})^{-2} X_{n} P_{n}^{\top} \tilde{u}_{i} = \frac{1}{m} \sum_{j,k=1}^{r} \theta_{j} \theta_{k} \langle \tilde{u}_{i}, u_{j} \rangle \langle \tilde{u}_{i}, u_{k} \rangle v_{j}^{\top} X_{n}^{\top} (S_{n} - \tilde{\lambda}_{i} I_{n})^{-2} X_{n} v_{k} ,$$

$$c_{n} = \frac{\tilde{\sigma}_{i}}{\sqrt{m}} \tilde{u}_{i}^{\top} P_{n} X_{n}^{\top} (S_{n} - \tilde{\lambda} I_{n})^{-2} P_{n} \tilde{v}_{i} = \frac{\tilde{\sigma}_{i}}{\sqrt{m}} \sum_{j,k=1}^{r} \theta_{j} \theta_{k} \langle \tilde{u}_{i}, u_{j} \rangle \langle \tilde{v}_{i}, v_{k} \rangle v_{j}^{\top} X_{n}^{\top} (S_{n} - \tilde{\lambda}_{i} I_{n})^{-2} u_{k} .$$
(59)

By (5) and Corollary 1,

$$\begin{aligned} \left| \tilde{\lambda}_i \frac{d}{dz} s_n(\tilde{\lambda}_i) - \bar{\lambda}_i \frac{d}{dz} \bar{s}_n(\bar{\lambda}_i) \right| &\leq \left| \tilde{\lambda}_i - \bar{\lambda}_i \right| \left| \frac{d}{dz} s_n(\tilde{\lambda}_i) \right| + \left| \bar{\lambda}_i \right| \left| \frac{d}{dz} (s_n(\tilde{\lambda}_i) - s_n(\bar{\lambda}_i)) \right| + \left| \bar{\lambda}_i \right| \left| \frac{d}{dz} (s_n(\bar{\lambda}_i) - \bar{s}_n(\bar{\lambda}_i)) \right| \\ &\lesssim n^{-2\ell} \beta_n^{-1/2} + n^{-2\ell} \beta_n^{-1} + n^{-\ell} \beta_n^{-1}. \end{aligned}$$

Together with (30) of Lemma 7,

$$a_{n} = \sum_{j,k=1}^{r} \theta_{j} \theta_{k} \langle \tilde{v}_{i}, v_{j} \rangle \langle \tilde{v}_{i}, v_{k} \rangle \Big( \delta_{jk} \bar{\lambda}_{i} \frac{d}{dz} \bar{s}_{n}(\bar{\lambda}_{i}) + O(n^{-\ell} \beta_{n}^{-1}) \Big)$$
  
$$= \sum_{j=1}^{r} \theta_{j}^{2} |\langle \tilde{v}_{i}, v_{j} \rangle|^{2} \bar{\lambda}_{i} \frac{d}{dz} \bar{s}_{n}(\bar{\lambda}_{i}) + O(n^{-\ell} \beta_{n}^{-1/2} \|V_{n}^{\top} \tilde{v}_{i}\|_{1}^{2}).$$
(60)

Similarly, using

$$\frac{1}{m}\operatorname{tr}(S_n-\tilde{\lambda}_i)^{-2}S_n = \frac{1}{m}\operatorname{tr}\left((S_n-\tilde{\lambda}_iI_n)^{-1}+\tilde{\lambda}_i(S_n-\tilde{\lambda}_iI_n)^{-2}\right) = \beta_n\left(s_n(\tilde{\lambda}_i)+\tilde{\lambda}_i\frac{d}{dz}s_n(\tilde{\lambda}_i)\right),$$

(31), (33), and (34), it may be shown that

...

$$b_{n} = \sum_{j,k=1}^{r} \theta_{j} \theta_{k} \langle \tilde{u}_{i}, u_{j} \rangle \langle \tilde{u}_{i}, u_{k} \rangle \Big( \delta_{jk} \beta_{n} \Big( \bar{s}_{n}(\bar{\lambda}_{i}) + \bar{\lambda}_{i} \frac{d}{dz} \bar{s}_{n}(\bar{\lambda}_{i}) \Big) + O(n^{-\ell}) \Big)$$

$$= \sum_{j=1}^{r} \theta_{j}^{2} |\langle \tilde{u}_{i}, u_{j} \rangle|^{2} \beta_{n} \Big( \bar{s}_{n}(\bar{\lambda}_{i}) + \bar{\lambda}_{i} \frac{d}{dz} \bar{s}_{n}(\bar{\lambda}_{i}) \Big) + O(n^{-\ell} \sqrt{\beta_{n}})$$

$$|c_{n}| \lesssim n^{-\ell} \beta_{n}^{-1/2} \sum_{j,k=1}^{r} \theta_{j} \theta_{k} |\langle \tilde{u}_{i}, u_{j} \rangle \langle \tilde{v}_{i}, v_{k} \rangle| \lesssim n^{-\ell}.$$
(61)

Thus,

$$\sum_{j=1}^{r} \theta_{j}^{2} |\langle \tilde{v}_{i}, v_{j} \rangle|^{2} \bar{\lambda}_{i} \frac{d}{dz} \bar{s}_{n}(\bar{\lambda}_{i}) + \sum_{j=1}^{r} \theta_{j}^{2} |\langle \tilde{u}_{i}, u_{j} \rangle|^{2} \beta_{n} \Big( \bar{s}_{n}(\bar{\lambda}_{i}) + \bar{\lambda}_{i} \frac{d}{dz} \bar{s}_{n}(\bar{\lambda}_{i}) \Big) = 1 + O \Big( n^{-\ell} (1 + \beta_{n}^{-1/2} \|V_{n}^{\top} \tilde{v}_{i}\|_{1}^{2}) \Big).$$
(62)

Terms on the left-hand side with  $j \neq i$  are negligible: by (53),

$$\begin{split} \left| \sum_{j \neq i} \theta_j^2 |\langle \tilde{v}_i, v_j \rangle|^2 \bar{\lambda}_i \frac{d}{dz} \bar{s}_n(\bar{\lambda}_i) \right| &\lesssim n^{-4\ell} (\beta_n^{-1/2} |\langle \tilde{v}_i, v_i \rangle|^2 + \sqrt{\beta_n}) \\ \left| \sum_{j \neq i} \theta_j^2 |\langle \tilde{u}_i, u_j \rangle|^2 \beta_n \Big( \bar{s}_n(\bar{\lambda}_i) + \bar{\lambda}_i \frac{d}{dz} \bar{s}_n(\bar{\lambda}_i) \Big) \right| &\lesssim \sqrt{\beta_n} \end{split}$$

implying

$$\theta_i^2 |\langle \tilde{v}_i, v_i \rangle|^2 \bar{\lambda}_i \frac{d}{dz} \bar{s}_n(\bar{\lambda}_i) + \theta_i^2 |\langle \tilde{u}_i, u_i \rangle|^2 \beta_n \left( \bar{s}_n(\bar{\lambda}_i) + \bar{\lambda}_i \frac{d}{dz} \bar{s}_n(\bar{\lambda}_i) \right) = 1 + O\left( n^{-\ell} (1 + \beta_n^{-1/2} \|V_n^\top \tilde{v}_i\|_1^2) + \sqrt{\beta_n} \right).$$

$$\tag{63}$$

Substitution of (52) into (63) yields

$$\begin{aligned} |\langle \tilde{v}_{i}, v_{i} \rangle|^{2} \left( \theta_{i}^{2} \bar{\lambda}_{i} \frac{d}{dz} \bar{s}_{n}(\bar{\lambda}_{i}) + \theta_{i}^{4} \bar{\lambda}_{i} \bar{s}_{n}(\bar{\lambda}_{i})^{2} \beta_{n} \left( \bar{s}_{n}(\bar{\lambda}_{i}) + \bar{\lambda}_{i} \frac{d}{dz} \bar{s}_{n}(\bar{\lambda}_{i}) \right) \right) \\ &= |\langle \tilde{v}_{i}, v_{i} \rangle|^{2} \left( \theta_{i}^{2} \int \frac{\bar{\lambda}_{i}}{(t - \bar{\lambda}_{i})^{2}} d\overline{F}_{n}(t) + \theta_{i}^{4} \bar{\lambda}_{i} \bar{s}_{n}(\bar{\lambda}_{i})^{2} \beta_{n} \int \frac{t}{(t - \bar{\lambda}_{i})^{2}} d\overline{F}_{n}(t) \right) \\ &= 1 + O \left( n^{-\ell} (1 + \beta_{n}^{-1/2} \| V_{n}^{\top} \tilde{v}_{i} \|_{1}^{2}) + \sqrt{\beta_{n}} \right). \end{aligned}$$

$$\tag{64}$$

(64) is included to facilitate comparison with Theorem 2.10 of [9]. In the regime where  $\beta_n \rightarrow \beta > 0$  and  $\theta_i > 1$  is constant, the term multiplying  $|\langle \tilde{v}_i, v_i \rangle|^2$  converges to a positive limit, and  $|\langle \tilde{v}_i, v_i \rangle|^2$  converges to the limit's inverse. Here, since

$$\left|\theta_i^2|\langle \tilde{u}_i, u_i\rangle|^2 \beta_n \Big(\bar{s}_n(\bar{\lambda}_i) + \bar{\lambda}_i \frac{d}{dz} \bar{s}_n(\bar{\lambda}_i)\Big)\right| \lesssim \sqrt{\beta_n} \,,$$

(53) and (63) imply

$$\theta_{i}^{2} |\langle \tilde{v}_{i}, v_{i} \rangle|^{2} \bar{\lambda}_{i} \frac{d}{dz} \bar{s}_{n}(\bar{\lambda}_{i}) = 1 + O\left(n^{-\ell} (1 + \beta_{n}^{-1/2} \|V_{n}^{\top} \tilde{v}_{i}\|_{1}^{2}) + \sqrt{\beta_{n}}\right).$$
(65)

Evaluated at  $z = \bar{\lambda}_i$ , Lemma 5 is tight up to constants:  $\bar{\lambda}_i \frac{d}{dz} \bar{s}_n(\bar{\lambda}_i) \simeq \beta_n^{-1}$ . Thus, for  $j \neq i$ ,

$$(1 - Cn^{-\ell})|\langle \tilde{v}_i, v_i \rangle|^2 \lesssim \sqrt{\beta_n}, \qquad |\langle \tilde{v}_i, v_i \rangle| \lesssim \beta_n^{1/4}, \qquad |\langle \tilde{v}_i, v_j \rangle| \lesssim n^{-2\ell} \beta_n^{1/4}.$$
(66)

The third inequality follows from (53). We have established (7): right singular vectors asymptotically do not correlate with right signal vectors.

Now, by (52), (65), and (66), we obtain

$$\left|\langle \tilde{u}_i, u_i \rangle\right|^2 = (\bar{\sigma}_i \bar{s}_n(\bar{\lambda}_i) \theta_i \langle \tilde{v}_i, v_i \rangle)^2 + O(n^{-4\ell}) = \frac{\bar{s}_n(\bar{\lambda}_i)^2}{\frac{d}{dz} \bar{s}_n(\bar{\lambda}_i)} + O(n^{-\ell} + \sqrt{\beta_n}).$$
(67)

Direct calculation yields

$$\bar{s}_n(\bar{\lambda}_i) = \frac{-(\tau_i^2 + \tau_i^{-2}) + \sqrt{(\tau_i^2 + \tau_i^{-2})^2 - 4}}{2\sqrt{\beta_n}} + O(1)$$
$$\frac{d}{dz}\bar{s}_n(\bar{\lambda}_i) = -\frac{1}{2\beta_n\bar{\lambda}_i} \left(1 - \frac{\tau_i^2 + \tau_i^{-2}}{\sqrt{(\tau_i^2 + \tau_i^{-2})^2 - 4}}\right) + O(\beta_n^{-1/2})$$

from which we conclude (6):

$$|\langle \tilde{u}_i, u_i \rangle|^2 = 1 - \tau_i^4 + O(n^{-\ell} + \sqrt{\beta_n}).$$
(68)

Therefore,  $|\langle \tilde{u}_i, u_i \rangle|^2 \xrightarrow{a.s.} 1 - \tau_i^{-4}$ . For  $j \neq i$ ,  $|\langle \tilde{u}_i, u_j \rangle| \leq n^{-2\ell}$  follows from (52).  $\Box$ 

#### 5. Convergence rate of the Stieltjes transform

This section develops bounds on  $|s_n(z) - \bar{s}_n(z)|$  and  $|\frac{d}{dz}(s_n(z) - \bar{s}_n(z))|$  based on the work of Zhidong Bai, Jack Silverstein, Jiang Hu, and Wang Zhou, in particular, Section 4.1 of [2] and Section 8 of [4].

**Theorem 3.** Let  $\eta > 0$  and  $u_n \ge 1 + (2 + \eta)\sqrt{\beta_n}$  be a bounded sequence. For any  $\ell < 1/4$ , almost surely,

$$\sup_{|z-u_n| \le n^{-1/4} \sqrt{\beta_n}} |s_n(z) - \bar{s}_n(z)| \lesssim n^{-\ell} \beta_n^{-1/2} \,. \tag{69}$$

**Corollary 1.** Let  $\eta > 0$  and  $u_n \ge 1 + (2 + \eta)\sqrt{\beta_n}$  be a bounded sequence. For any  $\ell < 1/8$ , almost surely,

$$\sup_{|z-u_n|\le n^{-1/8}\sqrt{\beta_n}} \left| \frac{d}{dz} \left( s_n(z) - \bar{s}_n(z) \right) \right| \lesssim n^{-\ell} \beta_n^{-1}.$$
(70)

These bounds are novel in that (1) dependence on  $\beta_n$  is more carefully tracked and (2) the bound applies to  $z \ge 1 + (2 + \eta)\sqrt{\beta_n}$  (rather than to z with  $\Im(z) > 1/\sqrt{m}$ ). (2) is critical since the proof of Theorem 2 requires bounding  $|s_n(z) - \bar{s}_n(z)|$  and  $|\frac{d}{dz}(s_n(z) - \bar{s}_n(z))|$  along a contour centered on the real line. Furthermore, (2) above enables improvement of previous bounds using the concentration of the maximum eigenvalue of  $S_n$ . For example, in place of deterministic bounds such as

$$\frac{1}{n}\operatorname{tr}(S_n - zI_n)^{-1}(S_n - \bar{z}I_n)^{-1} \le \Im(z)^{-2}$$

we will argue using Lemma 15 that  $\frac{1}{n}$ tr( $S_n - zI_n$ )<sup>-1</sup>( $S_n - \bar{z}I_n$ )<sup>-1</sup>  $\leq \beta_n^{-1}$  with high probability. As set forth in Section 2.2, the entries of  $X_n$  are assumed truncated and normalized as in Lemma 15. For sake of completeness, we note that Theorem 3 and Corollary 1 hold for non-truncated noise as well; the effects of truncation and normalization may be shown to be negligible using Lemma 16.

Let  $x_k^{\top}$  denote the *k*th-row of  $X_n$ ,  $X_{nk}$  consist of the remaining n - 1 rows, and  $E_k(\cdot)$  denote the conditional expectation given  $\{x_{ij}, i \leq k, j \leq m\}$ . Additionally, we define

$$s_{kk} = \frac{1}{m} \sum_{j=1}^{m} |x_{kj}|^{2} \qquad S_{nk} = \frac{1}{m} X_{nk} X_{nk}^{\top},$$

$$\alpha_{k} = X_{nk} x_{k} \qquad \sigma_{k} = \text{tr}(S_{n} - zI_{n})^{-1} - \text{tr}(S_{nk} - zI_{n-1})^{-1},$$

$$\gamma_{k} = (E_{k} - E_{k-1})\sigma_{k} \qquad b_{k} = \frac{1}{1 - z - \beta_{n} - \frac{z}{m} \text{tr}(S_{nk} - zI_{n-1})^{-1}},$$

$$\bar{b}_{k} = \frac{1}{s_{kk} - z - \frac{1}{m^{2}} \alpha_{k}^{\top}(S_{nk} - zI_{n-1})^{-1} \alpha_{k}},$$

$$\bar{b}_{n} = \frac{1}{1 - z - \beta_{n} - \beta_{n} z E_{n}(z)},$$

$$\bar{c}_{k} = s_{kk} - 1 - \frac{1}{m^{2}} \alpha_{k}^{\top}(S_{nk} - zI_{n-1})^{-1} \alpha_{k} + \frac{1}{m} \text{tr}(S_{nk} - zI_{n-1})^{-1} \alpha_{k} + \frac{1}{m},$$

$$\bar{c}_{k} = s_{kk} - 1 + \beta_{n} + \beta_{n} z E_{n}(z) - \frac{1}{m^{2}} \alpha_{k}^{\top}(S_{nk} - zI_{n-1})^{-1} \alpha_{k}.$$
(71)

Standard properties of these definitions are stated below (see Theorem A.5, Lemma 8.18, and (8.4.19) of [4]).

**Lemma 11.** *For*  $z \in \mathbb{C}^+$ *,* 

$$\sigma_k = \frac{1 + \frac{1}{m^2} \alpha_k^\top (S_{nk} - zI_{n-1})^{-2} \alpha_k}{S_{kk} - z - \frac{1}{m^2} \alpha_k^\top (S_{nk} - zI_{n-1})^{-1} \alpha_k} \qquad |\sigma_k| \le \frac{1}{\Im(z)} \qquad |\tilde{b}_k| \le \frac{1}{\Im(z)} \qquad |\tilde{b}_n| \le \frac{1}{\sqrt{\beta_n |z|}}.$$

Furthermore, for  $\Im(z) \geq \frac{1}{\sqrt{m}}$ ,  $|b_k| \lesssim (\beta_n |z|)^{-1/2}$ .

In the following lemmas, we will use  $u_n$  and  $v_n$  to refer to the real and imaginary parts of a complex sequence  $z_n$ . The subscripts will be suppressed for notational lightness. Lemmas 12 and 13 respectively bound  $|s_n(z) - E\bar{s}_n(z)|$  and  $|Es_n(z) - \bar{s}_n(z)|$  for z with imaginary part at least  $1/\sqrt{m}$ . The proof of Theorem 3 extends the bound to the real axis.

**Lemma 12.** Let z = u + iv be a bounded sequence where  $u \ge 1 + (2 + \eta)\sqrt{\beta_n}$  and  $v \ge 1/\sqrt{m}$ . Then,

$$|\mathsf{E}|s_n(z)-\mathsf{E}s_n(z)|^2 \lesssim \frac{1}{n^2}\Big(\frac{1}{\beta_n}+\frac{1}{v^2}\Big).$$

**Proof.** Begin with a standard decomposition of  $s_n(z) - Es_n(z)$ :

$$s_n(z) - \mathrm{E}s_n(z) = \frac{1}{n} \sum_{k=1}^n (\mathrm{E}_k - \mathrm{E}_{k-1}) \big( \mathrm{tr}(S_n - zI_n)^{-1} - \mathrm{tr}(S_{nk} - zI_{n-1})^{-1} \big)$$
  
=  $\frac{1}{n} \sum_{k=1}^n \gamma_k$ .

As  $\{\gamma_k\}_{k=1}^n$  forms a martingale difference sequence, the Burkholder inequality (Lemma 2.12 of [4]) yields

$$\mathrm{E}|s_n(z)-\mathrm{E}s_n(z)|^2\lesssim rac{1}{n^2}\sum_{k=1}^n\mathrm{E}|\gamma_k|^2\,.$$

This martingale decomposition, together with the bound  $|\gamma_k| \leq 2/v$ , appear in proofs of convergence to the Marchenko–Pastur distribution as  $\beta_n \rightarrow \beta > 0$  [1,5]. Such proofs consider *z* fixed. Here and in [2,4], as *v* may decay rapidly, a tighter bound on the moments of  $\gamma_k$  is needed. Using the identity

$$S_{nk}(S_{nk} - zI_{n-1})^{-1} = (S_{nk} - zI_{n-1})^{-1}S_{nk} = I_{n-1} + z(S_{nk} - zI_{n-1})^{-1},$$
(72)

we obtain

$$\tilde{b}_k^{-1} - b_k^{-1} = \epsilon_k \qquad \tilde{b}_k = \frac{b_k}{1 + b_k \epsilon_k}.$$

Thus,

$$\sigma_{k} = \frac{b_{k}}{1 + b_{k}\epsilon_{k}} \left( 1 + \frac{1}{m^{2}}\alpha_{k}^{\top}(S_{nk} - zI_{n-1})^{-2}\alpha_{k} \right) = -\sigma_{k}b_{k}\epsilon_{k} + b_{k} \left( 1 + \frac{1}{m^{2}}\alpha_{k}^{\top}(S_{nk} - zI_{n-1})^{-2}\alpha_{k} \right)$$

and

$$\gamma_{k} = -(E_{k} - E_{k-1})\sigma_{k}b_{k}\epsilon_{k} + \frac{1}{m^{2}}(E_{k} - E_{k-1})b_{k}\alpha_{k}^{\top}(S_{nk} - zI_{n-1})^{-2}\alpha_{k}$$
  
=  $-(E_{k} - E_{k-1})\sigma_{k}b_{k}\epsilon_{k} + \frac{1}{m^{2}}E_{k}b_{k}\left[\alpha_{k}^{\top}(S_{nk} - zI_{n-1})^{-2}\alpha_{k} - \operatorname{tr}(X_{nk}^{\top}(S_{nk} - zI_{n-1})^{-2}X_{nk})\right].$  (73)

Here, we have used that  $b_k$  does not depend on  $x_k$ , so  $(E_k - E_{k-1})b_k = 0$ . By Lemmas 11 and 17,

$$\frac{1}{m^4} \operatorname{Eb}_k^2 \left| \alpha_k^\top (S_{nk} - zI_{n-1})^{-2} \alpha_k - \operatorname{tr}(X_{nk}^\top (S_{nk} - zI_{n-1})^{-2} X_{nk}) \right|^2 \lesssim \frac{1}{m^4 \beta_n} \operatorname{Etr} X_{nk}^\top (S_{nk} - zI_{n-1})^{-2} X_{nk} (X_{nk}^\top (S_{nk} - \bar{z}I_{n-1})^{-2} X_{nk})^\top \\
= \frac{1}{nm} \operatorname{Etr} S_{nk} (S_{nk} - zI_{n-1})^{-2} S_{nk} (S_{nk} - \bar{z}I_{n-1})^{-2} = \frac{1}{m} \operatorname{E} \left( \frac{1}{n} \sum_{i=1}^{n-1} \frac{\lambda_{k,i}^2}{|\lambda_{k,i} - z|^4} \right),$$
(74)

where  $\lambda_{k,1} \geq \cdots \geq \lambda_{k,n-1}$  denote the eigenvalues of  $S_{nk}$ . By (16) and a union bound,

$$\Pr(\lambda_{k,1} \ge 1 + (2 + \eta/2)\sqrt{\beta_n} \text{ for some } k) \lesssim \frac{1}{n}.$$

On the complement of this event, the integrand of (74) is bounded by  $16|z|/(\eta\sqrt{\beta_n})^4$ . In addition,

$$rac{\lambda_{k,i}^2}{|\lambda_{k,i}-z|^2} \leq 1 + rac{u^2}{v^2} \qquad rac{\lambda_{k,i}^2}{|\lambda_{k,i}-z|^4} \leq rac{|z|^2}{v^4}.$$

Therefore,

$$\frac{1}{m^4} E b_k^2 \left| \alpha_k^\top (S_{nk} - zI_{n-1})^{-2} \alpha_k - \operatorname{tr}(X_{nk}^\top (S_{nk} - zI_{n-1})^{-2} X_{nk}) \right|^2 \lesssim \frac{1}{m} \left( \frac{1}{\beta_n^2} + \frac{1}{nv^4} \right).$$
(75)

Note that [4] bounds terms such as (74) deterministically. Similarly,

$$E|\epsilon_{k}|^{2} \lesssim E|s_{kk} - 1|^{2} + \frac{1}{m^{2}} + \frac{1}{m^{4}} E|\alpha_{k}^{\top}(S_{nk} - zI_{n-1})^{-1}\alpha_{k} - \operatorname{tr}(X_{nk}^{\top}(S_{nk} - zI_{n-1})^{-1}X_{nk})|^{2} \\ \lesssim \frac{1}{m} + \frac{1}{m^{2}} \operatorname{Etr} S_{nk}(S_{nk} - zI_{n-1})^{-1}S_{nk}(S_{nk} - \bar{z}I_{n-1})^{-1} \lesssim \frac{1}{m} \left(1 + \frac{1}{mv^{2}}\right).$$

$$(76)$$

By Lemma 11 and (76),

$$\mathsf{E}|\sigma_k b_k \epsilon_k|^2 \lesssim \frac{1}{v^2 \beta_n} \mathsf{E}|\epsilon_k|^2 \lesssim \frac{1}{nv^2} \,. \tag{77}$$

Thus, using (73)–(77) and  $v \ge 1/\sqrt{m}$ ,

$$\mathrm{E}|s_n(z)-\mathrm{E}s_n(z)|^2 \lesssim \frac{1}{n^2}\sum_{k=1}^n \mathrm{E}|\gamma_k|^2 \lesssim \frac{1}{n^2}\Big(\frac{1}{\beta_n}+\frac{1}{v^2}\Big). \quad \Box$$

**Lemma 13.** Let z = u + iv be a bounded sequence where  $u \ge 1 + (2 + \eta)\sqrt{\beta_n}$  and  $v \ge 1/\sqrt{m}$ . Then,

$$|\mathrm{E}s_n(z)-\bar{s}_n(z)|\lesssim \frac{1}{nv}.$$

**Proof.** Recall definitions (71). We have

$$\tilde{b}_k^{-1} - \bar{b}_n^{-1} = \bar{\epsilon}_k \qquad \tilde{b}_k - \bar{b}_n = -\bar{b}_n^2 \bar{\epsilon}_k + \bar{b}_n^2 \tilde{b}_k \bar{\epsilon}_k^2$$

Together with Theorem A.4 of [4], this leads to the following decomposition of  $s_n(z)$ :

$$s_n(z) = \frac{1}{n} \sum_{k=1}^n \tilde{b}_k = \bar{b}_n + \frac{1}{n} \sum_{k=1}^n (-\bar{b}_n^2 \bar{\epsilon}_k + \bar{b}_n^2 \tilde{b}_k \bar{\epsilon}_k^2).$$
(78)

(8.3.12) of [4] proves the moment bound  $|E\bar{\epsilon}_k| \leq 1/(mv)$ . The second moment of  $\bar{\epsilon}_k$  is bounded as follows:

$$|\bar{\epsilon}_k|^2 \lesssim |\bar{\epsilon}_k - |\bar{\epsilon}_k| |X_{nk}|^2 + ||\bar{\epsilon}_k| |X_{nk}| - ||\bar{\epsilon}_k|^2 + ||\bar{\epsilon}_k|^2$$

$$\tag{79}$$

A bound on  $E|\bar{\epsilon}_k - E(\bar{\epsilon}_k|X_{nk})|^2 = E|\epsilon_k - m^{-1}|^2 \le 2(E|\epsilon_k|^2 + m^{-2})$  is given by (76). By (72) and Lemma 11,

$$\begin{aligned} \mathsf{E}|\mathsf{E}(\bar{\epsilon}_{k}|X_{nk}) - \mathsf{E}\bar{\epsilon}_{k}|^{2} &= \frac{1}{m^{2}}\mathsf{E}\big|\mathsf{tr}(S_{nk} - zI_{n-1})^{-1}S_{nk} - \mathsf{E}\mathsf{tr}(S_{nk} - zI_{n-1})^{-1}S_{nk}\big|^{2} &= \frac{|z|^{2}}{m^{2}}\mathsf{E}\big|\mathsf{tr}(S_{nk} - zI_{n-1})^{-1} - \mathsf{E}\mathsf{tr}(S_{nk} - zI_{n-1})^{-1}\big|^{2} \\ &\leq |z|^{2}\beta_{n}^{2}\mathsf{E}|s_{n}(z) - \mathsf{E}s_{n}(z)|^{2} + \frac{2|z|^{2}}{m^{2}v^{2}}. \end{aligned}$$

$$\tag{80}$$

By (78)-(80) and Lemma 12,

$$|\mathsf{E}s_{n}(z) - \bar{b}_{n}| \lesssim \frac{1}{n\beta_{n}} \sum_{k=1}^{n} |\mathsf{E}\bar{\epsilon}_{k}| + \frac{1}{n\nu\beta_{n}} \sum_{k=1}^{n} \mathsf{E}|\bar{\epsilon}_{k}|^{2} \lesssim \frac{1}{n\nu} + \frac{1}{\nu\beta_{n}} \Big(\frac{1}{m} + \frac{1}{m^{2}\nu^{2}} + \beta_{n}^{2}\mathsf{E}|s_{n}(z) - \mathsf{E}s_{n}(z)|^{2}\Big) \lesssim \frac{1}{n\nu} \,. \tag{81}$$

The above bound describes an equation that is quadratic in  $Es_n(z)$ :

$$\operatorname{Es}_{n}(z) - \frac{1}{1 - z - \beta_{n} - \beta_{n} z \operatorname{Es}_{n}(z)} = \delta_{n}$$
(82)

where  $\delta_n := Es_n(z) - \bar{b}_n$  obeys  $|\delta_n| \leq 1/(nv) \leq 1/\sqrt{n\beta_n}$  (note that  $\bar{s}_n(z)$  is a solution of the related equation  $s = (1 - z - \beta_n z s)^{-1}$ ). By (3.3.8)-(3.3.13) of [4], the unique solution satisfying the requirement that  $\Im(Es_n(z)) > 0$  (recall that  $z \in \mathbb{C}^+$ ) is given by

$$\mathrm{E}s_n(z) = \frac{1}{2\beta_n z} \Big( 1 - z - \beta_n + \beta_n z \delta_n + \sqrt{(z + \beta_n - 1 + \beta_n z \delta_n)^2 - 4\beta_n z} \Big).$$

Thus,

$$|\operatorname{Es}_{n}(z) - \bar{s}_{n}(z)| = \frac{1}{2} \left| \delta_{n} + \frac{2(z + \beta_{n} - 1)\delta_{n} + \beta_{n}z\delta_{n}^{2}}{\sqrt{(z + \beta_{n} - 1)^{2} - 4\beta_{n}z} + \sqrt{(z + \beta_{n} - 1 + \beta_{n}z\delta_{n})^{2} - 4\beta_{n}z}} \right| \\ \leq \frac{|\delta_{n}|}{2} \left( 1 + \frac{|2(z + \beta_{n} - 1) + \beta_{n}z\delta_{n}|}{|\sqrt{(z + \beta_{n} - 1)^{2} - 4\beta_{n}z} + \sqrt{(z + \beta_{n} - 1 + \beta_{n}z\delta_{n})^{2} - 4\beta_{n}z}|} \right) \lesssim \frac{1}{nv} . \quad \Box$$
(83)

**Proof of Theorem 3.** Let z satisfy  $|z - u| \le n^{-1/4}\sqrt{\beta_n}$  and  $z' = u + in^{-1/4}\sqrt{\beta_n}$ . We will decompose  $|s_n(z) - \bar{s}_n(z)|$  as

$$|s_n(z) - \bar{s}_n(z)| \le |s_n(z) - s_n(z')| + |s_n(z') - \bar{s}_n(z')| + |\bar{s}_n(z) - \bar{s}_n(z')|$$

and bound each term.  $|s_n(z) - s_n(z')|$  is bounded in the proof of Lemma 6. By Lemmas 12 and 13,

$$|s_n(z') - \bar{s}_n(z')| \lesssim \frac{1}{n^\ell \sqrt{\beta_n}},$$
(84)

almost surely. As *u* is bounded away from the support of  $\overline{F}_n$  by a constant multiple of  $\sqrt{\beta_n}(17)$ ,

$$\begin{split} |\bar{s}_n(z) - \bar{s}_n(z')| &\leq \int \Big| \frac{1}{\lambda - z} - \frac{1}{\lambda - z'} \Big| d\overline{F}_n(\lambda) = \int \frac{|z - z'|}{|\lambda - z||\lambda - z'|} d\overline{F}_n(\lambda) \\ &\lesssim \frac{1}{n^{1/4} \sqrt{\beta_n}} \cdot \quad \Box \end{split}$$
(85)

**Proof of Corollary 1.** Consider *z* satisfying  $|z - u| \le n^{-1/8}\sqrt{\beta_n}$ . Let  $v = n^{-1/8}\sqrt{\beta_n}$  and z' = u + iv. We will decompose  $|\frac{d}{dz}(s_n(z) - \bar{s}_n(z))|$  as

$$\left|\frac{d}{dz}(s_n(z)-\bar{s}_n(z))\right| \leq \left|\frac{d}{dz}(s_n(z)-s_n(z'))\right| + \left|\frac{d}{dz}(s_n(z')-\bar{s}_n(z'))\right| + \left|\frac{d}{dz}(\bar{s}_n(z)-\bar{s}_n(z'))\right|$$

and bound each term. For  $1 \le \alpha \le n$ , by the spacing bound (18),

$$\left|\frac{1}{(\lambda_{\alpha}-z)^2}-\frac{1}{(\lambda_{\alpha}-z')^2}\right|=\frac{|z-z'||2\lambda_{\alpha}-z-z'|}{|\lambda_{\alpha}-z|^2|\lambda_{\alpha}-z'|^2}\leq\frac{32}{\eta^3n^{1/8}\beta_n}.$$

Thus, almost surely,

$$\frac{d}{dz} \left( s_n(z) - s_n(z') \right) \bigg| \le \frac{1}{n} \sum_{\alpha=1}^n \left| \frac{1}{(\lambda_\alpha - z)^2} - \frac{1}{(\lambda_\alpha - z')^2} \right| \lesssim \frac{1}{n^{1/8} \beta_n} \,. \tag{86}$$

Note that

$$\left|\frac{1}{(\lambda-z')^2}-\frac{1}{(\lambda-u)^2+v^2}\right|\leq \frac{2v}{|\lambda-u|^3}, \qquad \frac{1}{(\lambda-u)^2+v^2}=\Im\Big(\frac{1}{v(\lambda-z)}\Big).$$

Hence,

$$\begin{split} \left| \frac{d}{dz} \big( s_n(z') - \bar{s}_n(z') \big) \right| &= \left| \int \frac{1}{(\lambda - z')^2} dF_n(\lambda) - \int \frac{1}{(\lambda - z')^2} d\overline{F}_n(\lambda) \right| \\ &\leq \left| \int \frac{1}{(\lambda - z')^2} - \frac{1}{(\lambda - u)^2 + v^2} (dF_n - d\overline{F}_n)(\lambda) \right| + \frac{1}{v} \left| \int \frac{1}{\lambda - z'} (dF_n - d\overline{F}_n)(\lambda) \right| \\ &\leq 2v \int \frac{1}{|\lambda - u|^3} (dF_n + d\overline{F}_n)(\lambda) + \frac{1}{v} |s_n(z') - \bar{s}_n(z')| \,. \end{split}$$

We bound the first term above as in Lemma 5 and the second using Theorem 3: almost surely,

$$\left|\frac{d}{dz}\left(s_n(z')-\bar{s}_n(z')\right)\right| \lesssim \frac{v}{\beta_n^{3/2}} + \frac{1}{n^{\ell+1/8}v\sqrt{\beta_n}} \,. \tag{87}$$

Finally, by (17),

$$\left|\frac{d}{dz}(\bar{s}_n(z)-\bar{s}_n(z'))\right| \leq \int \left|\frac{1}{(\lambda-z)^2}-\frac{1}{(\lambda-z')^2}\right| d\bar{F}_n(\lambda) \lesssim \frac{1}{n^{1/8}\beta_n} \,. \quad \Box$$
(88)

#### Acknowledgments

The author is grateful to David Donoho for continuous encouragement and detailed comments on drafts of this paper. The author would also like to thank the anonymous referees for their insightful suggestions and comments. This work was partially supported by NSF DMS grants 1407813, 1418362, and 1811614.

#### Appendix

**Lemma 14** (Theorem 1 of [11]). Let  $X_n = (x_{ij} : 1 \le i \le n, 1 \le j \le m)$  be an array of i.i.d. real-valued random variables with  $Ex_{11} = 0$ ,  $Ex_{11}^2 = 1$ , and  $Ex_{11}^4 < \infty$ . Suppose that as  $n \to \infty$ ,  $\beta_n = n/m_n \to 0$ . Define

$$A_n = \frac{1}{2\sqrt{\beta_n}} \left( \frac{1}{m} X_n X_n^\top - I_n \right).$$

Then,

$$\lambda_{\max}(A_n) \xrightarrow{a.s.} 1,$$

where  $\lambda_{\max}(A_n)$  represents the largest eigenvalue of  $A_n$ .

**Lemma 15** (Theorem 2 of [11]). In the setting of Lemma 14, let  $\widehat{X}_n = (\widehat{x}_{ij} : 1 \le i \le n, 1 \le j \le m)$  and  $Y_n = (y_{ij} : 1 \le i \le n, 1 \le j \le m)$  denote  $X_n$  after truncation and normalization, respectively:

$$\hat{x}_{ij} = x_{ij}I(|x_{ij}| \le \delta_n (nm)^{1/4}), \qquad y_{ij} = \frac{\hat{x}_{ij} - E\hat{x}_{11}}{\nu},$$

where  $\nu^2 = E(\hat{x}_{11} - E\hat{x}_{11})^2$  and  $\delta_n$  is a sequence constructed in Section 2 of [11], satisfying  $\delta_n \to 0$  and  $\delta_n(nm)^{1/4} \to \infty$ . Define

$$\widetilde{A}_n = \frac{1}{2\sqrt{\beta_n}} \left( \frac{1}{m} Y_n Y_n^\top - I_n \right).$$

Then, for any  $\eta$ ,  $\ell > 0$ ,

$$\Pr(\lambda_{\max}(\widetilde{A}_n) \ge 1 + \eta) = o(n^{-\ell})$$

Furthermore,  $Pr(X_n \neq \widehat{X}_n \text{ i.o.}) = 0$ ,

$$|\mathrm{E}\hat{x}_{11}| \lesssim rac{1}{(nm)^{3/4}}, \qquad |\nu^2 - 1| = o\Big(rac{1}{\sqrt{nm}}\Big).$$

**Lemma 16.** In the setting of Lemma 15, let  $P_n \in \mathbb{R}^{n \times m}$  have bounded operator norm and consider the singular value decompositions

$$P_n + \frac{1}{\sqrt{m}} X_n = U \Lambda V^\top \qquad P_n + \frac{1}{\sqrt{m}} Y_n = \widetilde{U} \widetilde{\Lambda} \widetilde{V}^\top$$

where  $\Lambda = \operatorname{diag}(\sigma_1, \ldots, \sigma_n)$  and  $\widetilde{\Lambda} = \operatorname{diag}(\widetilde{\sigma}_1, \ldots, \widetilde{\sigma}_n)$ . Almost surely,

$$\sup_{1\leq i\leq n}|\sigma_i^2-\tilde{\sigma}_i^2|\lesssim \frac{1}{\sqrt{nm}}.$$

Moreover, for fixed  $i \leq n$ , if  $\min(\sigma_i^2 - \sigma_{i-1}^2, \sigma_i^2 - \sigma_{i+1}^2) \asymp \sqrt{\beta_n}$ , where  $\sigma_0 := \infty$  and  $\sigma_n := -\infty$ , then

$$1-|(U^{\top}\widetilde{U})_{ii}|\lesssim \frac{1}{n^2} \qquad 1-|(V^{\top}\widetilde{V})_{ii}|\lesssim \frac{1}{n^2}.$$

**Proof.** By Lemmas 14 and 15, almost surely eventually,  $||X_n||_2 \le 2\sqrt{m}$ ,  $||Y_n||_2 \le 2\sqrt{m}$ , and

$$\frac{1}{\sqrt{m}} \|X_n - Y_n\|_2 = \frac{1}{\sqrt{m}} \|\widehat{X}_n - Y_n\|_2 = \frac{1}{\sqrt{m}} \|\widehat{X}_n - \nu Y_n - (1 - \nu)Y_n\|_2 
\leq \frac{1}{\sqrt{m}} \|(\widehat{E}\widehat{x}_{11})\mathbf{1}_n\mathbf{1}_m^\top\|_2 + \frac{\nu - 1}{\sqrt{m}} \|Y_n\|_2 \lesssim \frac{1}{\sqrt{nm}}.$$
(89)

The first claim of the lemma follows from (89) and Weyl's inequality:

$$|\sigma_i^2 - \tilde{\sigma}_i^2| = |\sigma_i - \tilde{\sigma}_i| |\sigma_i + \tilde{\sigma}_i| \le \frac{1}{\sqrt{m}} ||X_n - Y_n||_2 |\sigma_1 + \tilde{\sigma}_1|.$$

The second claim follows from the Davis–Kahan theorem (Corollary 3 of [18]).

**Lemma 17** (Lemma 2.7 of [3]). Let A be an  $n \times n$  nonrandom matrix and  $x = (x_1, \ldots, x_n)^{\top}$  be a random vector of independent entries. Assume that  $Ex_i = 0$ ,  $E|x_i|^2 = 1$ , and  $E|x_j|^\ell \le v_\ell$ . Then, for any  $\ell \ge 1$ ,

$$|\mathbf{E}|x^*Ax - trA|^{\ell} \lesssim_{\ell} (\nu_4 trA^*A)^{\ell/2} + \nu_{2\ell} tr(A^*A)^{\ell/2}$$

**Lemma 18.** Let A be an  $n \times m$  nonrandom matrix and  $x \in \mathbb{C}^n$ ,  $y \in \mathbb{C}^m$  random vectors. Assume x and y are independent with entries satisfying the moment conditions of Lemma 17. Then, for any  $\ell > 2$ ,

$$E|x^*Ay|^{\ell} \lesssim_{\ell} (1+\nu_4^{\ell/4}+\nu_\ell) \big( (\nu_4 \operatorname{tr}(A^*A)^2)^{\ell/4}+\nu_\ell \operatorname{tr}(A^*A)^{\ell/2}+(\operatorname{tr} A^*A)^{\ell/2} \big).$$

**Proof.** Condition on *y* and apply Lemma 17.  $\Box$ 

Lemmas 19 and 20 are elementary, following from the Leibniz determinant formula.

**Lemma 19.** For any  $n \times n$  matrices A and B,

 $|\det A - \det B| \leq_n (||A||_{\infty} + ||B||_{\infty})^{n-1} ||A - B||_{\infty}.$ 

**Lemma 20.** For any  $n \times n$  matrices A(z) and B(z).

$$\left|\frac{d}{dz}(\det A - \det B)\right| \lesssim_n \|A\|_{\infty}^{n-1} \left\|\frac{d}{dz}(A - B)\right\|_{\infty} + (\|A\|_{\infty} + \|B\|_{\infty})^{n-2} \left\|\frac{d}{dz}B\right\|_{\infty} \|A - B\|_{\infty},$$

where  $\frac{d}{dz}B$  denotes the entrywise derivative of the matrix. Additionally,

$$\frac{d}{dz}\det A\bigg|\lesssim_n\|A\|_{\infty}^{n-1}\bigg\|\frac{d}{dz}A\bigg\|_{\infty}.$$

The remaining lemmas pertain to the right singular vectors of the noise matrix. Consider an array  $X_n$  satisfying assumption A1 and a vector v satisfying assumptions A2 places on right signal vectors. Let W be a matrix containing as columns the first *n* (normalized) eigenvectors of  $X_n^{\top}X_n$ .

**Lemma 21.** Let  $S_n = \frac{1}{m}X_nX_n^{\top}$  and  $A_n = \frac{1}{m}X_n^{\top}S_n^{-1}X_n = WW^{\top}$ . Almost surely,

$$v^{\top}A_nv = \|W^{\top}v\|_2^2 \lesssim \beta_n \log(n)$$

The proof requires the following two lemmas:

**Lemma 22.** Let  $n/m \le \beta$  for some  $\beta \in (0, 1]$ . There exist constants  $c_0, c_1, c_2$  depending only on  $Ex_{11}^4$  such that with probability at least  $1 - c_0 \log(e/\beta) \exp(-c_1\beta m)$ ,

$$\frac{1}{\sqrt{m}}\sigma_{\min}(X_n)\geq 1-c_2\sqrt{\beta}.$$

**Proof.** This is one case of Theorem 1.3 of [13].

**Lemma 23.** Let  $\widetilde{A}_n = \frac{1}{m} X_n^{\top} (S_n + \frac{1}{m} I_n)^{-1} X_n$ . The fourth moment condition on the entries of  $\sqrt{m}v$  implies

$$\mathsf{E}(v^{\top}\widetilde{A}_nv) \leq \beta_n \qquad Var(v^{\top}\widetilde{A}_nv) \lesssim \frac{\beta_n^2}{n}.$$

Proof of Lemma 21. Let SD denote standard deviation. By Lemma 23, eventually,

$$\Pr(v^{\top}\widetilde{A}_{n}v \geq \beta_{n}\log(n)) \leq \Pr(v^{\top}\widetilde{A}_{n}v \geq \mathsf{E}(v^{\top}\widetilde{A}_{n}v) + \sqrt{n}\log(n) \cdot SD(v^{\top}\widetilde{A}_{n}v)) \leq \frac{1}{n\log^{2}(n)}$$

Summability of the right-hand-side gives  $v^{\top} \widetilde{A}_n v \leq \beta_n \log(n)$  almost surely eventually. Let  $\lambda_1 \geq \cdots \geq \lambda_n$  denote the eigenvalues of  $S_n$  and  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ . Observe that the eigenvalues of  $\widetilde{A}_n = W \Lambda^{1/2} (\Lambda + \frac{1}{m} I_n)^{-1} \Lambda^{1/2} W^{\top}$  are  $\lambda_j / (\lambda_j + \frac{1}{m})$ ,  $j = 1, \dots, n$ , joined by a zero eigenvalue with multiplicity m - n. A first consequence of Lemma 22 is  $\lambda_n \gtrsim 1$  almost surely. Hence,

$$v^{\top}A_nv \leq v^{\top}\widetilde{A}_nv + \|v\|_2^2\lambda_{\max}(A_n - \widetilde{A}_n) \leq v^{\top}\widetilde{A}_nv + \frac{\|v\|_2^2}{1 + m\lambda_n}$$
$$\lesssim \beta_n \log(n)$$

almost surely.  $\Box$ 

**Proof of Lemma 23.** As  $X_n$  and v are independent,

$$E(v^{\top}\widetilde{A}_{n}v)^{2} = \sum_{j_{1},j_{2},j_{3},j_{4}=1}^{m} E(v_{j_{1}}v_{j_{2}}v_{j_{3}}v_{j_{4}})E(\widetilde{A}_{j_{1}j_{2}}\widetilde{A}_{j_{3}j_{4}}).$$
(90)

Any term  $E(v_{j_1}v_{j_2}v_{j_3}v_{j_4})$  containing a singleton index vanishes, leaving the following terms:

$$j_{1} = j_{2} = j_{3} = j_{4} : \qquad \frac{\nu_{4}}{m^{2}} \sum_{j} E\widetilde{A}_{jj}^{2}$$

$$j_{1} = j_{2}, j_{3} = j_{4}, j_{1} \neq j_{3} : \qquad \frac{1}{m^{2}} \sum_{j_{1}, j_{3}} E(\widetilde{A}_{j_{1}j_{1}}\widetilde{A}_{j_{3}j_{3}}),$$

$$j_{1} = j_{3}, j_{2} = j_{4}, j_{1} \neq j_{2} : \qquad \frac{1}{m^{2}} \sum_{j_{1}, j_{2}} E\widetilde{A}_{j_{1}j_{2}}^{2},$$

$$j_{1} = j_{4}, j_{2} = j_{3}, j_{1} \neq j_{2} : \qquad \frac{1}{m^{2}} \sum_{j_{1}, j_{2}} E\widetilde{A}_{j_{1}j_{2}}^{2},$$

$$(91)$$

where  $v_4 = m^2 E v_1^4 < \infty$ . Let  $x_j$  denote the *j*th column of  $X_n$  and  $B_j = S_n + \frac{1}{m}(I_n - x_j x_j^\top)$ . By the Sherman–Morrison formula,

$$\begin{split} \mathbf{E}\widetilde{A}_{jj}^{2} &= \frac{1}{m^{2}} \mathbf{E} \left( \mathbf{x}_{j}^{\top} (S_{n} + \frac{1}{m} I_{n})^{-1} \mathbf{x}_{j} \right)^{2} = \frac{1}{m^{2}} \mathbf{E} \left( \mathbf{x}_{j}^{\top} B_{j}^{-1} \mathbf{x}_{j} - \frac{(\mathbf{x}_{j}^{\top} B_{j}^{-1} \mathbf{x}_{j})^{2}}{m + \mathbf{x}_{j}^{\top} B_{j}^{-1} \mathbf{x}_{j}} \right)^{2} \\ &= \frac{1}{m^{2}} \mathbf{E} \left( \frac{\mathbf{x}_{j}^{\top} B_{j}^{-1} \mathbf{x}_{j}}{1 + \frac{1}{m} \mathbf{x}_{j}^{\top} B_{j}^{-1} \mathbf{x}_{j}} \right)^{2} \le \frac{1}{m^{2}} \mathbf{E} (\mathbf{x}_{j}^{\top} B_{j}^{-1} \mathbf{x}_{j})^{2} \le \frac{1}{m^{2}} \mathbf{E} \|\mathbf{x}_{j}\|_{2}^{4} \mathbf{E} \lambda_{\max} (B_{j}^{-1})^{2} \end{split}$$

where the final inequality follows from the independence of  $x_j$  and  $B_j$ . Lemma 22 and the almost-sure lower bound  $\lambda_{\min}(B_j) \ge 1/m$  imply  $E\lambda_{\max}(B_j^{-1})^2 \lesssim 1$ . Moreover,  $E||x_j||_2^4 = n\nu_4 + (n^2 - n)\nu_2^2$ . Thus,

$$\mathbf{E}|\widetilde{A}_{ij}|^2 \lesssim \beta_n^2 \,. \tag{92}$$

Next, consider terms of the third and fourth types. Let  $B_{j_1,j_2} = S_n + \frac{1}{m}(I_n - x_{j_1}x_{j_1}^\top - x_{j_2}x_{j_2}^\top)$ . By the Woodbury formula,

$$\begin{split} \mathsf{E}\widetilde{A}_{j_{1},j_{2}}^{2} &= \frac{1}{m^{2}}\mathsf{E}\big(x_{j_{1}}^{\top}(S_{n} + \frac{1}{m}I_{n})^{-1}x_{j_{2}}\big)^{2} = \frac{1}{m^{2}}\mathsf{E}\Big(\frac{x_{j_{1}}^{\top}B_{j_{1},j_{2}}^{-1}x_{j_{2}}}{(1 + \frac{1}{m}x_{j_{1}}^{\top}B_{j_{1},j_{2}}^{-1}x_{j_{1}})(1 + \frac{1}{m}x_{j_{2}}^{\top}B_{j_{1},j_{2}}^{-1}x_{j_{2}}) - (\frac{1}{m}x_{j_{1}}^{\top}B_{j_{1},j_{2}}^{-1}x_{j_{2}})^{2}}\right)^{2} \\ &\leq \frac{1}{m^{2}}\mathsf{E}(x_{j_{1}}^{\top}B_{j_{1},j_{2}}^{-1}x_{j_{2}})^{2} \leq \frac{1}{m^{2}}\mathsf{E}\langle x_{j_{1}}, x_{j_{2}}\rangle^{2}\mathsf{E}\lambda_{\max}(B_{j_{1},j_{2}}^{-1})^{2}. \end{split}$$

The second-to-last inequality follows from the Cauchy–Schwarz inequality. The final step uses the independence of  $B_{j_1,j_2}$  from  $x_{j_1}$  and  $x_{j_2}$ . We have that  $E\lambda_{\max}(B_{j_1,j_2}^{-1})^2 \leq 1$  and  $E\langle x_{j_1}, x_{j_2} \rangle^2 = n$ . Therefore,

$$\mathrm{E}\widetilde{A}_{j_1,j_2}^2\lesssim rac{eta_n}{m}\,.$$

As the non-zero eigenvalues of  $\widetilde{A}_n = W \Lambda^{1/2} (\Lambda + \frac{1}{m} I_n)^{-1} \Lambda^{1/2} W^{\top}$  are  $\lambda_j / (\lambda_j + \frac{1}{m}), j \in \{1, \dots, n\}$ ,

$$\begin{split} \mathsf{E}(v^{\top}\widetilde{A}_{n}v) &= \frac{1}{m}\mathsf{E}\mathsf{tr}\widetilde{A}_{n} = \frac{1}{m}\sum_{j=1}^{n}\mathsf{E}\Big(\frac{\lambda_{j}}{\lambda_{j}+\frac{1}{m}}\Big) \leq \beta_{n}, \\ \mathsf{E}\big|(\mathsf{tr}\widetilde{A}_{n})^{2} - n^{2}\big| \leq \sum_{j_{1},j_{2}=1}^{n}\mathsf{E}\Big|\frac{1+m(\lambda_{j_{1}}+\lambda_{j_{2}})}{(1+m\lambda_{j_{1}})(1+m\lambda_{j_{2}})}\Big| \lesssim n\beta_{n}. \end{split}$$

We have used that by Lemma 22, there exists c > 0 such that with high probability,  $\lambda_n \ge c$ . On this event,

$$\left|\frac{1+m(\lambda_{j_1}+\lambda_{j_2})}{(1+m\lambda_{j_1})(1+m\lambda_{j_2})}\right| = \frac{\lambda_{j_1}^{-1}\lambda_{j_2}^{-1}+m(\lambda_{j_1}^{-1}+\lambda_{j_2}^{-1})}{\lambda_{j_1}^{-1}\lambda_{j_2}^{-1}+m(\lambda_{j_1}^{-1}+\lambda_{j_2}^{-1})+m^2} \le \frac{c^{-2}+2c^{-1}m}{c^{-2}+2c^{-1}m+m^2}$$

Thus,

$$\frac{1}{m^2} \sum_{j_1, j_3} \mathsf{E}(\widetilde{A}_{j_1 j_1} \widetilde{A}_{j_3 j_3}) - (\mathsf{E}v^\top \widetilde{A}_n v)^2 \bigg| = \frac{Var(\mathrm{tr}\widetilde{A}_n)}{m^2} \le \frac{1}{m^2} |\mathsf{E}(\mathrm{tr}\widetilde{A}_n)^2 - n^2| + \frac{1}{m^2} |(\mathsf{E}\mathrm{tr}\widetilde{A}_n)^2 - n^2| \le \frac{\beta_n^2}{m}.$$
(94)

 $Var(v^{\top}\widetilde{A}_n v) \lesssim n^{-1}\beta_n^2$  follows from (90)–(94).  $\Box$ 

#### References

R. Adamczak, On the Marchenko-Pastur and circular laws for some classes of random matrices with dependent entries, Electron. J. Probab. 16 (2011) 1068–1095.

- [2] Z. Bai, J. Hu, W. Zhou, Convergence rates to the Marchenko–Pastur type distribution, Stoch. Process. Appl. 122 (2012) 68–92.
- [3] Z. Bai, J. Silverstein, No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices, Ann. Probab. 26 (1998) 316-345.
- [4] Z. Bai, J. Silverstein, Spectral Analysis of Large Dimensional Random Matrices, Springer, New York, 2010.
- [5] Z. Bai, W. Zhou, Large sample covariance matrices without independence structures in columns, Statist. Sinica 18 (2008) 425-442.
- [6] J. Baik, G. Ben Arous, S. Péché, Phase transition of the largest eigenvalue for non-null complex sample covariance matrices, Ann. Probab. 33 (2005) 1643–1697.
- [7] J. Baik, J. Silverstein, Eigenvalues of large sample covariance matrices of spiked population models, J. Multivariate Anal. 97 (6) (2006) 1382–1408.
   [8] F. Benaych-Georges, R. Nadakuditi, The eigenvalues and eigenvectors of low rank perturbations of large random matrices, Adv. Math. 227 (1) (2011) 494–521.
- [9] F. Benaych-Georges, R. Nadakuditi, The singular values and vectors of low rank perturbations of large rectangular random matrices, J. Multivariate Anal. 111 (2012) 120–135.
- [10] A. Bloemendal, A. Knowles, H. Yau, J. Yin, On the principal components of sample covariance matrices, Probab. Theory Related Fields 164 (2016) 459–552.
- [11] B. Chen, G. Pan, Convergence of the largest eigenvalue of normalized sample covariance matrices when p and n both tend to infinity with their ratio converging to zero, Bernoulli 18 (4) (2012) 1405–1420.
- [12] M. Gavish, D. Donoho, The optimal hard threshold for singular values is  $4/\sqrt{3}$ , IEEE Trans. Inform. Theory 60 (2014) 5040–5053.
- [13] V. Koltchinskii, S. Mendelson, Bounding the smallest singular value of a random matrix without concentration, Int. Math. Res. Not. IMRN (2015) 12991–13008.
- [14] A. Montanari, E. Richard, A statistical model for tensor PCA, Proc. 27th Int. Conf. Neural Inf. Process. Syst. 2 (2014) 2897–2905.
- [15] J. Novembre, et al., Genes mirror geography within Europe, Nature 456 (2008) 98–101.
- [16] D. Paul, Asymptotics of sample eigenstructure for a large dimensional spiked covariance model, Statist. Sinica 17 (2007) 1617-1642.
- [17] L. Wang, D. Paul, Limiting spectral distribution of renormalized separable sample covariance matrices when  $p/n \rightarrow 0$ , J. Multivariate Anal. 126 (2014) 25–52.
- [18] Y. Yu, T. Wang, R. Samworth, A useful variant of the Davis-Kahan theorem for statisticians, Biometrika 102 (2015) 315-323.