Electron. Commun. Probab. **28** (2023), article no. 23, 1–17.
https://doi.org/10.1214/23-ECP527
ISSN: 1083-589X

ELECTRONIC COMMUNICATIONS in PROBABILITY

Marchenko-Pastur law for a random tensor model*

Pavel Yaskov†

Abstract

We study the limiting spectral distribution of large-dimensional sample covariance matrices associated with the order-d symmetric random tensors formed by products of d variables chosen from n independent standardized random variables. We find optimal sufficient conditions for this distribution to be the Marchenko-Pastur law in the case $d = d(n)$ and $n \to \infty$. Our conditions reduce to $d^2 = o(n)$ when the variables have uniformly bounded fourth moments. The proofs are based on a new concentration inequality for quadratic forms in symmetric random tensors and a law of large numbers for elementary symmetric random polynomials.

Keywords: random matrices; random tensors; sample covariance matrices; symmetric random polynomials.
MSC2020 subject classifications: 60B20.
Submitted to ECP on December 27, 2021, final version accepted on April 28, 2023.

1 Introduction

The paper studies the limiting spectral distribution (LSD) of large-dimensional sample covariance matrices associated with symmetric random tensor models. Namely, we consider sample covariance matrices of the form

$$\widehat{\Sigma}_N = \frac{1}{N}\sum_{k=1}^{N} \mathbf{x}_{pk}\mathbf{x}_{pk}^{\top}, \tag{1.1}$$

where $\{\mathbf{x}_{pk}\}_{k=1}^{N}$ are i.i.d. copies of a random vector $\mathbf{x}_p$ in $\mathbb{R}^p$ following one of the two models defined below.

Definition 1.1. Let X be a random vector in $\mathbb{R}^n$ with independent zero-mean, unit-variance entries X_α, $\alpha = 1, \dots, n$, and let $d \in \{1, \dots, n\}$. Then a random vector $\mathbf{x}_p$ in $\mathbb{R}^p$ follows the *symmetric random tensor model* (SRT model) with parameters (d, n)

> if $p = n^d$ and $\mathbf{x}_p$ could be obtained by vectorizing the symmetric tensor $X^{\otimes d}$ (for some X as above), i.e. the entries of $\mathbf{x}_p$ could be indexed by sequences $(\alpha_1, \dots, \alpha_d) \in \{1, \dots, n\}^d$ and defined as products $\prod_{k=1}^{d} X_{\alpha_k}$,

and the *restricted symmetric random tensor model* (RSRT model) with parameters (d, n)

*This work was supported by the Russian Science Foundation under grant no. 18-71-10097, https://www.rscf.ru/project/18-71-10097/, https://rscf.ru/project/21-71-03017/.
†Steklov Mathematical Institute of RAS, Moscow, Russia. E-mail: yaskov@mi-ras.ru

if $p = \binom{n}{d}$ and the entries of $\mathbf{x}_p$ could be indexed by d-element subsets $i \subseteq \{1, \ldots, n\}$ and defined as products $\prod_{\alpha \in i} X_\alpha$.

The RSRT model always corresponds to isotropic random vectors $\mathbf{x}_p$, i.e. such that $\mathbb{E}\mathbf{x}_p\mathbf{x}_p^\top = I_p$. In contrast, the SRT model is a highly correlated data model. To the best of our knowledge, the SRT model has not been systematically studied before in the context of the Marchenko-Pastur (MP) law. In this paper we will study the SRT model by reducing it to the RSRT model.

The RSRT model was investigated by Bryson, Vershynin, and Zhao [6], who showed that LSD of $\widehat{\Sigma}_N$ is the MP distribution[1] if $N \to \infty$, $\binom{n}{d}/N$ tends to a positive constant, the fourth moments of $X_\alpha = X_\alpha(d, n)$ are uniformly bounded, and $d^3 = o(n)$. However, they conjectured that the optimal condition is $d^2 = o(n)$. We prove this conjecture. As a corollary, we also obtain a similar result for the SRT model.

The RSRT model appeared in the study of the spectrum of inner-product kernel matrices $(h((X^{(k)})^\top X^{(l)}))_{k,l=1}^N$ in the polynomial regime with $n^d \asymp N$ for given $d \in \mathbb{N}$, where $(X^{(k)})_{k=1}^N$ are i.i.d. random vectors in $\mathbb{R}^n$ and $h : [-1, 1] \to \mathbb{R}$ is some square-integrable function. As shown in [25] and [31], the latter problem plays a crucial role in deriving the predictive risk of kernel ridge regression. To see where the RSRT model appears, let the data vectors $X^{(k)}$ be uniformly distributed on the unit sphere S^{n-1} of $\mathbb{R}^n$, then, as in [25] and [31], one could write an inner-product kernel matrix as a weighted series of the matrices $Y_d Y_d^\top$, $d \in \mathbb{N}$, where Y_d is a $N \times p$ matrix having the rows $(Y_{d,1}(X^{(k)}), \ldots, Y_{d,p}(X^{(k)}))$, $1 \leqslant k \leqslant N$, with $Y_{d,l} : S^{n-1} \to \mathbb{R}$ being the l-th spherical harmonic of degree d in n dimensions and $p = p(n, d)$ being the total number of such harmonics.[2] Now, the key observation is that, for large n (when d is fixed), up to a unitary transformation, almost all functions in the set $\{Y_{d,l}(x)\}_{l=1}^p$, $x = (x_\alpha)_{\alpha=1}^n \in S^{n-1}$, are monomials of the form

$$C_{d,n} \prod_{\alpha \in i} x_\alpha,$$

where i is as in Definition 1.1 and $C_{d,n} > 0$ is a normalizing factor. Using this observation and assuming $n^d \asymp N$ for fixed d, Xiao and Pennington [31] first proved the MP law for the RSRT model with X being uniformly distributed over the sphere of radius $\sqrt{n}$ in $\mathbb{R}^n$, and then for $Y_d Y_d^\top$ (see also Theorem 2 in [25]). Our methods allow to prove the corresponding result in the case $d^2 = o(n)$, for details, see Remark 2.6 in Section 2.

Much more is known about the non-symmetric random tensor model defined below.

Definition 1.2. Let X be a random vector in $\mathbb{R}^n$ with independent zero-mean, unit-variance entries X_α, $\alpha = 1, \ldots, n$, and let $d \in \mathbb{N}$. Then a random vector $\mathbf{x}_p$ in $\mathbb{R}^p$ follows the *non-symmetric random tensor model* (RT model) with parameters (d, n) if $p = n^d$ and $\mathbf{x}_p$ could be obtained by vectorizing the tensor $X^{(1)} \otimes \ldots \otimes X^{(d)}$, where $X^{(k)} = \big(X_\alpha^{(k)}\big)_{\alpha=1}^n$, $k = 1, \ldots, d$, are independent copies of X.

The model (1.1) with $\mathbf{x}_{pk}$ following the RT model appeared in quantum information theory and was studied by Ambainis, Harrow, and Hastings [4] as a quantum analog of the classical probability balls-into-bins problem. To see the analogy, let $\mathbf{x}_p/\sqrt{N}$ be uniformly distributed over the set $\{(1, 0, \ldots, 0), (0, 1, \ldots, 0), \ldots, (0, 0, \ldots, 1)\}$. Then $\widehat{\Sigma}_N$ is a diagonal matrix, whose spectrum represents a random allocation of N balls into p bins. The quantum analog of the balls-into-bins problem is to take $\mathbf{x}_p/\sqrt{N}$ being a random unit vector (uniformly distributed on the unit sphere in $\mathbb{C}^p$) and to consider the spectrum of $\widehat{\Sigma}_N$.[3] More generally, one could take $\mathbf{x}_p/\sqrt{N}$ to be random product states in $(\mathbb{C}^n)^{\otimes d}$ or, equivalently, to follow the RT model with X being a random unit vector. This

[1] See the definition of the MP distribution and the statement of the MP law in Section 2.

[2] When $n \to \infty$ and d is fixed, $p(n, d) \asymp n^d$.

[3] In the complex-valued case, $\mathbf{x}_{pk}^\top$ should be replaced by $\mathbf{x}_{pk}^*$ in (1.1).

setting was studied in [4], along with the case where X is a complex standard Gaussian vector (up to a factor of $n^{-1/2}$). It was shown in [4] that the expected moments of the empirical spectral distribution of $\widehat{\Sigma}_N$ converge to the corresponding moments of the MP distribution with parameter $\rho > 0$, when $N \to \infty$, d is fixed, and $N/n^d \to \rho$.

Lytova [22] extended the above result by showing that LSD of $\widehat{\Sigma}_N$ in the setting of [4] is the MP distribution a.s. in the case $d = o(n)$ (see Theorem 1.2 and Remark 4.1 in [22]). In fact, Lytova obtained a more general result on LSD of weighted sample covariance matrices

$$\Sigma_N = \frac{1}{N}\sum_{k=1}^{N} \tau_k \mathbf{x}_{pk}\mathbf{x}_{pk}^*,$$

where $\{\tau_k\}_{k=1}^{\infty}$ are real numbers and $\{\mathbf{x}_{pk}\}_{k=1}^{N}$ are i.i.d. copies of $\mathbf{x}_p$ following the RT model generated by a centered isotropic random vector X in $\mathbb{C}^n$, whose entries have uniformly bounded fourth moments and quadratic forms in which concentrate around their means in L_2 (see (A) in Section 2). Recently, Collins, Yao, and Yuan [7] proved that $d = o(n)$ is the optimal growth condition in the MP law for the RT model. They assumed the RT models generated by i.i.d. random variables with all moments finite and used the method of moments to show that, when d/n tends to a positive constant, the moment sequences of empirical spectral distributions of Σ_N have an explicit almost sure limit, which is different from the limit in the case of $d = o(n)$.

In the present paper, the proofs are based on a general version of the MP theorem and a new concentration inequality for quadratic forms in symmetric random tensors.

Concentration properties of quadratic forms provide a powerful tool to study the asymptotic behaviour of empirical spectral distributions of sample covariance and related random matrices (see [3], [5], [8], [12], [14], [17], [26], [35]). In particular, [33] gives necessary and sufficient conditions for the MP theorem in terms of weak concentration of certain quadratic forms. For various models of data, the concentration properties are established in [5], [6], [22], [26], [32], [35], [36], among others.

Lytova [22] obtained concentration inequalities for quadratic forms in non-symmetric random tensors $X^{(1)} \otimes \ldots \otimes X^{(d)}$ generated by d independent copies of an isotropic random vector X in $\mathbb{R}^n$, quadratic forms in which concentrate around their means. In the simplest L_2 case, the results of [22] are derived from the Efron-Stein inequality for the variance of a function in $\{X^{(k)}\}_{k=1}^{d}$. Recently, Vershynin [30] proved corresponding exponential concentration inequalities, assuming that X has independent subgaussian entries.

In this paper, we derive L_2 concentration inequalities for quadratic forms in symmetric random tensors with optimal dependence on d (at least, when $d = O(n)$). Our results improve those of [6] and have some relevance to random chaoses. However, as noted in [6], known concentration inequalities for random chaoses [1], [2], [13], [18], [19], [20] exhibit an unspecified (possibly exponential) dependence on the degree d, which is too bad for our problem. Also, comparing to the non-symmetric case, the strong dependence structure of the RSRT (or SRT) model highly complicates its analysis. In particular, for the quadratic form $\|\mathbf{x}_p\|^2$ given by the squared l_2 norm, $\|\mathbf{x}_p\|^2 = \|\bigotimes_{k=1}^{n} X^{(k)}\|^2$ is the product $\prod_{k=1}^{d}\|X^{(k)}\|^2$ of independent random variables in the case of the RT model and $\|\mathbf{x}_p\|^2$ is the elementary symmetric polynomial of order d in the squared entries of X in the case of the RSRT model. The latter plays a key role in the context of the MP theorem for the RSRT model, as the necessary condition for the theorem in the isotropic case is given by $p^{-1}\|\mathbf{x}_p\|^2 \to 1$ in probability when $p \to \infty$ (see Section 2). This condition could be viewed as a law of large numbers for elementary symmetric polynomials in independent nonnegative random variables. The asymptotic behaviour of such polynomials was thoroughly studied in [9], [16], [23], [29]. Using the saddle-point approximation method of [16], we find necessary and sufficient conditions (in terms of d and n) for the above law of large numbers in the i.i.d. case.

The paper is structured as follows. Section 2 and 3 contain main results concerning the RSRT model and the SRT model, respectively. The proofs are deferred to Section 4. Auxiliary results are proved in the Supplementary Material.

2 Main results: the RSRT model

Let us introduce some notation. For $p \in \mathbb{N}$, let $\mathbf{x}_p$ be a random vector in $\mathbb{R}^p$. For $A \in \mathbb{R}^{p\times p}$ (with $\mathbb{R}^{p\times p}$ being the set of real $p \times p$ matrices), denote its spectral norm by $\|A\|$ and, for symmetric A, let $\mu_A := p^{-1}\sum_{i=1}^p \delta_{\lambda_i}$ be its empirical spectral distribution, where $\{\lambda_i\}_{i=1}^p$ is the spectrum of A (we allow $\lambda_i = \lambda_j$ for $i \neq j$) and δ_λ is a Dirac measure with mass at $\lambda \in \mathbb{R}$. Also, put $\mathbb{C}_+ := \{z \in \mathbb{C} : \operatorname{Im}(z) > 0\}$ and let $I_p \in \mathbb{R}^{p\times p}$ be the identity matrix. Denote further by $\mathcal{B}(\mathbb{R}_+)$ the Borel σ-algebra of $\mathbb{R}_+$, by $|S|$ the cardinality of a set S, by $\mathbb{1}(S)$ the indicator function of S, and by $[n]_d$ the set $\{i \subseteq \{1,\dots,n\} : |i| = d\}$. For a set of random variables X_n, write $X_n = o_{\mathbb{P}}(1)$ if $X_n \to 0$ in probability as $n \to \infty$, and $X_n = O_{\mathbb{P}}(1)$ if X_n are stochastically bounded. All random elements below will be defined on the same probability space.

Recall the definition of the Marchenko-Pastur (MP) distribution μ_ρ ($\rho > 0$), i.e.

$$d\mu_\rho = \max\{1 - 1/\rho, 0\}\, d\delta_0 + \frac{\sqrt{(a_+ - x)(x - a_-)}}{2\pi x\rho}\mathbb{1}\big(a_- \leqslant x \leqslant a_+\big)\, dx \text{ for } a_\pm = (1 \pm \sqrt{\rho})^2,$$

and a version of the MP theorem (going back to [24]) under the following general assumption that quadratic forms in $\mathbf{x}_p$ weakly concentrate around some values:

(A) $(\mathbf{x}_p^\top A_p \mathbf{x}_p - \operatorname{tr}(A_p))/p \xrightarrow{\mathbb{P}} 0$ as $p \to \infty$ for all sequences of symmetric positive semidefinite $A_p \in \mathbb{R}^{p\times p}$ with $\|A_p\| \leqslant 1$.

Theorem 2.1. *Let $p = p(N) \in \mathbb{N}$ satisfy $p/N \to \rho > 0$ as $N \to \infty$. Let also $\mathbf{x}_p$ be a random vector in $\mathbb{R}^p$ for every p. If (A) holds, then*

$$\mathbb{P}(\mu_{\widehat{\Sigma}_N} \to \mu_\rho \text{ weakly, } N \to \infty) = 1. \tag{2.1}$$

Furthermore, if (2.1) holds and $\mathbb{E}\mathbf{x}_p\mathbf{x}_p^\top = I_p$ for all p, then $\mathbf{x}_p^\top\mathbf{x}_p/p \xrightarrow{\mathbb{P}} 1$ as $p = p(N) \to \infty$.

The first part of the theorem follows from Theorem 2.1 in [34]. The second part of the theorem follows from Theorem 1.1 in [33]. To apply Theorem 2.1, one has to check (A). For the RSRT model, this could be done via the following concentration inequality, which is our first main result (for its proof, see Section 4).

Theorem 2.2. *Let $d, n \in \mathbb{N}$, $n \geqslant 16d$, and $p = \binom{n}{d}$. If $\mathbf{x}_p$ is a random vector in $\mathbb{R}^p$ following the RSRT model and $A_p \in \mathbb{R}^{p\times p}$, then*

$$\operatorname{Var}(\mathbf{x}_p^\top A_p \mathbf{x}_p) \leqslant p \operatorname{tr}(A_p A_p^\top) \begin{cases} (1 + K_n d/n)^d - 1 & \text{if } A_p \text{ is diagonal,} \\ (1 + 2K_n d/n)^d (16d/n)(d \wedge 8) & \text{if } A_p \text{ is zero-diagonal,} \\ 64 K_n d^2/n & \text{if } A_p \text{ is arbitrary, } 2K_n d^2 \leqslant n, \end{cases}$$

where $K_n = \max\{\mathbb{E}X_\alpha^4 : 1 \leqslant \alpha \leqslant n\}$ with X_α constituting $\mathbf{x}_p$ according to Definition 1.1.

Theorem 2.2 improves the corresponding upper bound in Theorem 1.9 in [6] by a factor of $n^{-1/3}$ (up to some constants). The latter follows from

$$K_n d^2/n = (K_n^{1/2} d/n^{1/3})^2/n^{1/3} \quad \text{and} \quad \operatorname{tr}(A_p A_p^\top) \leqslant \|A_p A_p^\top\| \operatorname{tr}(I_p) = p\|A_p\|^2.$$

Note also that if X_α are i.i.d. over α, then $\operatorname{Var}(\mathbf{x}_p^\top \mathbf{x}_p) \geqslant p^2(K_n - 1)d^2/n$ by Theorem 5.2 of Hoeffding [15]. This shows the sharpness of our bound.

Theorem 2.2 guarantees that (A) holds for any sequence of random vectors $\mathbf{x}_p$ following the RSRT model with $p = \binom{n}{d}$ and $X_\alpha = X_\alpha(d,n)$, when $K_n d^2 = o(n)$ and $n \to \infty$ (here d, n may depend on some parameter N that goes to infinity). This along with Theorem 2.1 gives a version of the MP theorem for the RSRT model under the 4th moment condition, extending Theorem 1.5 of [6], where it is assumed that $d^3 = o(n)$ and $K_n = O(1)$. We can state even a more general result.

Theorem 2.3. *Let $d = d(N), n = n(N) \in \mathbb{N}$, $N \in \mathbb{N}$, satisfy $d \leqslant n$ and $\binom{n}{d}/N \to \rho > 0$, whereinafter all limits are with respect to $N \to \infty$. Assume also that, for each $p = \binom{n}{d}$, $\mathbf{x}_p$ is a random vector in $\mathbb{R}^p$ that follows the RSRT model with parameters (d,n) and $X_\alpha = X_\alpha(d,n)$, $\alpha = 1, \ldots, n$. Then (2.1) follows from (2.2), where*

$$d\mathbb{E}X_\alpha^2 \mathbb{1}(dX_\alpha^2 > n) \to 0 \text{ and } \frac{d^2}{n}\mathbb{E}X_\alpha^4 \mathbb{1}(dX_\alpha^2 \leqslant n) \to 0 \text{ uniformly in } \alpha \in \{1, \ldots, n\}. \quad (2.2)$$

Conversely, (2.1) implies (2.2) if, for all d, n, $\{X_\alpha\}_{\alpha=1}^n$ are independent copies of a random variable X not depending on d, n and such that $\mathbb{E}X = 0$, $\mathbb{E}X^2 = 1 > \mathbb{P}(X^2 = 1)$.

Theorem 2.3 is proved in Section 4. The sufficiency part of the theorem follows from Theorem 2.1 and 2.2. The necessity part follows from Theorem 2.1 and the following law of large numbers (LLN) for elementary symmetric polynomials of the form

$$S_n^{(d)} := \sum_{1 \leqslant i_1 < \ldots < i_d \leqslant n} Z_{i_1} \cdots Z_{i_d}, \quad d, n \in \mathbb{N}.$$

Theorem 2.4. *Let $Z, Z_1, Z_2, \ldots$ be i.i.d. nonnegative nondegenerate random variables. If $\mathbb{E}Z = 1$ and $d = d(n) \in \{1, \ldots, n\}$ for $n \in \mathbb{N}$, then the following are equivalent as $n \to \infty$:*

(i) $S_n^{(d)}/\binom{n}{d} \xrightarrow{\mathbb{P}} 1$,

(ii) $d(S_n/n - 1) \xrightarrow{\mathbb{P}} 0$ *for* $S_n = S_n^{(1)}$,

(iii) $d\mathbb{E}Z\mathbb{1}(dZ > n) \to 0$ *and* $d^2\mathbb{E}Z^2\mathbb{1}(dZ \leqslant n) = o(n)$.

Theorem 2.4 is proved in Section 4. It gives necessary conditions for LLN for the U-statistic $U_n^{(d)} := S_n^{(d)}/\binom{n}{d}$. The asymptotic distribution of this U-statistic is thoroughly studied in [9], [16], [23], [29] for various asymptotic regimes. As shown in [9], under linear norming, the distribution may differ significantly for the cases $d^2 = o(n)$, $d^2 \sim cn$, and $d^2/n \to \infty$. However, to the best of our knowledge, there is no result giving (i) $\Rightarrow$ (ii) under no assumptions on d, n, and higher-order moments of Z.[4] To prove (i) $\Rightarrow$ (ii), we first show that (i) implies $d^2 = o(n)$ and then use the representation for $S_n^{(d)}$ from the next lemma, which proof is omitted as it closely follows the proof of Theorem in [16].[5]

Lemma 2.5. *Under the conditions of Theorem 2.3, let $d^2 = o(n)$ and $d \to \infty$. Then*

$$\ln \frac{S_n^{(d)}}{\binom{n}{d}} = \sum_{k=1}^n \ln(Z_k/\rho + 1) - d + d\ln(\rho d/n) + o_{\mathbb{P}}(1), \quad n \to \infty, \quad (2.3)$$

where $\rho = \rho(n, Z_1, \ldots, Z_n)$ is the unique solution of the equation $\sum_{k=1}^n \rho/(Z_k + \rho) = n - d$ if such solution exists and $\rho = 1$ otherwise.

Remark 2.6. Under the assumptions of Theorem 2.3, one could show that the MP law (2.1) holds in the case where (2.2) is replaced by $d^2 = o(n)$ and the random vector

[4] If $U_n^{(d)} \to 1$ in L_2, one can prove (ii) by using the following fact from the theory of U-statistics: $d(S_n/n - 1)$ is the Hájek projection of $U_n^{(d)} - 1$ in L_2, i.e. the projection on the set of sums $\sum_{k=1}^n f_k(Z_k)$, in particular, $\mathbb{E}|d(S_n/n-1)|^2 \leqslant \mathbb{E}|U_n^{(d)} - 1|^2$. However, this argument works well only for L_2 convergence.

[5] For positive Z, Lemma 2.5 follows from (8) in [29]. For general case, the proof of the lemma is given in the preprint version of this paper arXiv:2111.04296.

$X = (X_\alpha(d,n))_{\alpha=1}^n$ generating $\mathbf{x}_p$ have the form $X = \sqrt{n}\bar{X}/\|\bar{X}\|$, where $\bar{X} = \bar{X}(d,n)$ is a standard normal vector in $\mathbb{R}^n$. In the latter case, X will be uniformly distributed in the sphere of radius $\sqrt{n}$ in $\mathbb{R}^n$. In view of Theorem 2.1, to prove the MP law, one needs to verify (A). This could be done by noting the following (when $n \to \infty$ and $d^2 = o(n)$):

(1) $\mathbf{x}_p = \bar{\mathbf{x}}_p/\|\bar{X}/\sqrt{n}\|^d$ for $\bar{\mathbf{x}}_p$ following the RSRT model generated by $\bar{X}$ (here $p = \binom{n}{d}$),
(2) $\bar{\mathbf{x}}_p$ satisfies (A), as follows from Theorem 2.2,
(3) $(\|\bar{X}\|^2/n)^d \xrightarrow{\mathbb{P}} 1$, as follows from Theorem 2.3,
(4) $\mathbf{x}_p$ satisfies (A), as follows from (1)–(3).

For fixed d, such MP law was proved by Xiao and Pennington [31] (see the proof of Theorem 5 in [31]), who used this result to derive the limiting spectral distribution of inner-product kernel matrices (for details, see the Introduction).

3 Main results: the SRT model

In this section, we will give a version of the MP law for the SRT model. We need to introduce one more random tensor model.

Definition 3.1. Let X be a random vector in $\mathbb{R}^n$ with independent zero-mean, unit variance entries X_α, $\alpha = 1,\dots,n$, and let $d \in \{1,\dots,n\}$. Then a random vector $\mathbf{x}_p$ in $\mathbb{R}^p$ follows the *reduced symmetric random tensor model* (reduced SRT model) with parameters (d,n)

if $p = \binom{n+d-1}{d}$ and the entries of $\mathbf{x}_p$ could be indexed by sequences $(d_\alpha)_{\alpha=1}^n \in \{0,\dots,d\}^n$ with $\sum_{\alpha=1}^n d_\alpha = d$ and defined as products $\prod_{\alpha=1}^n X_\alpha^{d_\alpha}/\sqrt{d_\alpha!}$.

Note that for any $\mathbf{x}_p$ following the SRT model with parameters (d,n) and $p = n^d$ and having the entries $\prod_{k=1}^d X_{\alpha_k}$ with $(\alpha_k)_{k=1}^d \in \{1,\dots,n\}^d$, one can construct a unique vector $\bar{\mathbf{x}}_{\bar{p}}$ following the reduced SRT model with parameters (d,n) and $\bar{p} = \binom{n+d-1}{d}$ and having the entries

$$\frac{\prod_{k=1}^d X_{\alpha_k}}{\prod_{\alpha=1}^n \sqrt{d_\alpha!}} = \prod_{\alpha=1}^n \frac{X_\alpha^{d_\alpha}}{\sqrt{d_\alpha!}}, \quad \text{for} \quad \alpha_1 \leqslant \ldots \leqslant \alpha_d \quad \text{and} \quad d_\alpha = \sum_{k=1}^d \mathrm{I}(\alpha_k = \alpha).$$

Let $\widehat{\Sigma}_N$ ($N \in \mathbb{N}$) be a $p \times p$ sample covariance matrix constructed from the sample $\{\mathbf{x}_{pk}\}_{k=1}^N$ of i.i.d. copies of $\mathbf{x}_p$ as above and let $\bar{\Sigma}_N$ be a $\bar{p} \times \bar{p}$ sample covariance matrix $\bar{\Sigma}_N$ constructed from the sample $\{\bar{\mathbf{x}}_{pk}\}_{k=1}^N$, i.e. from i.i.d. copies of $\bar{\mathbf{x}}_p$.

Proposition 3.2. *Let $\widehat{\Sigma}_N$ and $\bar{\Sigma}_N$ be as above. If $\{\lambda_k\}_{k=1}^p$ and $\{\bar{\lambda}_k\}_{k=1}^{\bar{p}}$ are the sets of eigenvalues of $\widehat{\Sigma}_N$ and $\bar{\Sigma}_N$, respectively, where $\lambda_1 \geqslant \ldots \geqslant \lambda_p$ and $\bar{\lambda}_1 \geqslant \ldots \geqslant \bar{\lambda}_{\bar{p}}$. Then*

$$\lambda_k = d!\bar{\lambda}_k, \quad 1 \leqslant k \leqslant \bar{p}, \quad \textit{and} \quad \lambda_k = 0, \quad \bar{p} < k \leqslant p.$$

The proposition shows that one could study the spectrum of $\bar{\Sigma}_N$ (in the reduced SRT model) instead of the spectrum of $\widehat{\Sigma}_N$ (in the SRT model) because of the one-to-one correspondence between them. The main result of this section is the following MP law for the reduced SRT model.

Theorem 3.3. *Let $d = d(N), n = n(N) \in \mathbb{N}$, $N \in \mathbb{N}$, satisfy $d \leqslant n$ and $\binom{n+d-1}{d}/N \to \rho > 0$, whereinafter all limits are with respect to $N \to \infty$. Assume also that, for each $p = \binom{n+d-1}{d}$, $\mathbf{x}_p$ is a random vector in $\mathbb{R}^p$ that follows the reduced SRT model with parameters d,n and $X_\alpha = X_\alpha(d,n)$, $\alpha = 1,\dots,n$. Then* (2.1) *follows from* (2.2).

As in Theorem 2.3, we expect (2.2) to be a sharp sufficient condition. We could not formally verify it by using Theorem 2.1, because, in general, the vector $\mathbf{x}_p$ following the reduced SRT model is not isotropic. Moreover, the second moment of some entries of $\mathbf{x}_p$ could be infinite. The proof of Theorem 3.3 is based on Theorem 2.3, i.e. the corresponding result for the RSRT model.

Remark 3.4. Theorem 3.3 and Proposition 3.2 give a version of the MP law for the SRT model under the sharp sufficient condition $d^2 = o(n)$, when random variables $X_\alpha(d,n)$ generating the SRT model have uniformly bounded fourth moments. As shown by Collins, Yao, and Yuan [7], $d = o(n)$ is the corresponding optimal growth condition in the MP law for the RT model. As noted in the Introduction, Collins et al. [7] assumed the RT models generated by i.i.d. random variables with all moments finite and used the method of moments to show that, when d/n tends to a positive constant, the moment sequences of empirical spectral distributions of $\bar{\Sigma}_N$ have an explicit almost sure limit, which is different from the limit in the case of $d = o(n)$. We expect a similar result in the case of the SRT model when d^2/n tends to a positive constant. However, our method of the proof is based on Theorem 2.1, which does not allow to get such a result. One could try to use the method of moments as in [7] to analyse the case $d^2 = O(n)$, but we expect this will be much more involved comparing with the RT model, as the SRT model is generated by much fewer independent random variables (dn vs n). We leave this question for future research.

4 Proofs

Proof of Theorem 2.2. For brevity, we will write $\mathbf{x}$, A, K instead of $\mathbf{x}_p$, A_p, K_n. Let further $\mathbf{x} = (x_i)$, $A = (a_{ij})$, where i, j run all elements of $[n]_d$ ($= \{i \subseteq \{1, \dots, n\} : |i| = d\}$). Assume w.l.o.g. that A is non-zero (the result is trivial otherwise). First, consider the case of diagonal A. We have $\operatorname{Var}(\mathbf{x}^\top A\mathbf{x}) = \sum a_{ii}a_{jj}\operatorname{Cov}(x_i^2, x_j^2)$, where the sum is over all $i, j \in [n]_d$. Note that $\operatorname{Cov}(x_i^2, x_j^2) = 0$ if $i \cap j = \varnothing$ and

$$|\operatorname{Cov}(x_i^2, x_j^2)| \leqslant \mathbb{E}x_i^2x_j^2 = \prod_{\alpha\in i\Delta j} \mathbb{E}X_\alpha^2 \prod_{\beta\in i\cap j} \mathbb{E}X_\beta^4 \leqslant K^{|i\cap j|} \text{ for any } i, j.$$

The latter along with the Cauchy inequality yields

$$\operatorname{Var}(\mathbf{x}^\top A\mathbf{x}) \leqslant \frac{1}{2} \sum_{i,j: i\cap j\neq\varnothing} (a_{ii}^2 + a_{jj}^2)K^{|i\cap j|} = \sum_{i,j: i\cap j\neq\varnothing} a_{ii}^2 K^{|i\cap j|} = \sum_{t=1}^{d} K^t \sum_i a_{ii}^2 \sum_{j: |i\cap j| = t} 1.$$

If $|i \cap j| = t(\leqslant d)$ is fixed, then there are $\binom{d}{t}$ choices for choosing $i \cap j$ for any given i and $\binom{n-d}{d-t}$ choices for choosing $j \setminus i$ for any given i and $i \cap j$. Therefore, the very last sum is equal to $\binom{d}{t}\binom{n-d}{d-t}$. As $d \leqslant n$, we see that

$$\frac{\binom{n-d}{d-t}}{\binom{n}{d}} = \frac{d!/(d-t)!}{n!/(n-t)!}\frac{(n-d)!/(n-d-(d-t))!}{(n-t)!/(n-d)!} \leqslant \prod_{s=0}^{t-1}\frac{d-s}{n-s}\prod_{r=0}^{d-t-1}\frac{n-d-r}{n-t-r} \leqslant \left(\frac{d}{n}\right)^t 1^{d-t}.$$

Combining the above bounds and recalling that $p = \binom{n}{d}$, we derive that

$$\operatorname{Var}(\mathbf{x}^\top A\mathbf{x}) \leqslant p \operatorname{tr}(AA^\top) \sum_{t=1}^{d} \binom{d}{t} (Kd/n)^t = p \operatorname{tr}(AA^\top)((1 + Kd/n)^d - 1).$$

Consider the case of zero-diagonal A. Let $\Delta_t := \sum a_{ij}x_ix_j \mathbb{1}(|i \cap j| = t)$, $t \in [0, d) \cap \mathbb{Z}_+$, where the sum is over $i, j \in [n]_d$. By the triangle inequality for the norm $\|\cdot\|_2 = \sqrt{\mathbb{E}|\cdot|^2}$,

$$\sqrt{\operatorname{Var}(\mathbf{x}^\top A\mathbf{x})} = \sqrt{\mathbb{E}|\mathbf{x}^\top A\mathbf{x}|^2} = \Big\|\sum_{t=0}^{d-1}\Delta_t\Big\|_2 \leqslant \sum_{t=0}^{d-1}\|\Delta_t\|_2 = \sum_{t=0}^{d-1}\sqrt{\mathbb{E}|\Delta_t|^2}.$$

Let us estimate $\mathbb{E}|\Delta_t|^2$ for any fixed t. By definition,

$$\mathbb{E}|\Delta_t|^2 = \sum_{i,j,k,l: |i\cap j| = |k\cap l| = t} a_{ij}a_{kl}\mathbb{E}x_ix_jx_kx_l.$$

The product $x_ix_jx_kx_l$ is equal to $\prod_{\alpha\in\Lambda_1} X_\alpha \prod_{\beta\in\Lambda_2} X_\beta^2 \prod_{\gamma\in\Lambda_3} X_\gamma^3 \prod_{\delta\in\Lambda_4} X_\delta^4$, where $\Lambda_c = \Lambda_c(i,j,k,l)$, $c = 1,\ldots,4$, contains all $\alpha \in i\cup j\cup k\cup l$ that are covered by exactly c sets among i,j,k,l. If $|\Lambda_1| \neq 0$, then $\mathbb{E}x_ix_jx_kx_l = 0$. When $|\Lambda_1| = 0$, it follows from the independence of $\{X_\alpha\}_{\alpha=1}^n$ and the Cauchy–Schwarz inequality that

$$|\mathbb{E}x_ix_jx_kx_l| = \prod_{\gamma\in\Lambda_3} |\mathbb{E}X_\gamma^3| \prod_{\delta\in\Lambda_4} \mathbb{E}X_\delta^4 \leqslant K^{|\Lambda_4|} \prod_{\gamma\in\Lambda_3} \sqrt{\mathbb{E}X_\gamma^4 \mathbb{E}X_\gamma^2} \leqslant K^s,$$

where $s := |\Lambda_3|/2 + |\Lambda_4| \in \mathbb{Z}_+$ if $|\Lambda_1| = 0$ by $|\Lambda_1| + 2|\Lambda_2| + 3|\Lambda_3| + 4|\Lambda_4| = 4d$. We get that

$$\mathbb{E}|\Delta_t|^2 \leqslant \sum_{s\in\mathbb{Z}_+} K^s \sum_{(i,j,k,l)\in\Gamma(s,t)} |a_{ij}a_{kl}|$$

for $\Gamma(s,t) := \{(i,j,k,l) : |i\cap j| = |k\cap l| = t, |\Lambda_1| = 0, |\Lambda_3|/2 + |\Lambda_4| = s\}$ and $\Lambda_c = \Lambda_c(i,j,k,l)$, $c = 1,3,4$.

By symmetry, for all given i,j with $|i\cap j| = t$, the sets $\{(k,l) : (i,j,k,l) \in \Gamma(s,t)\}$ have the same cardinality. Let us denote this cardinality by $\gamma(s,t)$. By the Cauchy inequality,

$$\sum_{(i,j,k,l)\in\Gamma(s,t)} |a_{ij}a_{kl}| \leqslant \frac{1}{2} \sum_{(i,j,k,l)\in\Gamma(s,t)} (a_{ij}^2 + a_{kl}^2) = \sum_{(i,j,k,l)\in\Gamma(s,t)} a_{ij}^2 = \gamma(s,t) \sum_{i,j:|i\cap j|=t} a_{ij}^2.$$

Combining the above bounds (for all t) along with the Cauchy-Schwartz inequality gives

$$\sqrt{\mathrm{Var}(\mathbf{x}^\top A\mathbf{x})} \leqslant \sum_{t=0}^{d-1} \sqrt{\sum_{i,j:|i\cap j|=t} a_{ij}^2} \sqrt{\sum_{s\in\mathbb{Z}_+} K^s\gamma(s,t)} \leqslant \sqrt{\mathrm{tr}(AA^\top)} \sqrt{\sum_{t=0}^{d-1}\sum_{s\in\mathbb{Z}_+} K^s\gamma(s,t)}.$$

Lemma 4.1. *Under the above notations, for all $t \in [0,d)\cap\mathbb{Z}_+$ and $s \in \mathbb{Z}_+$, we have*

$$\gamma(s,t) \leqslant 2^{4(d-t)} \binom{n}{d} \binom{t}{s} 2^s (d/n)^{d-(t-s)} \mathbb{1}(s \leqslant t).$$

Lemma 4.1 is proved in the Supplementary Material. For $p = \binom{n}{d}$, it implies that

$$\frac{\mathrm{Var}(\mathbf{x}^\top A\mathbf{x})}{p\,\mathrm{tr}(AA^\top)} \leqslant \sum_{t=0}^{d-1} \left(\frac{2^4 d}{n}\right)^{d-t} \sum_{s=0}^{t} \binom{t}{s} \left(\frac{2Kd}{n}\right)^s = \sum_{t=0}^{d-1} \left(\frac{16d}{n}\right)^{d-t} \left(1 + \frac{2Kd}{n}\right)^t.$$

If S is the last sum, then $S \leqslant (16d^2/n)(1+2Kd/n)^d$. Also, as $n \geqslant 16d$ and $K \geqslant (\mathbb{E}X_1^2)^2 = 1$,

$$S = \frac{16d}{n} \frac{(1+2Kd/n)^d - (16d/n)^d}{1+2Kd/n - 16d/n} \leqslant \frac{16d}{n} \frac{(1+2Kd/n)^d}{1-14d/n} \leqslant \frac{2^7 d}{n}(1+2Kd/n)^d.$$

This proves the desired bound for zero-diagonal A. The variance bound for an arbitrary matrix A follows from the corresponding bounds for $A_0 = (a_{ij}\mathbb{1}(i=j))_{i,j\in[n]_d}$ and $A_1 = A - A_0$ along with the inequalities $\sqrt{\mathrm{Var}(\mathbf{x}^\top A\mathbf{x})} \leqslant \sqrt{\mathrm{Var}(\mathbf{x}^\top A_0\mathbf{x})} + \sqrt{\mathrm{Var}(\mathbf{x}^\top A_1\mathbf{x})}$, $K \geqslant 1$, $(1+x)^d \leqslant e^{dx}$ and $(1+x)^d - 1 \leqslant dxe^{dx}$, $x \geqslant 0$. □

Proof of Theorem 2.3. Suppose (2.2) holds. To apply Theorem 2.1, we will verify (A). For each $p = \binom{n}{d}$, let $A_p \in \mathbb{R}^{p\times p}$ be a positive semidefinite symmetric matrix with $\|A_p\| \leqslant 1$. Consider a random vector $\mathbf{y}_p$ defined as $\mathbf{x}_p$ in Definition 1.1 (RSRT model) with X_α replaced by $Y_\alpha := X_\alpha \mathbb{1}(dX_\alpha^2 \leqslant n)$. As $\{\mathbf{x}_p^\top A_p\mathbf{x}_p \neq \mathbf{y}_p^\top A_p\mathbf{y}_p\} \subseteq \bigcup_{\alpha=1}^n \{dX_\alpha^2 > n\}$, (2.2) yields that

$$\mathbb{P}(\mathbf{x}_p^\top A_p\mathbf{x}_p \neq \mathbf{y}_p^\top A_p\mathbf{y}_p) \leqslant \sum_{\alpha=1}^n \mathbb{P}(dX_\alpha^2 > n) \leqslant n \max_{1\leqslant\alpha\leqslant n} \mathbb{E}(dX_\alpha^2/n)\mathbb{1}(dX_\alpha^2 > n) =: L_{n,d} \to 0.$$

Thus, it sufficient to check (A) for $\mathbf{x}_p$ replaced by $\mathbf{y}_p$. We will do it by showing that for $\mathbf{z}_p$ defined as $\mathbf{x}_p$ in Definition 1.1 (RSRT model) with X_α replaced by $Z_\alpha := Y_\alpha - \mathbb{E}Y_\alpha$, the following holds:

(a) $(\mathbf{z}_p^\top A_p \mathbf{z}_p - \mathbf{y}_p^\top A_p \mathbf{y}_p)/p \xrightarrow{\mathbb{P}} 0$,

(b) $(\mathbf{z}_p^\top A_p \mathbf{z}_p - \mathbb{E}\mathbf{z}_p^\top A_p \mathbf{z}_p)/p \xrightarrow{\mathbb{P}} 0$,

(c) $(\mathbb{E}\mathbf{z}_p^\top A_p \mathbf{z}_p - \mathrm{tr}(A_p))/p \xrightarrow{\mathbb{P}} 0$.

Let us prove (a). Set $\mathbf{z}_{p,0} := \mathbf{y}_p$ and define $\mathbf{z}_{p,k}$, $k = 1, \dots, n$, as $\mathbf{y}_p$ above with Y_α replaced by Z_α for $\alpha = 1, \dots, k$. We have

$$\mathbb{E}|\mathbf{z}_p^\top A_p \mathbf{z}_p - \mathbf{y}_p^\top A_p \mathbf{y}_p| \leqslant \mathbb{E}|\mathbf{z}_{p,n}^\top A_p \mathbf{z}_{p,n} - \mathbf{z}_{p,0}^\top A_p \mathbf{z}_{p,0}| \leqslant \sum_{k=1}^n \mathbb{E}\Delta_{p,k},$$

where $\Delta_{p,k} := |\mathbf{z}_{p,k}^\top A_p \mathbf{z}_{p,k} - \mathbf{z}_{p,k-1}^\top A_p \mathbf{z}_{p,k-1}| = |\|A_p^{1/2}\mathbf{z}_{p,k}\|^2 - \|A_p^{1/2}\mathbf{z}_{p,k-1}\|^2|$. To estimate $\mathbb{E}\Delta_{p,k}$ for any given k, we will use the inequality

$$\mathbb{E}|\xi_0^2 - \xi_1^2| \leqslant \sqrt{\mathbb{E}(\xi_0 - \xi_1)^2}\sqrt{\mathbb{E}(\xi_0 + \xi_1)^2} \leqslant 2 \max_{q=0,1} \sqrt{\mathbb{E}\xi_q^2}\sqrt{\mathbb{E}(\xi_0 - \xi_1)^2}$$

valid for any ξ_0, ξ_1 in L_2. Taking $\xi_q = \|A_p^{1/2}\mathbf{z}_{p,k-q}\|$ and using that $\|A_p^{1/2}\| = \sqrt{\|A_p\|} \leqslant 1$, we get $|\xi_0 - \xi_1| \leqslant \|A_p^{1/2}(\mathbf{z}_{p,k} - \mathbf{z}_{p,k-1})\| \leqslant \|\mathbf{z}_{p,k} - \mathbf{z}_{p,k-1}\|$ and $\mathbb{E}\xi_q^2 \leqslant \mathbb{E}\|\mathbf{z}_{p,k-q}\|^2 \leqslant p$ as

$$\mathbb{E}Z_\alpha^2 = \mathrm{Var}(Y_\alpha) \leqslant \mathbb{E}Y_\alpha^2 = \mathbb{E}X_\alpha^2 \mathbb{1}(dX_\alpha^2 \leqslant n) \leqslant \mathbb{E}X_\alpha^2 = 1. \tag{4.1}$$

Thus, $\mathbb{E}\Delta_{p,k} \leqslant 2\sqrt{p}\sqrt{\mathbb{E}\|\mathbf{z}_{p,k} - \mathbf{z}_{p,k-1}\|^2}$. The entries $\mathbf{z}_{p,k} - \mathbf{z}_{p,k-1}$ have the form

$$-\mathbb{1}(k \in i) \prod_{\alpha\colon \alpha \in i,\, 1 \leqslant \alpha < k} Z_\alpha \cdot \mathbb{E}Y_k \cdot \prod_{\alpha\colon \alpha \in i,\, k < \alpha \leqslant n} Y_\alpha$$

for $i \in [n]_d (= \{j \subseteq \{1, \dots, n\} : |i| = d\})$, here $\prod_{i \in \varnothing} = 1$. By (4.1),

$$\mathbb{E}\|\mathbf{z}_{p,k} - \mathbf{z}_{p,k-1}\|^2 \leqslant (\mathbb{E}Y_k)^2 \sum_{i \in [n]_d : k \in i} \mathbb{1}(k \in i) = (\mathbb{E}Y_k)^2 \binom{n-1}{d-1} =$$

$$= (\mathbb{E}Y_k)^2 \binom{n}{d} \frac{d}{n} = \frac{pd}{n}(\mathbb{E}X_k \mathbb{1}(dX_k^2 \leqslant n))^2.$$

It follows from $\mathbb{E}X_k = 0$, that $\mathbb{E}X_k \mathbb{1}(dX_k^2 \leqslant n) = -\mathbb{E}X_k \mathbb{1}(dX_k^2 > n)$ and

$$|\mathbb{E}X_k \mathbb{1}(dX_k^2 \leqslant n)| \leqslant \mathbb{E}|X_k| \mathbb{1}(dX_k^2 > n) \leqslant \sqrt{d/n}\, \mathbb{E}X_k^2 \mathbb{1}(dX_k^2 > n) \tag{4.2}$$

Combining the above estimates yields

$$\mathbb{E}|\mathbf{z}_p^\top A_p \mathbf{z}_p - \mathbf{y}_p^\top A_p \mathbf{y}_p| \leqslant \sum_{k=1}^n \mathbb{E}\Delta_{p,k} \leqslant \frac{2pd}{n} \sum_{k=1}^n \mathbb{E}X_k^2 \mathbb{1}(dX_k^2 > n) \leqslant 2pL_{n,d} = o(p).$$

This implies (a).

Let us prove (b). Applying Theorem 2.2 to $\bar{\mathbf{z}}_p$ defined as $\mathbf{x}_p$ in Definition 1.1 (RSRT model) with X_α replaced by $Z_\alpha/\sqrt{\mathbb{E}Z_\alpha^2}$, we get

$$\mathrm{Var}(\mathbf{z}_p^\top A_p \mathbf{z}_p) = \mathrm{Var}(\bar{\mathbf{z}}_p^\top D_p^{1/2} A_p D_p^{1/2} \bar{\mathbf{z}}_p) \leqslant \frac{64 K_n p d^2}{n} \mathrm{tr}((D_p^{1/2} A_p D_p^{1/2})^2) \text{ if } 2K_n d^2 \leqslant n,$$

where $D_p := \mathbb{E}\mathbf{z}_p \mathbf{z}_p^\top \in \mathbb{R}^{p \times p}$ is diagonal with diagonal entries $\prod_{\alpha \in i} \mathbb{E}Z_\alpha^2$, $i \in [n]_d$, and

$$K_n := \max_{1 \leqslant \alpha \leqslant n} \frac{\mathbb{E}Z_\alpha^4}{(\mathbb{E}Z_\alpha^2)^2} \leqslant \frac{16 \max_{1 \leqslant \alpha \leqslant n} \mathbb{E}X_\alpha^4 \mathbb{1}(dX_\alpha^2 \leqslant n)}{\min_{1 \leqslant \alpha \leqslant n} (\mathbb{E}Z_\alpha^2)^2},$$

here we have used that $\mathbb{E}Z_\alpha^4 = \mathbb{E}(Y_\alpha - \mathbb{E}Y_\alpha)^4 \leqslant 8(\mathbb{E}Y_\alpha^4 + (\mathbb{E}Y_\alpha)^4) \leqslant 16\mathbb{E}Y_\alpha^4$. By (4.2) and $L_{n,d} = o(1)$, we have uniformly in α, $\mathbb{E}Y_\alpha = \mathbb{E}X_\alpha \mathbb{1}(dX_\alpha^2 \leqslant n) = o((dn)^{-1/2})$,

$$\mathbb{E}Z_\alpha^2 = \mathbb{E}Y_\alpha^2 - (\mathbb{E}Y_\alpha)^2 = 1 - \mathbb{E}X_\alpha^2 \mathbb{1}(dX_\alpha^2 > n) + o((dn)^{-1}) = 1 + o(d^{-1}) \tag{4.3}$$

As a result, by (2.2), $d^2 K_n/n = o(1)/(1+o(d^{-1})) = o(1)$. In addition,

$$\operatorname{tr}((D_p^{1/2} A_p D_p^{1/2})^2) \leqslant p\|(D_p^{1/2} A_p D_p^{1/2})^2\| \leqslant p\|A_p\|^2 \|D_p^{1/2}\|^4 \leqslant p\|D_p^{1/2}\|^4 =$$

$$= p \max_{i\in[n]_d} \Big(\prod_{\alpha\in i} \mathbb{E}Z_\alpha^2\Big)^2 \leqslant p\Big(\max_{1\leqslant\alpha\leqslant n} \mathbb{E}Z_\alpha^2\Big)^{2d} = p(1+o(d^{-1}))^{4d} = p(1+o(1)).$$

Finally, we conclude that $K_n d^2 = o(n)$ and $\operatorname{Var}(\mathbf{z}_p^\top A_p \mathbf{z}_p) \leqslant p^2 o(1)$, which implies (b).

The relation (c) follows from $\mathbb{E}\mathbf{z}_p^\top A_p \mathbf{z}_p = \operatorname{tr}(A_p D_p) = \operatorname{tr}(A_p) + \operatorname{tr}(A_p(D_p - I_p))$ and $\operatorname{tr}(A_p(D_p - I_p)) = o(p)$. Here the last equality could be derived from Von Neumann's trace inequality and (4.3) as follows:

$$p^{-1}|\operatorname{tr}(A_p(D_p - I_p))| \leqslant \|A_p\|\|D_p - I_p\| \leqslant \|D_p - I_p\| =$$

$$= \max_{i\in[n]_d} \Big|\prod_{\alpha\in i} \mathbb{E}Z_\alpha^2 - 1\Big| \leqslant 1 - \min_{1\leqslant\alpha\leqslant n} (\mathbb{E}Z_\alpha^2)^d = 1 - (1+o(d^{-1}))^d = o(1),$$

where we have also used that $\mathbb{E}Z_\alpha^2 = \operatorname{Var}(Y_\alpha) \leqslant \mathbb{E}Y_\alpha^2 \leqslant \mathbb{E}X_\alpha^2 = 1$ for all α. We have verified the sufficiency part of the theorem.

Let us prove the necessity part, i.e. (2.1) $\Rightarrow$ (2.2). Let (2.1) holds and, for each d, n, $X_\alpha = X_\alpha(d,n)$ are independent copies of X with $\mathbb{E}X = 0$, $\mathbb{E}X^2 = 1 > \mathbb{P}(X^2 = 1)$. So, by Theorem 2.1, $\mathbf{x}_p^\top \mathbf{x}_p/p \to 1$ in probability. Taking $Z = X^2$ in Theorem 2.4, we see that

$$\mathbf{x}_p^\top \mathbf{x}_p/p \text{ and } S_n^{(d)}/\tbinom{n}{d} \text{ from Theorem 2.4 have the same distribution.}$$

As convergence in distribution to a constant implies convergence in probability to the same constant, (i) in Theorem 2.4 holds. By (i) $\Rightarrow$ (iii) in Theorem 2.4, (2.2) holds. ☐

Proof of Theorem 2.4. First, we will show that any of the conditions (i), (ii), (iii) implies that $d^2 = o(n)$. Indeed, if (iii) holds, then $\mathbb{E}Z\mathbb{1}(dZ \leqslant n) = 1 - \mathbb{E}Z\mathbb{1}(dZ > n) = 1 - o(d^{-1})$ and $d^2(1-o(1)) = d^2(\mathbb{E}Z\mathbb{1}(dZ \leqslant n))^2 \leqslant d^2\mathbb{E}Z^2\mathbb{1}(dZ \leqslant n) = o(n)$, hereinafter all limits are with respect to $n \to \infty$. To prove that both (i) and (ii) imply $d^2 = o(n)$, we introduce one more condition:

(iv) $\mathbb{P}(d(S_n/n - 1) < -\varepsilon) \to 0$ for all $\varepsilon > 0$.

Obviously, (ii) $\Rightarrow$ (iv). Also, (i) $\Rightarrow$ (iv). Indeed, if (i) holds, then

$$d(S_n/n - 1) \geqslant d\ln(S_n/n) \geqslant \ln\left(S_n^{(d)}/\binom{n}{d}\right) = o_{\mathbb{P}}(1), \tag{4.4}$$

where we have used Maclaurin's inequality $S_n/n \geqslant \big(S_n^{(d)}/\tbinom{n}{d}\big)^{1/d}$ (see (12.3) in [28]).

Let us prove that (iv) $\Rightarrow$ $d^2 = o(n)$. By Theorem 1 in [10] (or Theorem 1.1 in [11]), $\mathbb{P}(S_n - n \leqslant 1) \geqslant 1/13$ for all n. By standard inequalities for concentration functions (see Theorem 2.22 in [27]), there exists $C > 0$ such that $\mathbb{P}(1 - \lambda \leqslant S_n - n \leqslant 1) \leqslant C(\lambda+1)/\sqrt{n}$ for all $\lambda \geqslant 0$ and n. Taking $\lambda = \gamma\sqrt{n} + 1$ for small enough $\gamma > 0$, we will guarantee that $\mathbb{P}(-\gamma\sqrt{n} \leqslant S_n - n \leqslant 1) \leqslant 1/26$ for all large enough n. As a result, we will get that $\mathbb{P}(S_n - n \leqslant -\gamma\sqrt{n}) \geqslant 1/26 + o(1)$. If (iv) holds, then $\mathbb{P}(S_n - n \leqslant -\varepsilon n/d) \to 0$ for any $\varepsilon > 0$. The latter is possible only if, for any fixed $\varepsilon > 0$, $\varepsilon n/d > \gamma\sqrt{n}$ for all large enough n, i.e. $n \geqslant n_0(\varepsilon)$. The latter means that $\sqrt{n} = o(n/d)$ or, equivalently, $d^2 = o(n)$.

Assume further that $d^2 = o(n)$. Letting $b_n := n/d$, we see that $b_n \to \infty$ and $nb_n^{-2} \to 0$. By classical weak laws of large numbers (see Theorem 2 in [15] with a discussion above it and page 317 in [21]), $(S_n - n)/b_n = o_{\mathbb{P}}(1)$ if and only if

$$n\mathbb{P}(|Z-1| > b_n) + nb_n^{-2}\mathbb{E}(Z-1)^2\mathbb{1}(|Z-1| \leqslant b_n) + nb_n^{-1}|\mathbb{E}(Z-1)\mathbb{1}(|Z-1| \leqslant b_n)| \to 0.$$

The last condition could simplified as follows. As $Z \geqslant 0$ a.s. and $b_n \to \infty$, then $Z > 1$ when $|Z-1| > b_n$ for any large enough n. This and $\mathbb{E}Z = 1$ imply that (for large n)

$$n\mathbb{P}(|Z-1| > b_n) = n\mathbb{P}(Z-1 > b_n) \leqslant nb_n^{-1}\mathbb{E}(Z-1)\mathbb{1}((Z-1) > b_n) =$$

$$= -nb_n^{-1}\mathbb{E}(Z-1)\mathbb{1}((Z-1) \leqslant b_n).$$

We also have that $nb_n^{-2}\mathbb{E}(Z-1)^2\mathbb{1}(|Z-1| \leqslant b_n, Z-1 < 0) \leqslant nb_n^{-2} = o(1)$. Therefore, the above results yield that (ii) $\Leftrightarrow n\mathbb{E}\phi((Z-1)/b_n) \to 0$, where

$$\phi(x) = \frac{x^2}{2}\mathbb{1}(x \in [0,1]) + \Big(x - \frac{1}{2}\Big)\mathbb{1}(x > 1), \quad x \in \mathbb{R}.$$

Using that ϕ is non-decreasing and convex, we derive

$$0 \leqslant \mathbb{E}\phi(Z/b_n) - \mathbb{E}\phi((Z-1)/b_n) \leqslant \mathbb{E}\phi'(Z/b_n)/b_n =$$

$$= b_n^{-2}\mathbb{E}Z\mathbb{1}(Z \leqslant b_n) + b_n^{-1}\mathbb{P}(Z > b_n) \leqslant 2b_n^{-2}\mathbb{E}Z = 2b_n^{-2} = o(n^{-1}).$$

In addition, $\kappa(x)/2 \leqslant \phi(x) \leqslant \kappa(x)$ for $\kappa(x) = x^2\mathbb{1}(x \in [0,1]) + x\mathbb{1}(x > 1)$ and all $x \in \mathbb{R}$. Thus, (ii) $\Leftrightarrow n\mathbb{E}\phi((Z-1)/b_n) \to 0 \Leftrightarrow n\mathbb{E}\phi(Z/b_n) \to 0 \Leftrightarrow$ (iii).

Let us now show that (i) $\Leftrightarrow$ (ii). By the above arguments, we could assume that $d^2 = o(n)$. Also, suppose for a moment that $d \to \infty$. Then, by Lemma 2.5, we have the representation (2.3). Denote further by $\bar{\mathbb{E}}f(Z)$ the empirical mean $n^{-1}\sum_{k=1}^n f(Z_k)$ for any function f. Also, let $\epsilon_n, \varepsilon_n, \delta_n, \gamma_n$ be random sequences tending to 0 in probability, not necessarily the same at each occurrence. It is straightforward to show that

$$\rho \xrightarrow{\mathbb{P}} \infty \quad \text{and} \quad \rho d/n \xrightarrow{\mathbb{P}} 1$$

by noting that

$$\bar{\mathbb{E}}\frac{\rho}{Z+\rho} \xrightarrow{\mathbb{P}} 1 \quad \text{and} \quad \frac{\rho d}{n} = \bar{\mathbb{E}}Z\frac{\rho}{Z+\rho}$$

(see the proof of Lemma 2.5 in arXiv:2111.04296). Using the well-known inequalities $\ln(x) \leqslant x-1$ and $x \leqslant (x+1)\ln(x+1) \leqslant x(2+x)/2$ valid for every $x \geqslant 0$, we see that by the definition of ρ,

$$\frac{d}{n} = \bar{\mathbb{E}}\frac{Z/\rho}{Z/\rho+1} \leqslant \bar{\mathbb{E}}\ln\Big(\frac{Z}{\rho}+1\Big) \leqslant \bar{\mathbb{E}}\frac{(Z/\rho)(2+Z/\rho)}{2(Z/\rho+1)} = \frac{d}{n} + \bar{\mathbb{E}}\frac{Z^2/(2\rho)}{Z+\rho} = \frac{d}{n} + \Big(\frac{1}{2}+\gamma_n\Big)\bar{\mathbb{E}}\frac{dZ^2/n}{Z+\rho}$$

and

$$d\ln\frac{\rho d}{n} = (1+\varepsilon_n)d\Big(\frac{\rho d}{n}-1\Big) = (1+\varepsilon_n)\Big(d\bar{\mathbb{E}}\frac{Z\rho}{Z+\rho} - d\Big) = (1+\varepsilon_n)\Big(d\Big(\frac{S_n}{n}-1\Big) - d\,\bar{\mathbb{E}}\frac{Z^2}{Z+\rho}\Big).$$

As a result, when $d \to \infty$, we get from Lemma 2.5 that for $n \geqslant 1$,

$$(1+\varepsilon_n)\Big(d(S_n/n-1) - d\bar{\mathbb{E}}\frac{Z^2}{Z+\rho}\Big) + \delta_n \leqslant \ln\frac{S_n^{(d)}}{\binom{n}{d}}, \tag{4.5}$$

$$\ln\frac{S_n^{(d)}}{\binom{n}{d}} \leqslant (1+\varepsilon_n)d(S_n/n-1) - (1/2+\epsilon_n)d\bar{\mathbb{E}}\frac{Z^2}{Z+\rho} + \delta_n. \tag{4.6}$$

Suppose (ii) holds and $d \to \infty$. Then (iii) holds and, by $\rho d/n \xrightarrow{\mathbb{P}} 1$,

$$d\bar{\mathbb{E}}\frac{Z^2}{Z+\rho} \leqslant d\bar{\mathbb{E}}Z\mathbb{1}(dZ>n) + \frac{1}{\rho d/n}\frac{d^2}{n}\bar{\mathbb{E}}Z^2\mathbb{1}(dZ \leqslant n) \xrightarrow{\mathbb{P}} 0.$$

Therefore, in view of (4.5) and (4.6), (i) holds.

Suppose (ii) holds and $d \not\to \infty$. We need the following elementary fact:

if $\xi_n \xrightarrow{\mathbb{P}} 0$, then there exist (nonrandom) $\alpha_n \in \mathbb{N}$, $n \geqslant 1$, such that $\alpha_n \to \infty$ and $\alpha_n\xi_n \xrightarrow{\mathbb{P}} 0$.

Therefore, by (ii), there exists $d_* = d_*(n) \in \mathbb{N}$ such that $d(n) \leqslant d_*(n)$ for all n, $d = o(d_*)$ (in particular, $d_* \to \infty$), and (ii) holds for d replaced by d_*. As shown above, the latter implies that (i) holds for d replaced by d_*. By Maclaurin's inequality (see (12.3) in [28]),

$$\frac{S_n}{n} - 1 \geqslant \ln\frac{S_n}{n} \geqslant \frac{1}{d}\ln\frac{S_n^{(d)}}{\binom{n}{d}} \geqslant \frac{1}{d_*}\ln\frac{S_n^{(d_*)}}{\binom{n}{d_*}},$$

where $\ln 0 = -\infty$. Combining the above relations yields (i) (for $d = d(n)$).

Suppose (i) holds and $d \to \infty$. Then, noting that $(z + n/d)/(z+\rho)$ always lies between 1 and $n/(\rho d) = 1 + o_{\mathbb{P}}(1)$ for $z \geqslant 0$, we get from (4.6) that

$$d(S_n/n-1) - (1/2+\varepsilon_n)\bar{\mathbb{E}}\frac{dZ^2}{Z+n/d} \geqslant \delta_n. \tag{4.7}$$

It follows from (4.7) and $\bar{\mathbb{E}}Z^2/(Z+n/d) \geqslant \bar{\mathbb{E}}ZI(dZ>n)/2 + \bar{\mathbb{E}}Z^2I(dZ\leqslant n)/(2n/d)$ that

$$d(S_n/n-1) = d(\bar{\mathbb{E}}Z - \mathbb{E}Z) \geqslant (1/4+\epsilon_n)\Big(d\bar{\mathbb{E}}Z\mathbb{1}(dZ>n) + \frac{d^2}{n}\,\bar{\mathbb{E}}Z^2\mathbb{1}(dZ\leqslant n)\Big) + \delta_n. \tag{4.8}$$

As $\operatorname{Var}((d^2/n)\bar{\mathbb{E}}Z^2\mathbb{1}(dZ \leqslant n)) \leqslant d^2(d/n)^2\mathbb{E}Z^4\mathbb{1}(dZ\leqslant n)/n \leqslant (d^2/n)\mathbb{E}Z^2\mathbb{1}(dZ\leqslant n)$, the Chebyshev inequality yields

$$d(S_n/n-1) \geqslant (1/4+\epsilon_n)\frac{d^2}{n}\,\mathbb{E}Z^2\mathbb{1}(dZ\leqslant n) + O_{\mathbb{P}}(1)\sqrt{\frac{d^2}{n}\,\mathbb{E}Z^2\mathbb{1}(dZ\leqslant n)} + \delta_n.$$

Let us show that $d^2\mathbb{E}Z^2\mathbb{1}(dZ\leqslant n) = O(n)$. Suppose the contrary: for some $n_k \to \infty$ (as $k\to\infty$) and $d_k = d(n_k)$, $(d_k^2/n_k)\mathbb{E}Z^2\mathbb{1}(d_kZ\leqslant n_k) \to \infty$. If so, the right-hand side of the last inequality (with n_k, d_k replacing n, d) goes to infinity in probability. But this is impossible, because by Theorem 1 in [10],

$$\mathbb{P}(d_k(S_{n_k}/n_k - 1) \leqslant d_k/n_k) = \mathbb{P}(S_{n_k} \leqslant n_k + 1) \geqslant \frac{1}{13} \quad \text{for all } k.$$

Thus, $d^2\mathbb{E}Z^2\mathbb{1}(dZ\leqslant n) = O(n)$.

The variables $d(\bar{\mathbb{E}}Z\mathbb{1}(dZ\leqslant n) - \mathbb{E}Z\mathbb{1}(dZ\leqslant n))$ have uniformly bounded second moments not exceeding $(d^2/n)\mathbb{E}Z^2\mathbb{1}(dZ\leqslant n)$ and, hence, are bounded in probability. Therefore, (4.8) yields that

$$(3/4-\epsilon_n)d(\bar{\mathbb{E}}Z\mathbb{1}(dZ>n) - \mathbb{E}Z\mathbb{1}(dZ>n)) \geqslant (1/4+\epsilon_n)d\mathbb{E}Z\mathbb{1}(dZ>n) + O_{\mathbb{P}}(1). \tag{4.9}$$

The latter proves that $d\mathbb{E}Z\mathbb{1}(dZ>n) = O(1)$. Indeed, suppose the contrary: for some $n_k\to\infty$ and $d_k = d(n_k)$, $d_k\mathbb{E}Z\mathbb{1}(d_kZ>n_k)\to\infty$. If so, the right-hand side of (4.9) goes to infinity in probability. But this is impossible, as by Theorem 1 in [10],

$$\mathbb{P}(d(\bar{\mathbb{E}}Z\mathbb{1}(dZ>n) - \mathbb{E}Z\mathbb{1}(dZ>n)) \leqslant (d/n)\mathbb{E}Z\mathbb{1}(dZ>n)) \geqslant \frac{1}{13}$$

and $\mathbb{E}Z\mathbb{1}(dZ>n) \leqslant \mathbb{E}Z = 1$.

Let $\xi_n := d(S_n/n-1)$, $\eta_n := d(\bar{\mathbb{E}}Z\mathbb{1}(dZ \leqslant n)-\mathbb{E}Z\mathbb{1}(dZ \leqslant n))$, and $\zeta_n := d\bar{\mathbb{E}}Z\mathbb{1}(dZ > n)$. Obviously, $\mathbb{E}\xi_n = \mathbb{E}\eta_n = 0$, $\xi_n = \eta_n + \zeta_n - \mathbb{E}\zeta_n$, and $\zeta_n \geqslant 0$ a.s. As we have argued above, $\mathbb{P}(\xi_n < -\varepsilon) \to 0$ for all $\varepsilon > 0$ and $\mathbb{E}\eta_n^2, \mathbb{E}\zeta_n$ are bounded by some constant $K > 0$. To prove (ii), we only need to show that $\mathbb{P}(\xi_n > \varepsilon) \to 0$ for all $\varepsilon > 0$. The well-known formula for the expectation $\mathbb{E}\xi_n = \int_{\mathbb{R}_+} \mathbb{P}(\xi_n > x)dx - \int_{\mathbb{R}_+} \mathbb{P}(\xi_n < -x)dx$ gives

$$\int_0^\infty \mathbb{P}(\xi_n > x)dx = \int_0^\infty \mathbb{P}(\xi_n < -x)dx \leqslant \varepsilon + (M-\varepsilon)\mathbb{P}(\xi_n < -\varepsilon) + \int_M^\infty \mathbb{P}(\xi_n < -x)dx$$

for all $\varepsilon > 0$ and $M \geqslant \varepsilon$. When $x > 2K$, we have $\mathbb{E}\zeta_n - x/2 < 0$ and $\zeta_n - \mathbb{E}\zeta_n > -x/2$ a.s., therefore, for such x, $\xi_n < -x$ implies $\eta_n < -x/2$ and $\mathbb{P}(\xi_n < -x) \leqslant \mathbb{P}(\eta_n < -x/2)$. As a result, taking $M = \max\{2K, \varepsilon, 4K/\varepsilon\}$ and applying Chebyshev's inequality, we get

$$\int_0^\infty \mathbb{P}(\xi_n > x)dx \leqslant \varepsilon + (M-\varepsilon)\mathbb{P}(\xi_n < -\varepsilon) + \int_M^\infty \frac{4K dx}{x^2} \leqslant 2\varepsilon + M\mathbb{P}(\xi_n < -\varepsilon).$$

Thus, $\overline{\lim}_{n\to\infty} \int_0^\infty \mathbb{P}(\xi_n > x)dx \leqslant 2\varepsilon$ for all $\varepsilon > 0$. This is possible iff $\int_0^\infty \mathbb{P}(\xi_n > x)dx \to 0$. Thus, $\mathbb{P}(\xi_n > \varepsilon) \leqslant \varepsilon^{-1}\int_0^\infty \mathbb{P}(\xi_n > x)dx \to 0$ for all $\varepsilon > 0$. We have shown that (i) $\Rightarrow$ (ii) when $d \to \infty$.

Assume that (i) holds and $d \not\to \infty$. To prove (ii), suppose the contrary: (ii) does not hold, i.e. there are $\varepsilon, \delta > 0$ and $(n_k)_{k=1}^\infty$ such that $\mathbb{P}(|d_k(S_{n_k}/n_k - 1)| > \varepsilon) \geqslant \delta$ for all k (here $d_k = d(n_k)$) and $n_k \to \infty$ as $k \to \infty$. Such d_k are unbounded over k (if not, $d_k(S_{n_k}/n_k - 1) \to 0$ a.s. and in probability by the strong law of large numbers). As a result, there is a subsequence $(n_{k_l})_{l=1}^\infty$ such that $d_{k_l} \to \infty$ and $d_{k_l}(S_{n_{k_l}}/n_{k_l} - 1) \not\to 0$ in probability as $l \to \infty$. However, it is shown above that (i) and $d \to \infty$ imply (ii). The same argument shows that the latter holds for the subsequence $(n_{k_l})_{l=1}^\infty$ and $d_{k_l} = d(n_{k_l})$. We get the contradiction. Thus, (i) $\Rightarrow$ (ii). This finishes the proof of the theorem. □

Proof of Proposition 3.2. Let us reformulate the desired result in terms of the empirical spectral measures $\mu_{\widehat{\Sigma}_N}$ and $\mu_{d!\bar{\Sigma}_N}$: $\mu_{\widehat{\Sigma}_N} = \bar{p}\mu_{d!\bar{\Sigma}_N}/p + (p-\bar{p})\delta_0/p$. Denoting by $\mathcal{X}$ the $p \times N$ matrix, whose columns $\mathbf{x}_{p1}, \ldots, \mathbf{x}_{pN}$ form the sample corresponding to $\widehat{\Sigma}_N$. Define $\bar{\mathcal{X}}$ similarly from $\bar{\Sigma}_N$. By such definitions, $\widehat{\Sigma}_N = N^{-1}\mathcal{X}\mathcal{X}^\top$ and $\bar{\Sigma}_N = N^{-1}\bar{\mathcal{X}}\bar{\mathcal{X}}^\top$. In what follows, we will refer to entries of vectors and matrices by using vector indices. The dimension p corresponds to the vector index $(\alpha_1, \ldots, \alpha_d) \in [1,n]^d \cap \mathbb{N}^d$ and the dimension $\bar{p}$ corresponds to the vector index $(\bar{\alpha}_1, \ldots, \bar{\alpha}_d) \in [1,n]^d \cap \mathbb{N}^d$ with $\bar{\alpha}_1 \leqslant \ldots \leqslant \bar{\alpha}_d$.

For any $\mathbf{x}_p$ following the SRT model (Definition 1.1) with the entries $\prod_{k=1}^d X_{\alpha_k}$, $1 \leqslant \alpha_1, \ldots, \alpha_d \leqslant n$, let $\mathbf{z}_p$ be its z-sub-vector in $\mathbb{R}^{\bar{p}}$ formed by the entries $\prod_{k=1}^d X_{\alpha_k}$ with $\alpha_1 \leqslant \ldots \leqslant \alpha_d$. By definition, the $(\alpha_1, \ldots, \alpha_d)$-th entry of $\mathbf{x}_p$ is equal to the $(\alpha_{(1)}, \ldots, \alpha_{(d)})$-th entry of $\mathbf{z}_p$, hereinafter $\alpha_{(1)}, \ldots, \alpha_{(d)}$ are $\alpha_1, \ldots, \alpha_d$ ordered increasingly. Then $\mathcal{X} = A\mathcal{Z}$, where $\mathcal{Z}$ is the $\bar{p} \times N$ matrix, whose columns are z-sub-vectors corresponding to the columns of $\mathcal{X}$, and A is a $p \times \bar{p}$ matrix such that its $(\alpha_1, \ldots, \alpha_d)$-th row is equal to $(0, \ldots, 0, 1, 0, \ldots, 0)$ with 1 on the $(\alpha_{(1)}, \ldots, \alpha_{(d)})$-th place for all $\alpha_1, \ldots, \alpha_d$. For such A, the $\bar{p} \times \bar{p}$ matrix $A^\top A$ is diagonal and its $(\bar{\alpha}_1, \ldots, \bar{\alpha}_d)$-th diagonal entry (here $1 \leqslant \bar{\alpha}_1 \leqslant \ldots \leqslant \bar{\alpha}_d \leqslant n$) is equal to the number of all non-coinciding permutations of $(\bar{\alpha}_1, \ldots, \bar{\alpha}_d)$, i.e. the multinomial coefficient $d!/(d_1! \ldots d_n!)$ with $d_\alpha = \sum_{k=1}^d \mathrm{I}(\alpha = \bar{\alpha}_k)$.

As a result, we have $d!\bar{\mathcal{X}}^\top\bar{\mathcal{X}} = \mathcal{Z}^\top A^\top A\mathcal{Z}$ and

$$p\mu_{N^{-1}\mathcal{X}\mathcal{X}^\top} + (N-p)\delta_0 = N\mu_{N^{-1}\mathcal{X}^\top\mathcal{X}} =$$

$$= N\mu_{N^{-1}\mathcal{Z}^\top A^\top A\mathcal{Z}} = N\mu_{N^{-1}d!\bar{\mathcal{X}}^\top\bar{\mathcal{X}}} = \bar{p}\mu_{d!N^{-1}\bar{\mathcal{X}}\bar{\mathcal{X}}^\top} + (N-\bar{p})\delta_0.$$

The last identity yields the desired result. □

Proof of Theorem 3.3. First, let us derive some preliminary facts. Condition (2.2) implies $d^2 = o(n)$. Therefore,

$$1 \leqslant \binom{n+d-1}{d} \Big/ \binom{n}{d} \leqslant \Big(\frac{n+d}{n-d}\Big)^d = \exp((2+o(1))d^2/(n-d)) \to 1$$

and $q/N \to \rho$ with $q := \binom{n}{d}$. Also, as $\binom{n}{d} \leqslant \frac{n^d}{d!} \leqslant \binom{n+d-1}{d}$, we have

$$\frac{p}{n^d/d!} = \binom{n+d-1}{d} \Big/ \frac{n^d}{d!} \to 1.$$

As a result,

$$\frac{\|\mathbf{x}_p\|^2}{p} = (1+o(1))\frac{d!}{n^d} \sum_{\substack{d_1,\ldots,d_n \geqslant 0 \\ d_1+\ldots+d_n = d}} \prod_{\alpha=1}^{n} \frac{X_\alpha^{2d_\alpha}}{d_\alpha!} = (1+o(1))\Big(\frac{1}{n}\sum_{\alpha=1}^{n} X_\alpha^2\Big)^d.$$

By Theorem 2.4, condition (2.2) implies that $d\big(n^{-1}\sum_{\alpha=1}^n X_\alpha^2 - 1\big) \xrightarrow{\mathbb{P}} 0$. Therefore,

$$\ln\Big(\frac{1}{n}\sum_{\alpha=1}^{n} X_\alpha^2\Big)^d = d\ln\Big(\frac{1}{n}\sum_{\alpha=1}^{n} X_\alpha^2\Big) \xrightarrow{\mathbb{P}} 0 \quad \text{and} \quad \frac{\|\mathbf{x}_p\|^2}{p} \xrightarrow{\mathbb{P}} 1.$$

Assume w.l.o.g. that the last q $(= \binom{n}{d})$ entries of $\mathbf{x}_p$ have the form $\prod_{k=1}^d X_{\alpha_k}$, where $1 \leqslant \alpha_1 < \ldots < \alpha_d \leqslant n$. Let $\mathbf{y}_p$ be the sub-vector of $\mathbf{x}_p$ consisting of the first $p-q$ entries and let $\mathbf{z}_p$ the sub-vector of $\mathbf{x}_p$ consisting of the last q entries. By Theorem 2.4 and (2.2), $\|\mathbf{z}_p\|^2 = q + o_{\mathbb{P}}(q)$, hereinafter $o_{\mathbb{P}}(\cdot)$ is small o in probability. We also have that

$$\frac{\|\mathbf{x}_p\|^2}{p} = \frac{\|\mathbf{y}_p\|^2}{p} + \frac{\|\mathbf{z}_p\|^2}{p} = \frac{\|\mathbf{y}_p\|^2}{p} + (1+o_{\mathbb{P}}(1))\frac{q}{p} = \frac{\|\mathbf{y}_p\|^2}{p} + 1 + o_{\mathbb{P}}(1).$$

Hence,

$$\|\mathbf{y}_p\|^2 = o_{\mathbb{P}}(p). \tag{4.10}$$

In view of Theorem 2.1, to prove (2.1), we need to check (A). Consider a sequence of positive semi-definite matrices $A_p \in \mathbb{R}^{p\times p}$ with $p = \binom{n+d-1}{d}$ and $\|A_p\| = O(1)$ as well as an $\mathbb{R}^p$-valued Gaussian vector $X = X(p)$ that has zero mean, $\mathbb{E}XX^\top = A_p$ and is independent of $\mathbf{x}_p$. Writing X as (Y,Z) and $\mathbf{x}_p$ as $(\mathbf{y}_p, \mathbf{z}_p)$, where $Y, \mathbf{y}_p$ are the corresponding sub-vectors in $\mathbb{R}^{p-q}$ and Z, z_p are the corresponding sub-vectors in $\mathbb{R}^q$. Hence, by the Cauchy-Schwartz inequality,

$$|\mathbf{x}_p^\top A_p \mathbf{x}_p - \operatorname{tr}(A_p)| = |\mathbb{E}((\mathbf{x}_p^\top X)^2|\mathbf{x}_p) - \mathbb{E}X^\top X| \leqslant$$
$$\leqslant |\mathbb{E}((\mathbf{y}_p^\top Y)^2|\mathbf{x}_p) - \mathbb{E}Y^\top Y| + |\mathbb{E}((\mathbf{z}_p^\top Z)^2|\mathbf{x}_p) - \mathbb{E}Z^\top Z| + 2|\mathbb{E}((\mathbf{y}_p^\top Y)(\mathbf{z}_p^\top Z)|\mathbf{x}_p)| \leqslant$$
$$\leqslant |\mathbb{E}((\mathbf{y}_p, Y)^2|\mathbf{x}_p) - \mathbb{E}Y^\top Y| + |\mathbb{E}((\mathbf{z}_p, Z)^2|\mathbf{x}_p) - \mathbb{E}Z^\top Z| + 2\sqrt{\mathbb{E}((\mathbf{y}_p^\top Y)^2|\mathbf{x}_p)\mathbb{E}(\mathbf{z}_p^\top Z)^2|\mathbf{x}_p)}.$$

As a result, denoting by B_p and C_p the matrices $\mathbb{E}YY^\top$ and $\mathbb{E}ZZ^\top$, respectively, we get

$$|\mathbf{x}_p^\top A_p \mathbf{x}_p - \operatorname{tr}(A_p)| \leqslant |\mathbf{y}_p^\top B_p \mathbf{y}_p - \operatorname{tr}(B_p)| + 2\sqrt{\mathbf{y}_p^\top B_p \mathbf{y}_p}\sqrt{\mathbf{z}_p^\top C_p \mathbf{z}_p} + |\mathbf{z}_p^\top C_p \mathbf{z}_p - \operatorname{tr}(C_p)|.$$

The matrices B_p and C_p are positive semi-definite principal sub-matrices of A_p, in particular, $\|B_p\|, \|C_p\| \leqslant \|A_p\| = O(1)$ and $\operatorname{tr}(B_p) = O(p-q) = o(p)$ (as $q/p \to 1$). As was shown in the proof of Theorem 2.3, (2.2) implies that $\mathbf{z}_p^\top C_p \mathbf{z}_p - \operatorname{tr}(C_p) = o_{\mathbb{P}}(q)$ and $\|\mathbf{z}_p\|^2 = q + o_{\mathbb{P}}(q)$. Therefore,

$$|\mathbf{x}_p^\top A_p \mathbf{x}_p - \operatorname{tr}(A_p)| \leqslant O(1)(\|\mathbf{y}_p\|^2 + o(p)) + O_{\mathbb{P}}(1)\|\mathbf{y}_p\|\sqrt{q} + o_{\mathbb{P}}(p) = o_{\mathbb{P}}(p),$$

where we have also used (4.10). This proves (A). The desired result now follows from Theorem 2.1. □

Supplementary material

Proof of Lemma 4.1. Setting $[n]_d := \{i \subseteq \{1,\dots,n\} : |i| = d\}$, fix arbitrary $i_0, j_0 \subseteq [n]_d$ with $|i_0 \cap j_0| = t$. By definition, $\gamma(s,t)$ is the number of pairs (k_0,l_0) such that $k_0, l_0 \subseteq [n]_d$, $|k_0 \cap l_0| = t$, $|\Lambda_3(i_0,j_0,k_0,l_0)|/2 + |\Lambda_4(i_0,j_0,k_0,l_0)| = s$, and $|\Lambda_1(i_0,j_0,k_0,l_0)| = 0$. Let us count such pairs.

Set further $(i,j,k,l,u,v) = (i_0 \setminus j_0, j_0 \setminus i_0, k_0 \setminus l_0, l_0 \setminus k_0, i_0 \cap j_0, k_0 \cap l_0)$. By definition,

$$i,j,u \text{ are pairwise disjoint, } |i| = |j| = d-t, \text{ and } |u| = t \text{ (the same for } k,l,v). \tag{4.11}$$

Recall that $\Lambda_c(i_0,j_0,k_0,l_0)$, $c = 1,\dots,4$, contains all $\alpha \in i_0 \cup j_0 \cup k_0 \cup l_0$ that are covered by exactly c sets among i_0, j_0, k_0, l_0. Therefore,

$$|\Lambda_1| = 0 \quad \text{iff} \quad i \cup j \subseteq k \cup l \cup v \quad \text{and} \quad k \cup l \subseteq i \cup j \cup u,$$

$$|\Lambda_3| = |((i \cup j) \cap v) \cup (u \cap (k \cup l))| = |(i \cup j) \cap v| + |u \cap (k \cup l)|, \quad |\Lambda_4| = |u \cap v|.$$

Put $q := |(i \cup j) \cap v|$ and $r := |u \cap v|$. Let us show that it follows from $|\Lambda_1| = 0$ and $|i \cup j| = |k \cup l|$ that $|(i \cup j) \cap v| = |u \cap (k \cup l)| = q$. The relation $i \cup j \subseteq k \cup l \cup v$ implies that

$$(i \cup j) \setminus v = (i \cup j) \setminus (u \cup v) \subseteq (k \cup l \cup v) \setminus (u \cup v) = (k \cup l) \setminus u.$$

Similarly, $k \cup l \subseteq i \cup j \cup u$ implies that $(k \cup l) \setminus u \subseteq (i \cup j) \setminus v$. This proves that

$$q = |(i \cup j) \cap v| = |i \cup j| - |(i \cup j) \setminus v| = |k \cup l| - |(k \cup l) \setminus u| = |(k \cup l) \cap u|.$$

Combining the above relations gives $|\Lambda_3| = 2q$ and $s = q + r = |(i \cup j) \cap v| + |u \cap v| = |(i \cup j \cup u) \cap v| \leqslant |v| = t$. In particular, this proves that $\gamma(s,t) = 0$ when $s > t$.

Suppose $s \in \{0,\dots,t\}$. For given (i,j,u) satisfying (4.11) and $r \leqslant s$, let us compute the number of triples (k,l,v) satisfying (4.11) and such that $|\Lambda_1| = 0$, $|\Lambda_3| = 2(s-r)$, and $|\Lambda_4| = |u \cap v| = r$, where $\Lambda_c = \Lambda_c(i \cup u, j \cup u, k \cup v, l \cup v)$ for $c = 1,3,4$.

The number of possible choices of $v \cap u$ is $\binom{t}{r}$, as $|u| = t$ and $|u \cap v| = r$.

Given $v \cap u$, the number of possible choices of $v \cap (i \cup j)$ is $\binom{2(d-t)}{s-r}$, as $(i \cup j) \cap u = \varnothing$, $|i \cup j| = 2(d-t)$, and $|v \cap (i \cup j)| = s - r$.

Given $v \cap (i \cup j \cup u)$, the number of possible choices of $v \setminus (i \cup j \cup u)$ is $\binom{n-(2d-t)}{t-s}$, as $|i \cup j \cup u| = 2d - t$ and $|v \setminus (i \cup j \cup u)| = t - s$.

Given v, the number of possible choices of $(k \cup l) \cap u$ is $\binom{t-r}{s-r}$, as $(k \cup l) \cap u = (k \cup l) \cap (u \setminus v)$, $|u \setminus v| = t - r$, and $|(k \cup l) \cap u| = s - r$.

Given v and $(k \cup l) \cap u$, the number of possible choices of (k,l) is $\binom{2(d-t)}{d-t}$, as $k \cap l = \varnothing$ and, because of $|\Lambda_1| = 0$, k and l could be only composed from the elements of the set $((i \cup j) \setminus v) \cup ((k \cup l) \cap u)$, which has $2(d-t)$ elements in total.

Combining the above bounds and varying r, we deduce that

$$\gamma(s,t) = \sum_{r=0}^{s} \binom{t}{r} \binom{2(d-t)}{s-r} \binom{t-r}{s-r} \binom{n-2d+t}{t-s} \binom{2(d-t)}{d-t} \mathbb{1}(s \leqslant t).$$

As is well known, $\binom{2(d-t)}{m} \leqslant 2^{2(d-t)}$ for any m. Also, it follows from $s \leqslant t < d \leqslant n$ that

$$\frac{\binom{n-2d+t}{t-s}}{\binom{n}{d}} = \frac{\frac{(n-2d+t)!}{(n-2d+s)!}}{\frac{(n-d+(t-s))!}{(n-d)!}} \frac{\frac{d!}{(t-s)!}}{\frac{n!}{(n-d+(t-s))!}} = \prod_{z=1}^{t-s} \frac{n-2d+s+z}{n-d+z} \prod_{m=0}^{d-(t-s)-1} \frac{d-m}{n-m} \leqslant$$

$$\leqslant \prod_{m=0}^{d-(t-s)-1} \frac{d-m}{n-m} = \left(\frac{d}{n}\right)^{d-(t-s)} \prod_{m=0}^{d-(t-s)-1} \frac{1-m/d}{1-m/n} \leqslant \left(\frac{d}{n}\right)^{d-(t-s)},$$

where $\prod_{z=1}^{t-s}$ is equal to one if $t = s$. As a result, we get the desired bound

$$\frac{\gamma(s,t)}{M} \leqslant \sum_{r=0}^{s} \binom{t}{r} \binom{t-r}{s-r} = \binom{t}{s} \sum_{r=0}^{s} \binom{s}{r} = 2^s \binom{t}{s} \quad \text{for} \quad M := \frac{2^{4(d-t)} \binom{n}{d}}{(n/d)^{d-(t-s)}} \mathbb{1}(s \leqslant t). \qquad \square$$

References

[1] Adamczak, R.: Logarithmic Sobolev inequalities and concentration of measure for convex functions and polynomial chaoses, *Bull. Pol. Acad. Sci. Math.* **53**, (2005), 221–238. MR2163396

[2] Adamczak, R., and Latala, R.: Tail and moment estimates for chaoses generated by symmetric random variables with logarithmically concave tails, *Ann. Inst. Henri Poincaré Probab. Stat.* **48**, (2012) 1103–1136. MR3052405

[3] Adamczak, R.: Some remarks on the Dozier–Silverstein theorem for random matrices with dependent entries. *Random Matrices: Theory and Applications* **2**, (2013), 1250017. MR3077829

[4] Ambainis, A., Harrow, A.W., and Hastings, M.B.: Random tensor theory: Extending random matrix theory to mixtures of random product states. *Commun. Math. Phys.* **310**, (2012), 25–74. MR2885613

[5] Bai, Z., and Zhou, W.: Large sample covariance matrices without independence structures in columns. *Stat. Sinica* **18**, (2008), 425–442. MR2411613

[6] Bryson, J., Vershynin, R., and Zhao, H.: Marchenko–Pastur law with relaxed independence conditions. *Random Matrices: Theory and Applications* **10**, (2022), Article no. 2150040, 1–28. MR4379545

[7] Collins, B., Yao, J., and Yuan, W.: On spectral distribution of sample covariance matrices from large dimensional and large k-fold tensor products. *Electron. J. Probab.* **27**, Article no. 102, (2022), 1–18. MR4460273

[8] Dembczak-Kołodziejczyk, A., and Lytova, A.: On the empirical spectral distribution for certain models related to sample covariance matrices with different correlations. *Random Matrices: Theory and Applications* **11**, (2022), Article no. 2250030, 1–23. MR4440256

[9] van Es, A. J., and Helmers, R.: Elementary symmetric polynomials of increasing order. *Probab. Theory Related Fields* **80**, (1988), 21–35. MR0970469

[10] Feige, U.: On sums of independent random variables with unbounded variance and estimating the average degree in a graph. *SIAM Journal on Computing* **35**, (2006), 964–984. MR2203734

[11] Garnett, B.: Small deviations of sums of independent random variables. *J. Comb. Theory Ser. A.* **169**, (2020), 105119. MR3981219

[12] Girko, V., and Gupta, A.K.: Asymptotic behavior of spectral function of empirical covariance matrices. *Random Oper. and Stoch. Eqs.* **2**, (1994), 44–60. MR1276248

[13] Götze, F., Sambale, H., and Sinulis, A.: Concentration inequalities for polynomials in α-sub-exponential random variables. *Electronic Journal of Probability* **26**, 1–22. MR4247973

[14] Guédon, O., Lytova, A., Pajor, A., and Pastur, L.: The central limit theorem for linear eigenvalue statistics of the sum of independent random matrices of rank one. *Spectral Theory and Differential Equations. Amer. Math. Soc. Transl. Ser* **2**, (2014), 145–164. MR3307778

[15] Hall, P.: On the rate of convergence in the weak law of large numbers. *Ann. Probab.* **10**, (1982), 374–381. MR0647510

[16] Halász, G., and Szekely, G.J.: On the elementary symmetric polynomials of independent random variables. *Acta Math. Acad. Sci. H.* **28**, (1976), 397–400. MR0423491

[17] El Karoui, N.: Concentration of measure and spectra of random matrices: Applications to correlation matrices, elliptical distributions and beyond. *Annals of Applied Probability* **19**, (2009), 2362–2405. MR2588248

[18] Latala, R.: Estimates of moments and tails of Gaussian chaoses. *Ann. Probab.* **34**, (2006), 2315–2331. MR2294983

[19] Latala, R., and Lochowski, R.: Moment and tail estimates for multidimensional chaos generated by positive random variables with logarithmically concave tails. *Progress in Probability* **56**, (2003), 77–92. MR2073428

[20] Lehec J.: Moments of the Gaussian chaos. In *Seminaire de Probabilites XLIII*, Springer, Berlin, (2011), 327–340. MR2790379

[21] Loeve, M.: Probability Theory. 3rd ed. Van Nostrand, Princeton, (1963). MR0203748

[22] Lytova, A.: Central limit theorem for linear eigenvalue statistics for a tensor product version of sample covariance matrices. *J. Theor. Probab.* **31**, (2018), 1024–1057. MR3803923

[23] Major, P.: The limit behavior of elementary symmetric polynomials of i.i.d. random variables when their order tends to infinity. *Ann. Probab.* **27**, (1999), 1980–2010. MR1742897 MR1742897

[24] Marcenko, V.A., and Pastur, L.A.: Distribution of eigenvalues in certain sets of random matrices. *Mat. Sb. (N.S.)* **72**, (1967), 507–536. MR0208649

[25] Misiakiewicz, T.: Spectrum of inner-product kernel matrices in the polynomial regime and multiple descent phenomenon in kernel ridge regression. *arXiv preprint arXiv:2204.10425*, (2022).

[26] Pajor, A., and Pastur L.: On the limiting empirical measure of eigenvalues of the sum of rank one matrices with log-concave distribution. *Studia Math.* **195**, (2009), 11–29. MR2539559

[27] Petrov, V.V.: *Limit theorems of probability theory.* Clarendon Press, Oxford, (1995). MR1353441

[28] Steele, J.M.: *The Cauchy-Schwarz master class. An introduction to the art of mathematical inequalities.*Cambridge University Press, Cambridge, (2004). MR2062704

[29] Szekely, G.J.: A limit theorem for elementary symmetric polynomials of independent random variables. *Z. Wahrsch. Verw. Gebiete* **59**, (1982), 355–359. MR0721631

[30] Vershynin, R.: Concentration inequalities for random tensors. Bernoulli, 26(4), (2020), 3139–3162. MR4140540

[31] Xiao L, and Pennington J.: Precise learning curves and higher-order scaling limits for dot product kernel regression. *arXiv preprint arXiv:2205.14846*, 2022.

[32] Yaskov, P.: Variance inequalities for quadratic forms with applications. *Math. Methods Statist.* **24**, (2015), 309–319. MR3437388

[33] Yaskov, P.: Necessary and sufficient conditions for the Marchenko-Pastur theorem. *Electron. Commun. Probab.* **21**, Article no. 73, (2016), 1–8. MR3568347

[34] Yaskov, P.: A short proof of the Marchenko–Pastur theorem. *C. R. Math. Acad. Sci. Paris* **354**, (2016), 319–320. MR3463031

[35] Yaskov, P.: LLN for quadratic forms of long memory time series and its applications in random matrix theory. *J. Theor. Probab.* **31**, (2018), 2032–2055. MR3866606

[36] Yaskov, P.: Limiting spectral distribution for large sample covariance matrices with graph-dependent elements. *Theory Probab. Appl.* **67**, (2022), 375–388. MR4578767

Acknowledgments. The author would like to thank an anonymous referee for his/her valuable comments that greatly improved the paper.