

# On the Square Root of Wishart Matrices: Exact Distributions and Asymptotic Gaussian Behavior

Fengcheng Liu\*

Department of Probability and Statistics, School of Mathematical Sciences,  
University of Science and Technology of China

## Abstract

Random matrix theory has become a cornerstone in modern statistics and data science, providing fundamental tools for understanding high-dimensional covariance structures. Within this framework, the Wishart matrix plays a central role in multivariate analysis and related applications. This paper investigates both the exact and asymptotic distributions of the square root of a standard Wishart matrix. We first derive the exact distribution of the square root matrix. Then, by leveraging the Bartlett decomposition, we establish the joint asymptotic normality of the upper-triangular entries of the square root matrix. The resulting limiting distribution resembles that of a scaled Gaussian Wigner ensemble. Additionally, we quantify the rate of convergence using the 1-Wasserstein distance. To validate our theoretical findings, we conduct extensive Monte Carlo simulations, which demonstrate rapid convergence even with relatively low degrees of freedom. These results offer refined insights into the asymptotic behavior of random matrix functionals.

**Keywords:** Wishart distribution, matrix square root, asymptotic normality, convergence rate, Monte Carlo simulation

## 1 Introduction

The Wishart matrix (or ensemble) is a positive-definite random matrix following a Wishart distribution, which arises in a wide range of applications, including multivariate Gaussian modeling, stochastic volatility, random matrix theory, and computational statistics ([Anderson \(2003\)](#); [Jonsson \(1982\)](#)). The Bartlett decomposition([Bartlett \(1934\)](#)) of a Wishart matrix—typically regarded as a non-symmetric square root—expresses the matrix as the product of a lower-triangular matrix and its transpose. While the Bartlett decomposition has proven to be a powerful tool for developing theoretical properties of the Wishart distribution and for generating random positive-definite matrices (e.g., [Cao et al. \(2020\)](#); [Lee and Yoo \(2014\)](#)), the symmetric square root—which preserves the eigenbasis of the Wishart matrix—has received comparatively little attention. This paper addresses this gap by analyzing the exact distribution, the asymptotic distribution, and the convergence rate of the upper-triangular entries of the symmetric square root matrix.

---

\*Email: lfc2022@mail.ustc.edu.cn

To situate our work within the existing literature, we briefly review related results on the Wishart distribution and its matrix square roots. The Wishart distribution plays a central role in multivariate statistics, and its fundamental properties have been well established in classical works(Cao et al. (2020); Muirhead (2005)). It serves as the canonical distribution of sample covariance matrices and has been widely applied in hypothesis testing, estimation, and dimensionality-reduction methods.

In high-dimensional settings, the asymptotic behavior of sample covariance matrices drawn from Wishart laws has been extensively studied (Bai and Silverstein (2004); Johnstone (2001); Ledoit and Wolf (2004)). Much of this literature focuses on spectral quantities such as eigenvalues, traces, and log-determinants, which admit elegant limiting distributions under large-dimension asymptotics. Beyond eigenvalue analysis, recent studies have also examined fluctuations of matrix functionals and random matrix transformations (Couillet and Debbah (2011); Tao (2023)).

Research on matrix square roots of Wishart matrices has received comparatively less attention, though it appears in areas such as matrix decomposition algorithms, stochastic processes on manifolds, and covariance regularization. Notably, the Bartlett decomposition provides a tractable factorization of Wishart matrices that naturally yields expressions for their square roots. Building upon these foundational results, our work aims to derive the exact and asymptotic distributions of the square root of a standard Wishart matrix and to quantify its convergence properties.

We begin by recalling the definition of the Wishart distribution. The Wishart distribution is a probability distribution over symmetric positive-definite matrices. It can be viewed as a multivariate generalization of the chi-squared distribution and plays a fundamental role in multivariate statistics, particularly in the estimation of covariance matrices.

**Definition 1.1.** Let  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_m$  be independent and identically distributed (i.i.d.) random vectors from a multivariate normal distribution with mean vector  $\mathbf{0} \in \mathbb{R}^p$  and covariance matrix  $\Sigma \in \mathbb{R}^{p \times p} > 0$ , i.e.,  $\mathbf{X}_i \sim N_p(\mathbf{0}, \Sigma)$ ,  $i = 1, \dots, m$ . Define

$$W = \sum_{i=1}^m \mathbf{X}_i \mathbf{X}_i^\top.$$

Then  $W$  follows a Wishart distribution with  $m$  degrees of freedom and scale matrix  $\Sigma$ , denoted by

$$W \sim W_p(m, \Sigma).$$

When  $\Sigma = I_p$ , the Wishart distribution  $W_p(m, I_p)$  is often denoted by  $W_p(m)$  and referred to as the standard Wishart distribution. This special case is the primary focus of our study. Throughout this paper, we assume  $m \geq p$  to ensure that  $W$  is almost surely positive definite.

The symmetric square root of  $W$  is defined as the unique positive definite matrix  $V$  satisfying  $W = V^2$ , denoted by  $V = W^{\frac{1}{2}}$ . In many practical applications, it is the symmetric square root  $W^{\frac{1}{2}}$ , rather than the original Wishart matrix  $W$ , that serves as the key quantity of interest. For instance, the square root of a sample covariance matrix (which is proportional to the Wishart square root) is often used to generate samples with a prescribed covariance structure. In wireless communications, the Wishart square root also plays a vital role in optimal precoding for massive MIMO (multiple-input multiple-output) channel modeling.

We then introduce some notation. For any  $k \times k$  real-valued matrix  $G = (g_{sj})_{1 \leq s, j \leq k}$ , We denote

$(g_{11}, g_{22}, \dots, g_{kk}, g_{12}, \dots, g_{1k}, \dots, g_{k-1,k})^\top$  by  $G_{up}$ , and  $(g_{11}, g_{22}, \dots, g_{kk}, g_{21}, \dots, g_{k1}, \dots, g_{k,k-1})^\top$  by  $G_{lw}$ . The Frobenius norm of matrix  $G$  is defined as  $\|G\| = \sqrt{\sum_{s,j=1}^k g_{sj}^2}$ . It is easily verified that, whenever  $G$  is symmetric,  $\|G\|^2$  is equal to the square sum of  $G$ 's eigen values. Besides, for any two  $k \times k$  real-valued matrix  $A$  and  $B$ , the following inequalities hold:  $\|A + B\| \leq \|A\| + \|B\|$  and  $\|AB\| \leq \|A\| \cdot \|B\|$ .

The remainder of this paper is organized as follows. Section 2 derives the exact distribution of  $(W_p(m))^{\frac{1}{2}}$ . Section 3 presents the asymptotic law of  $(W_p(m))^{\frac{1}{2}}$ . Section 4 provides an upper bound for the rate of convergence. Section 5 reports the results of numerical experiments. Section 6 offers concluding remarks and further discussion.

## 2 Exact distribution of $(W_p(m))^{\frac{1}{2}}$

In this section, we derive the density function of  $(W_p(m))^{\frac{1}{2}}$ . We consider a more general case, namely the density of  $(W_p(m, \Sigma))^{\frac{1}{2}}$  for any positive definite  $p \times p$  matrix  $\Sigma$ . To achieve this, we first propose the density of  $W_p(m, \Sigma)$ , as shown in [Proposition 2.1](#).

**Proposition 2.1.** *Suppose that  $W \sim W_p(m, \Sigma)$ . Then,  $W$  has the following density function*

$$f(W) = \frac{1}{2^{mp/2} |\Sigma|^{m/2} \Gamma_p(\frac{m}{2})} |W|^{(m-p-1)/2} \exp\left(-\frac{1}{2} \text{tr}(\Sigma^{-1}W)\right), \quad \forall W \in \mathbb{R}^{p \times p}, W > 0,$$

where  $\Gamma_p(a) = \pi^{p(p-1)/4} \prod_{i=1}^p \Gamma(a - \frac{i-1}{2})$  denotes the multivariate Gamma function for  $a > (p-1)/2$ .

A proof of this proposition can be found in [Muirhead \(2005\)](#). We will compute the Jacobian  $J(W \rightarrow V := W^{1/2})$  of the matrix square-root transformation, based on the result in [Muirhead \(2005\)](#).

**Proposition 2.2.** *Let  $W$  be a real, symmetric, and positive definite  $p \times p$  matrix. Suppose it admits the spectral decomposition  $W = H\Lambda H^\top$ , where  $H \in \mathbb{R}^{p \times p}$  is an orthogonal matrix satisfying  $H^\top H = HH^\top = I_p$ , and  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_p)$  with ordered eigenvalues  $\lambda_1 \geq \dots \geq \lambda_p > 0$ . Then, the Jacobian of the transformation from  $W$  to its eigenvalue-eigenvector pair  $(\Lambda, H)$  is given by*

$$J(W \rightarrow (\Lambda, H)) = c_p \prod_{i < j} (\lambda_i - \lambda_j),$$

where  $c_p = \frac{2^p \pi^{p^2/2}}{\Gamma_p(p/2)}$ .

From this proposition, we derive the Jacobian  $J(W \rightarrow V := W^{1/2})$ , as shown in [Lemma 2.1](#).

**Lemma 2.1.** *Suppose that  $W_{p \times p} > 0$ . Then,*

$$J(W \rightarrow V := W^{1/2}) = \prod_{1 \leq i \leq j \leq p} (\delta_i + \delta_j) = 2^p \prod_{i=1}^p \delta_i \prod_{i < j} (\delta_i + \delta_j),$$

where  $\{\delta_j\}_{j=1}^p$  are the eigenvalues of  $V$ .

*Proof.* Suppose that  $W$  admits the spectral decomposition  $W = H\Lambda H^\top$ , where  $H \in \mathbb{R}^{p \times p}$  is an orthogonal matrix satisfying  $H^\top H = HH^\top = I_p$ , and  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_p)$  with ordered eigenvalues  $\lambda_1 \geq \dots \geq \lambda_p > 0$ . According to [Proposition 2.2](#), we have

$$J(W \rightarrow (\Lambda, H)) = c_p \prod_{i < j} (\lambda_i - \lambda_j), \quad c_p := \frac{2^p \pi^{p^2/2}}{\Gamma_p(p/2)}. \quad (2.1)$$

Let  $\Delta := \text{diag}(\delta_1, \delta_2, \dots, \delta_p) = \Lambda^{\frac{1}{2}} = \text{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \dots, \sqrt{\lambda_p})$ , It follows that

$$J(\Lambda \rightarrow \Delta) = \prod_{j=1}^p \left( \frac{d(\delta_j^2)}{d\delta_j} \right) = 2^p \prod_{j=1}^p \delta_j. \quad (2.2)$$

From the definition of the matrix square root, it follows that  $V = H\Lambda^{\frac{1}{2}}H^\top = H\Delta H^\top$ . By applying [Proposition 2.2](#) once more, we obtain

$$J((H, \Delta) \rightarrow V) = \frac{1}{J(V \rightarrow (H, \Delta))} = \frac{1}{c_p \prod_{i < j} (\delta_i - \delta_j)}. \quad (2.3)$$

Combining (2.1), (2.2) and (2.3),

$$\begin{aligned} J(W \rightarrow V) &= J(W \rightarrow (H, \Lambda)) \times J(\Lambda \rightarrow \Delta) \times J((H, \Delta) \rightarrow V) \\ &= c_p \prod_{i < j} (\delta_i^2 - \delta_j^2) \times (2^p \prod_{i=1}^p \delta_i) \times \frac{1}{c_p \prod_{i < j} (\delta_i - \delta_j)} \\ &= \prod_{i \leq j} (\delta_i + \delta_j). \end{aligned}$$

We complete the proof.  $\square$

The following theorem presents the main result of this section.

**Theorem 2.3.** *Suppose that  $W \sim W_p(m, \Sigma)$ , and let  $V = W^{1/2}$ . Then the probability density function of  $V$  is given by the following expression,*

$$f(V) = \frac{2^p}{2^{mp/2} \Gamma_p(m/2) |\Sigma|^{m/2}} |V|^{m-p} \exp\left(-\frac{1}{2} \text{tr}(\Sigma^{-1} V^2)\right) \prod_{i < j} (\delta_i + \delta_j),$$

where  $\delta_1, \dots, \delta_p$  are the eigenvalues of  $V$ .

*Proof.* Let the density function of  $W \sim W_p(m, \Sigma)$  be denoted by  $p(W)$ . From [Proposition 2.1](#), we have

$$p(W) = \frac{1}{2^{mp/2} |\Sigma|^{m/2} \Gamma_p(\frac{m}{2})} |W|^{(m-p-1)/2} \exp\left(-\frac{1}{2} \text{tr}(\Sigma^{-1} W)\right), \quad \forall W \in \mathbb{R}^{p \times p}, W > 0.$$

Since  $V = W^{\frac{1}{2}}$ , the density function of  $V$  can be expressed as

$$f(V) = p(V^2) J(W \rightarrow V).$$

By applying Lemma 2.1, we obtain

$$J(W \rightarrow V) = \prod_{1 \leq i \leq j \leq p} (\delta_i + \delta_j) = 2^p \prod_{i=1}^p \delta_i \prod_{i < j} (\delta_i + \delta_j) = 2^p |V| \prod_{i < j} (\delta_i + \delta_j).$$

Therefore,

$$\begin{aligned} f(V) &= \frac{1}{2^{mp/2} \Gamma_p(m/2) |\Sigma|^{m/2}} |V|^{m-p-1} \exp\left(-\frac{1}{2} \text{tr}(\Sigma^{-1} V^2)\right) \prod_{i \leq j} (\delta_i + \delta_j) \\ &= \frac{2^p}{2^{mp/2} \Gamma_p(m/2) |\Sigma|^{m/2}} |V|^{m-p} \exp\left(-\frac{1}{2} \text{tr}(\Sigma^{-1} V^2)\right) \prod_{i < j} (\delta_i + \delta_j), \end{aligned}$$

and we complete the proof.  $\square$

**Remark 1.** From Theorem 2.3, we further know that  $V = (W_p(m))^{\frac{1}{2}}$  has the following density function,

$$\frac{2^p}{2^{mp/2} \Gamma_p(m/2)} |V|^{m-p} \exp\left(-\frac{1}{2} \text{tr}(V^2)\right) \prod_{i < j} (\delta_i + \delta_j),$$

where  $\delta_1, \dots, \delta_p$  denote the eigenvalues of  $V$ . However, it is challenging to derive the convergence law of  $V$  directly from this exact density function.

### 3 Asymptotic distribution of $(W_p(m))^{\frac{1}{2}}$

Although the explicit form of the density function of  $(W_p(m))^{\frac{1}{2}}$  is available, the distributional behavior of  $(W_p(m))^{\frac{1}{2}}$  remains largely unexplored. Accordingly, in this section, we investigate the asymptotic distribution of the square root of a Wishart matrix  $W \sim W_p(m)$ .

We begin by introducing some additional notation. For two vectors  $\mathbf{x} = (x_1, \dots, x_k)^\top$  and  $\mathbf{y} = (y_1, \dots, y_k)^\top$  in  $\mathbb{R}^k$ , we write  $\mathbf{x} \leq \mathbf{y}$  if  $x_j \leq y_j$ ,  $\forall 1 \leq j \leq k$ . Any  $\mathbb{R}^k$  valued random vector  $\mathbf{X}$  is then associated with the distribution function  $F_{\mathbf{X}}(\mathbf{x}) = P(\mathbf{X} \leq \mathbf{x})$ . The notions of convergence in distribution and convergence in probability for sequences of random vectors are defined in the following way.

**Definition 3.1.** We say that  $\mathbf{X}_n$  converges in distribution to  $\mathbf{X}$ , denoted as  $\mathbf{X}_n \xrightarrow{d} \mathbf{X}$ , if  $F_{\mathbf{X}_n}(\mathbf{x}) \rightarrow F_{\mathbf{X}}(\mathbf{x})$  at all continuity points of  $F_{\mathbf{X}}$  as  $n \rightarrow \infty$ .

**Definition 3.2.** We say that  $\mathbf{X}_n$  converges in probability to  $\mathbf{X}$ , denoted as  $\mathbf{X}_n \xrightarrow{P} \mathbf{X}$ , if for every  $\epsilon > 0$ ,  $P(\|\mathbf{X}_n - \mathbf{X}\| > \epsilon) \rightarrow 0$ , as  $n \rightarrow \infty$ .

Next, we present our main result on the asymptotic distribution of  $(W_p(m))^{\frac{1}{2}}$ .

**Theorem 3.1.** Fix  $p \in \mathbb{N}^+$  and assume that  $W \sim W_p(m, I_p)$ ,  $m \geq p$ , and

$$V = W^{\frac{1}{2}} = \begin{pmatrix} v_{11} & v_{12} & \cdots & v_{1p} \\ v_{21} & v_{22} & \cdots & v_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ v_{p1} & v_{p2} & \cdots & v_{pp} \end{pmatrix}.$$

Then, as  $m \rightarrow \infty$ , the following result holds

$$(v_{11} - \sqrt{m}, \dots, v_{pp} - \sqrt{m}, v_{12}, \dots, v_{p-1,p})^\top \xrightarrow{d} \mathcal{N}_{p(p+1)/2} \left( \mathbf{0}, \begin{pmatrix} \frac{1}{2}I_p & 0 \\ 0 & \frac{1}{4}I_{p(p-1)/2} \end{pmatrix} \right). \quad (3.1)$$

**Remark 2.** A  $p \times p$  symmetric matrix  $G$  is called a Gaussian Wigner matrix if its diagonal entries are i.i.d  $N(0, 2)$  random variables and its upper-triangle entries are i.i.d  $N(0, 1)$  random variables that are independent with the diagonals. [Theorem 3.1](#) implies

$$W^{\frac{1}{2}} - \sqrt{m}I_p \xrightarrow{d} \frac{1}{2}G.$$

**Remark 3.** Since  $\sqrt{m - \frac{p}{4}} - \sqrt{m} \rightarrow 0$ , it follows that

$$W^{\frac{1}{2}} - \sqrt{m - \frac{p}{4}}I_p \xrightarrow{d} \frac{1}{2}G, \quad (3.2)$$

as a consequence of [Theorem 3.1](#) and [Proposition 3.2](#). Simulation results indicate that (3.2) provides a numerically closer approximation to the normal distribution.

[Theorem 3.1](#) can be established by applying the delta method, based on the known asymptotic distribution of the Wishart matrix ([Muirhead \(2005\)](#)). In this paper, however, we adopt an alternative approach that more readily facilitates the derivation of convergence rates. Specifically, we employ the Bartlett decomposition of the Wishart matrix ([Anderson \(2003\)](#)) together with the Slutsky lemma in high-dimensional settings ([Van der Vaart \(2000\)](#)), as demonstrated in the following two propositions.

**Proposition 3.2.** Let  $k \in \mathbb{N}^+$ . If

$$\mathbf{X}_n \xrightarrow{d} \mathbf{X} \in \mathbb{R}^k, \quad n \rightarrow \infty, \quad (3.3)$$

and

$$\mathbf{Y}_n \xrightarrow{P} \mathbf{c} := (c_1, c_2, \dots, c_k)^\top \in \mathbb{R}^k, \quad (3.4)$$

where  $\mathbf{c}$  is a constant vector, then,

$$\mathbf{X}_n + \mathbf{Y}_n \xrightarrow{d} \mathbf{X} + \mathbf{c}.$$

**Proposition 3.3. (Bartlett decomposition)** If the matrix  $L_m$  satisfies that

$$L_m = \begin{pmatrix} c_1 & 0 & 0 & \dots & 0 \\ n_{21} & c_2 & 0 & \dots & 0 \\ n_{31} & n_{32} & c_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ n_{p1} & n_{p2} & n_{p3} & \dots & c_p \end{pmatrix}, \quad (3.5)$$

where  $c_j^2 \sim \chi^2_{m-j+1}$ ,  $j = 1, 2, \dots, p$ , and  $n_{ij} \sim \mathcal{N}(0, 1)$ ,  $1 \leq j < i \leq p$  are mutually independent,

then

$$L_m L_m^\top \sim W_p(m).$$

Therefore, the distribution of the square root of  $W_p(m)$  can be equivalently characterized by that of  $(L_m L_m^\top)^{\frac{1}{2}}$ . To analyze this, we apply the Taylor expansion method to the square root of a positive definite matrix. Before turning to the matrix case, we first consider the one-dimensional case, which leads to the following lemma.

**Lemma 3.1.** *For any  $n \geq 2$ , we have*

$$\sum_{k=0}^n \binom{\frac{1}{2}}{k} \binom{\frac{1}{2}}{n-k} = 0,$$

where

$$\binom{\frac{1}{2}}{k} := \frac{\frac{1}{2}(\frac{1}{2}-1) \cdots (\frac{1}{2}-k+1)}{k!}, \quad \forall k \geq 1; \quad \binom{\frac{1}{2}}{0} := 1.$$

*Proof.* Applying the Taylor expansion with the Peano form of the remainder yields

$$(1+x)^{\frac{1}{2}} = \sum_{k=0}^n \binom{\frac{1}{2}}{k} x^k + o(x^n), \quad (x \rightarrow 0).$$

Since

$$1+x = \left( \sum_{k=0}^n \binom{\frac{1}{2}}{k} x^k + o(x^n) \right)^2, \quad (x \rightarrow 0),$$

and by the uniqueness of power series expansions, we get

$$\sum_{k=0}^s \binom{\frac{1}{2}}{k} \binom{\frac{1}{2}}{s-k} = 0, \quad s = 2, 3, \dots, n.$$

The proof is complete.  $\square$

Now, we extend Lemma 3.1 to matrix case.

**Lemma 3.2.** *Let  $X$  be a  $p \times p$  symmetric positive definite matrix. Suppose that  $\|X - I\| = \epsilon \leq \frac{1}{2}$ , then*

$$X^{\frac{1}{2}} = I + \frac{1}{2}(X - I) + R_X,$$

where  $R_X$  is a symmetric matrix satisfying

$$\|R_X\| \leq 2\epsilon^2.$$

*Proof.* For any  $\|X - I\| = \epsilon \leq \frac{1}{2}$ , we define  $(X - I)^0 := I$ , and

$$f_N(X) := \sum_{k=0}^N \binom{\frac{1}{2}}{k} (X - I)^k, \quad \forall N \geq 2.$$

Note that

$$\left\| \binom{\frac{1}{2}}{k} (X - I)^k \right\| \leq \| (X - I)^k \| \leq \frac{1}{2^k}, \quad \forall k \geq 1.$$

This implies that,

$$\sup_{l,s \geq N} |(f_l(X) - f_s(X))_{ij}| \leq \sup_{l,s \geq N} \|f_l(X) - f_s(X)\| \leq \frac{1}{2^N}, \quad \forall 1 \leq i, j \leq p, \quad N \geq 2.$$

It follows that, for  $\forall 1 \leq i, j \leq p$ ,  $\{(f_N(X))_{ij}\}_{N=2}^\infty$  is a Cauchy sequence in  $\mathbb{R}$ . Hence, each entry of  $f_N(X)$  converges to a finite limit when  $N$  goes to  $\infty$ , implying that  $f_N(X)$  converges in  $\|\cdot\|$  to a  $p \times p$  matrix, denoted by

$$f_\infty(X) := I + \sum_{k=1}^{\infty} \binom{\frac{1}{2}}{k} (X - I)^k.$$

Since each  $f_N(X)$  is symmetric,  $f_\infty(X)$  is also symmetric. We claim that:

$$f_\infty^2(X) = X, \quad f_\infty(X) > 0 \text{ (positive definite).} \quad (3.6)$$

The first statement of (3.6) follows directly from [Lemma 3.1](#). Indeed, we set

$$a_{X,0} = I; \quad a_{X,k} = \binom{\frac{1}{2}}{k} (X - I)^k, \quad \forall k \geq 1.$$

Then,

$$\sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \|a_{X,j} a_{X,k}\| \leq \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \frac{p}{2^{j+k}} < \infty.$$

Therefore, changing the order of summation does not affect the value of  $\sum_{j=0}^{\infty} \sum_{k=0}^{\infty} a_{X,j} a_{X,k}$ . This leads to

$$f_\infty^2(X) = \left( \sum_{j=0}^{\infty} a_{X,j} \right)^2 = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} a_{X,j} a_{X,k} = \sum_{s=0}^{\infty} \left( \sum_{j=0}^s a_{X,j} a_{X,s-j} \right) = I + X - I = X,$$

where we have applied results established in [Lemma 3.1](#). Next, we verify the second statement of (3.6). For any  $p$ -dimensional vector  $\mathbf{x}$  satisfying  $\|\mathbf{x}\| = 1$ , we have

$$\mathbf{x}^\top f_\infty(X) \mathbf{x} \geq 1 - \sum_{k=1}^{\infty} \left| \binom{\frac{1}{2}}{k} \right| \|X - I\|^k \geq 1 - \frac{1}{4} - \sum_{k=2}^{\infty} \frac{1}{2^k} > 0,$$

which implies that  $f_\infty(X) > 0$ . Therefore, (3.6) holds, and consequently,

$$X^{\frac{1}{2}} = f_\infty(X) = I + \sum_{k=1}^{\infty} \binom{\frac{1}{2}}{k} (X - I)^k.$$

Lastly, from the estimation below,

$$\sum_{k=2}^{\infty} \left\| \binom{\frac{1}{2}}{k} (X - I)^k \right\| \leq \| (X - I)^2 \| \left( \sum_{k=2}^{\infty} \frac{1}{2^{k-2}} \right) \leq 2\epsilon^2,$$

we obtain the desired result.  $\square$

The next lemma characterize the limiting distribution of the  $\chi$  random variable (i.e. the square root of a chi-squared variable), which corresponds to the case  $p = 1$  in [Theorem 3.1](#).

**Lemma 3.3.** *For any fixed  $x \in \mathbb{Z}$ , we have*

$$\sqrt{\chi_{m+x}^2} - \sqrt{m} \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{2}\right), \quad m \rightarrow \infty, \quad (3.7)$$

where, throughout,  $\chi_{m+x}^2$  denotes a random variable following a  $\chi_{m+x}^2$  distribution.

*Proof.* From the Central Limit Theorem, it follows that

$$\frac{\chi_{m+x}^2 - (m+x)}{\sqrt{m+x}} \xrightarrow{d} \mathcal{N}(0, 2), \quad m \rightarrow \infty.$$

By the law of large numbers, we have

$$\frac{\chi_{m+x}^2}{m+x} \xrightarrow{P} 1, \quad \frac{\sqrt{m+x}}{\sqrt{\chi_{m+x}^2 + \sqrt{m+x}}} \xrightarrow{P} \frac{1}{2}, \quad m \rightarrow \infty.$$

and thus, by Slutsky's lemma

$$\sqrt{\chi_{m+x}^2} - \sqrt{m+x} = \frac{\chi_{m+x}^2 - (m+x)}{\sqrt{\chi_{m+x}^2 + \sqrt{m+x}}} \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{2}\right). \quad (3.8)$$

Furthermore, noting that

$$\sqrt{m+x} - \sqrt{m} \rightarrow 0, \quad m \rightarrow \infty, \quad (3.9)$$

and applying Slutsky's lemma once again, we obtain (3.7) from (3.8) and (3.9). Hence, the proof is complete.  $\square$

We now turn to the proof of [Theorem 3.1](#).

*Proof of Theorem 3.1.* Define  $L_m$  as in (3.5), and set

$$T_m := L_m - \sqrt{m}I.$$

Then,  $T_m$  is a lower triangular matrix whose diagonal entries satisfy the following distributional properties:

$$(T_m)_{jj} \sim \sqrt{\chi_{m-j+1}^2} - \sqrt{m}, \quad j = 1, 2, \dots, p,$$

and whose strictly lower-triangular entries follow  $\mathcal{N}(0, 1)$ . Since the entries of  $(L_m)_{lw}$  are independent, it follows that the components of  $(T_m)_{lw}$  are also mutually independent. Now, we apply

[Lemma 3.2](#) to characterize the square root of  $L_m L_m^\top$ . Note that if  $\|T_m\| \leq \frac{\sqrt{m}}{5}$ , then

$$\left\| \frac{T_m + T_m^\top}{\sqrt{m}} + \frac{T_m T_m^\top}{m} \right\| \leq \frac{2\|T_m\|}{\sqrt{m}} + \frac{\|T_m\|^2}{m} < \frac{1}{2}.$$

Therefore, by [Lemma 3.2](#), we have

$$\left( I + \frac{T_m + T_m^\top}{\sqrt{m}} + \frac{T_m T_m^\top}{m} \right)^{\frac{1}{2}} = I + \frac{T_m + T_m^\top}{2\sqrt{m}} + \frac{T_m T_m^\top}{2m} + R,$$

where  $R$  is symmetric and satisfies

$$\|R\| \leq 2 \left( \frac{11}{5} \frac{\|T_m\|}{\sqrt{m}} \right)^2 < 10 \left( \frac{\|T_m\|}{\sqrt{m}} \right)^2.$$

Since  $\|\frac{T_m T_m^\top}{m}\| \leq \left(\frac{\|T_m\|}{\sqrt{m}}\right)^2$ , absorbing the term  $\frac{T_m T_m^\top}{2m}$  into the remainder yields

$$\left( I + \frac{T_m + T_m^\top}{\sqrt{m}} + \frac{T_m T_m^\top}{m} \right)^{\frac{1}{2}} = I + \frac{T_m + T_m^\top}{2\sqrt{m}} + R', \quad \|R'\| \leq 11 \left( \frac{\|T_m\|}{\sqrt{m}} \right)^2, \quad \text{if } \frac{\|T_m\|}{\sqrt{m}} \leq \frac{1}{5}.$$

Direct calculation further shows

$$\begin{aligned} (L_m L_m^\top)^{\frac{1}{2}} - \sqrt{m}I &= ((\sqrt{m}I + T_m)(\sqrt{m}I + T_m^\top))^{\frac{1}{2}} - \sqrt{m}I \\ &= \sqrt{m} \left( I + \frac{T_m + T_m^\top}{\sqrt{m}} + \frac{T_m T_m^\top}{m} \right)^{\frac{1}{2}} - \sqrt{m}I \\ &= 1_{\|T_m\| \leq \frac{\sqrt{m}}{5}} \left( \left( \frac{T_m + T_m^\top}{2} \right) + \sqrt{m}R' \right) + 1_{\|T_m\| > \frac{\sqrt{m}}{5}} ((L_m L_m^\top)^{\frac{1}{2}} - \sqrt{m}I) \\ &:= \frac{T_m + T_m^\top}{2} + 1_{\|T_m\| \leq \frac{\sqrt{m}}{5}} \sqrt{m}R' + 1_{\|T_m\| > \frac{\sqrt{m}}{5}} R'', \end{aligned} \tag{3.10}$$

where

$$R'' = 1_{\|T_m\| > \frac{\sqrt{m}}{5}} \left( (L_m L_m^\top)^{\frac{1}{2}} - \sqrt{m}I - \frac{T_m + T_m^\top}{2} \right).$$

Next, we evaluate  $\mathbb{E}[\|T_m\|^2]$ . For any  $j = 1, 2, \dots, p$ , we have

$$\mathbb{E} [(\chi_{m-j+1}^2 - m)^2] = 2(m - j + 1) + (j - 1)^2,$$

which implies that

$$\mathbb{E} (\sqrt{\chi_{m-j+1}^2} - \sqrt{m})^2 = \mathbb{E} \left( \frac{\chi_{m-j+1}^2 - m}{\sqrt{\chi_{m-j+1}^2} + \sqrt{m}} \right)^2 \leq \mathbb{E} \left( \frac{\chi_{m-j+1}^2 - m}{\sqrt{m}} \right)^2 = \frac{2(m - j + 1) + (j - 1)^2}{m}.$$

Therefore,

$$\begin{aligned}\mathbb{E}[\|T_m\|^2] &\leq \frac{p(p-1)}{2} + \sum_{j=1}^p \frac{2(m-j+1) + (j-1)^2}{m} \\ &< \frac{p^3}{3m} + \frac{p(p+3)}{2} \\ &\leq \frac{5}{6}p^2 + \frac{3}{2}p.\end{aligned}$$

It follows from Chebyshev's Inequality that, as  $m \rightarrow \infty$

$$P\left(\|T_m\| > \frac{\sqrt{m}}{5}\right) \leq \frac{25}{m} \left(\frac{5}{6}p^2 + \frac{3}{2}p\right) \rightarrow 0, \quad 1_{\|T_m\| > \frac{\sqrt{m}}{5}} R''_{up} \xrightarrow{P} \mathbf{0}. \quad (3.11)$$

In addition,

$$\mathbb{E}\left[1_{\|T_m\| \leq \frac{\sqrt{m}}{5}} \sqrt{m} \|R'\|\right] \leq 11\mathbb{E}\left[\frac{\|T_m\|^2}{\sqrt{m}}\right] \leq \frac{11}{\sqrt{m}} \left(\frac{5}{6}p^2 + \frac{3}{2}p\right) \rightarrow 0, \quad m \rightarrow \infty.$$

Since  $\|R'_{up}\| \leq \|R'\|$ , Chebyshev's Inequality further implies that

$$1_{\|T_m\| \leq \frac{\sqrt{m}}{5}} \sqrt{m} R'_{up} \xrightarrow{P} \mathbf{0}. \quad (3.12)$$

Note that the entries of  $\left(\frac{T_m + T_m^\top}{2}\right)_{up}$  are independently distributed. Specifically, the first  $p$  entries, which follow  $\sqrt{\chi_{m-j+1}^2} - \sqrt{m}$  for  $j = 1, 2, \dots, p$ , respectively, converge in distribution to  $\mathcal{N}(0, \frac{1}{2})$ , as established in [Lemma 3.3](#). Moreover, the remaining entries follow  $\mathcal{N}(0, \frac{1}{4})$ . Hence, we conclude that,

$$\left(\frac{T_m + T_m^\top}{2}\right)_{up} \xrightarrow{d} \mathcal{N}_{p(p+1)/2}\left(\mathbf{0}, \begin{pmatrix} \frac{1}{2}I_p & 0 \\ 0 & \frac{1}{4}I_{p(p-1)/2} \end{pmatrix}\right). \quad (3.13)$$

By selecting the upper triangular entries of [\(3.10\)](#), and combining [\(3.11\)](#), [\(3.12\)](#) and [\(3.13\)](#), we obtain

$$\left((L_m L_m^\top)^{\frac{1}{2}} - \sqrt{m} I\right)_{up} \xrightarrow{d} \mathcal{N}_{p(p+1)/2}\left(\mathbf{0}, \begin{pmatrix} \frac{1}{2}I_p & 0 \\ 0 & \frac{1}{4}I_{p(p-1)/2} \end{pmatrix}\right),$$

where we have applied Slutsky's lemma ([Proposition 3.2](#)). Finally, from [Proposition 3.3](#), we arrive at [\(3.1\)](#). The proof is complete.  $\square$

## 4 Convergence rate of $(W_p(m))^{\frac{1}{2}}$

In this section, we evaluate the convergence rate of  $W^{\frac{1}{2}}$ . Specifically, we elaborate on the proof of [Theorem 3.1](#) and derive an upper bound on the convergence rate in terms of the 1-Wasserstein distance, defined as follows.

**Definition 4.1.** For  $\mathbb{R}^d$ -valued random vectors  $\mathbf{X}$  and  $\mathbf{Y}$ , the 1-Wasserstein distance between them

is denoted by

$$d_W(\mathbf{X}, \mathbf{Y}) := \sup_{h: \text{Lip}(h) \leq 1} |\mathbb{E}[h(\mathbf{X}) - h(\mathbf{Y})]|.$$

Some useful properties of this distance are summarized below (we assume that  $\mathbf{X}$ ,  $\mathbf{Y}$ ,  $\mathbf{Z}$ ,  $\mathbf{T}$  are  $\mathbb{R}^d$ -valued random vectors for some  $d \in \mathbb{N}^+$ ):

$$(P_1) : d_W(a\mathbf{X}, a\mathbf{Y}) = |a|d_W(\mathbf{X}, \mathbf{Y}), \quad d_W(\mathbf{x} + \mathbf{X}, \mathbf{x} + \mathbf{Y}) = d_W(\mathbf{X}, \mathbf{Y}), \quad \forall a \in \mathbb{R}, \mathbf{x} \in \mathbb{R}^d.$$

$$(P_2) : d_W(\mathbf{X}, \mathbf{Y}) \leq d_W(\mathbf{X}, \mathbf{Z}) + d_W(\mathbf{Z}, \mathbf{Y}).$$

$$(P_3) : d_W(\mathbf{X} + \mathbf{Z}, \mathbf{Y} + \mathbf{Z}) = \sup_{h: \text{Lip}(h) \leq 1} |\mathbb{E}[h(\mathbf{X} + \mathbf{Z}) - h(\mathbf{Y} + \mathbf{Z})]| \leq \mathbb{E}[\|\mathbf{X} - \mathbf{Y}\|].$$

$$(P_4) : d_W(\mathbf{X} + \mathbf{Y}, \mathbf{Z}) \leq d_W(\mathbf{X}, \mathbf{Z}) + d_W(\mathbf{X} + \mathbf{Y}) \leq d_W(\mathbf{X}, \mathbf{Z}) + \mathbb{E}[\|\mathbf{Y}\|].$$

$$(P_5) : \text{If } \mathbf{X}, \mathbf{Y}, \mathbf{Z} \text{ and } \mathbf{T} \text{ are independent, then } d_W(\mathbf{X} + \mathbf{Y}, \mathbf{Z} + \mathbf{T}) \leq d_W(\mathbf{X}, \mathbf{Z}) + d_W(\mathbf{Y}, \mathbf{T}).$$

$$(P_6) : \text{If } X_1, X_2, \dots, X_k \text{ and } Y_1, Y_2, \dots, Y_k \text{ are independent one-dimensional random variables, then}$$

$$d_W((X_1, X_2, \dots, X_k)^\top, (Y_1, Y_2, \dots, Y_k)^\top) \leq \sum_{j=1}^k d_W(X_j, Y_j).$$

Properties  $(P_1)$  -  $(P_4)$  can be easily verified. Property  $(P_5)$  holds because

$$\begin{aligned} d_W(\mathbf{X} + \mathbf{Y}, \mathbf{Z} + \mathbf{T}) &\leq d_W(\mathbf{X} + \mathbf{Y}, \mathbf{Z} + \mathbf{Y}) + d_W(\mathbf{Z} + \mathbf{Y}, \mathbf{Z} + \mathbf{T}) \\ &= \sup_{h: \text{Lip}(h) \leq 1} |\mathbb{E}[\mathbb{E}[h(\mathbf{X} + \mathbf{Y}) - h(\mathbf{Z} + \mathbf{Y}) | \mathbf{Y}]]| \\ &\quad + \sup_{h: \text{Lip}(h) \leq 1} |\mathbb{E}[\mathbb{E}[h(\mathbf{Z} + \mathbf{Y}) - h(\mathbf{Z} + \mathbf{T}) | \mathbf{Z}]]| \\ &\leq d_W(\mathbf{X}, \mathbf{Z}) + d_W(\mathbf{Y}, \mathbf{T}). \end{aligned}$$

In the final step, we utilize the independence property, which ensures that the conditional distributions of  $(\mathbf{X}, \mathbf{Z}) | \mathbf{Y}$  and  $(\mathbf{Y}, \mathbf{T}) | \mathbf{Z}$  coincide with those of  $(\mathbf{X}, \mathbf{Z})$  and  $(\mathbf{Y}, \mathbf{T})$ , respectively. To prove  $(P_6)$ , note that the sets  $\{(0, \dots, X_j, \dots, 0)\}_{1 \leq j \leq k}$  and  $\{(0, \dots, Y_j, \dots, 0)\}_{1 \leq j \leq k}$  are independent. The result then follows by induction together with  $(P_5)$ .

Now, following the approach used in the proof of [Theorem 3.1](#), we derive an upper bound for the convergence rate of  $W^{\frac{1}{2}}$  in the 1-Wasserstein distance.

**Theorem 4.1.** Suppose that  $m \geq 2p$ , and

$$\mathbf{Z}' \sim \mathcal{N}_{p(p+1)/2} \left( \mathbf{0}, \begin{pmatrix} \frac{1}{2}I_p & 0 \\ 0 & \frac{1}{4}I_{p(p-1)/2} \end{pmatrix} \right).$$

Then,

$$d_W \left( \left( (W_p(m))^{\frac{1}{2}} - \sqrt{m}I \right)_{up}, \mathbf{Z}' \right) < \frac{1}{\sqrt{m}} (42p^{\frac{5}{2}} + 9p^{\frac{9}{4}} + 18p^2 + 75p^{\frac{3}{2}} + 15p^{\frac{5}{4}} + 56p) = O(p^{2.5}/\sqrt{m}).$$

The following lemma provides a bound in the one-dimensional case.

**Lemma 4.1.** Assume that  $X_1 \sim \chi_m^2$ . Then,

$$d_W\left(\sqrt{X_1} - \sqrt{m}, \mathcal{N}\left(0, \frac{1}{2}\right)\right) < \frac{16}{\sqrt{m}}.$$

*Proof.* By applying Lemma 3.2 to  $\frac{X_1}{m}$ , we obtain

$$\sqrt{X_1} - \sqrt{m} = \frac{X_1 - m}{2\sqrt{m}} + R_1,$$

where

$$\begin{aligned} |R_1| &\leq 2\sqrt{m} \left(\frac{X_1 - m}{m}\right)^2, \text{ if } \left|\frac{X_1 - m}{m}\right| \leq \frac{1}{2}, \\ |R_1| &= \left|\sqrt{X_1} - \sqrt{m} - \frac{X_1 - m}{2\sqrt{m}}\right| < \frac{X_1}{2\sqrt{m}}, \text{ if } \frac{X_1}{m} > \frac{3}{2}, \\ |R_1| &= \left|\sqrt{X_1} - \sqrt{m} - \frac{X_1 - m}{2\sqrt{m}}\right| \leq \frac{\sqrt{m}}{2}, \text{ if } \frac{X_1}{m} < \frac{1}{2}. \end{aligned}$$

We present different estimates for three distinct cases, as shown below:

$$\begin{aligned} \mathbb{E}\left[1_{\frac{X_1}{m} > \frac{3}{2}}|R_1|\right] &< \mathbb{E}\left[\frac{X_1}{2\sqrt{m}}1_{\frac{X_1}{m} > \frac{3}{2}}\right] \\ &= \int_{\frac{3}{2}m}^{\infty} \frac{1}{2\sqrt{m}} P(X_1 \geq x) dx + \frac{3\sqrt{m}}{4} P\left(X_1 - m > \frac{m}{2}\right) \\ &\leq \frac{1}{2\sqrt{m}} \int_{\frac{3}{2}m}^{\infty} \frac{\mathbb{E}[(X_1 - m)^2]}{(x - m)^2} dx + \frac{3\sqrt{m}}{4} P\left(X_1 - m > \frac{m}{2}\right) \\ &= \frac{2}{\sqrt{m}} + \frac{3\sqrt{m}}{4} P\left(X_1 - m > \frac{m}{2}\right). \end{aligned} \tag{4.1}$$

In addition,

$$\begin{aligned} \mathbb{E}\left[1_{\left|\frac{X_1 - m}{m}\right| \leq \frac{1}{2}}|R_1|\right] &\leq \mathbb{E}\left[\frac{2\sqrt{m}(X_1 - m)^2}{m^2}\right] = \frac{4}{\sqrt{m}}, \\ \mathbb{E}\left[1_{\frac{X_1}{m} < \frac{1}{2}}|R_1|\right] &\leq \frac{\sqrt{m}}{2} P\left(X_1 < \frac{m}{2}\right) \leq \frac{\sqrt{m}}{2} P\left(X_1 - m \leq -\frac{m}{2}\right). \end{aligned} \tag{4.2}$$

Chebyshev's inequality implies that

$$P\left(X_1 - m > \frac{m}{2}\right) + P\left(X_1 - m < -\frac{m}{2}\right) \leq \frac{\mathbb{E}[(X_1 - m)^2]}{\frac{m^2}{4}} = \frac{8}{m}. \tag{4.3}$$

(4.1), (4.2) and (4.3) together yield

$$\mathbb{E}\left[1_{\frac{X_1}{m} > \frac{3}{2}}|R_1|\right] + \mathbb{E}\left[1_{\frac{X_1}{m} < \frac{1}{2}}|R_1|\right] \leq \frac{8}{\sqrt{m}}. \tag{4.4}$$

From the Wasserstein distance bound in Reinert (1998), we have

$$\begin{aligned} d_W\left(\frac{X_1 - m}{2\sqrt{m}}, \mathcal{N}\left(0, \frac{1}{2}\right)\right) &= \frac{1}{\sqrt{2}}d_W\left(\frac{\chi_m^2 - m}{\sqrt{2m}}, \mathcal{N}(0, 1)\right) \\ &\leq \frac{1}{\sqrt{2m}} \left(2 + \mathbb{E}\left[\left|\frac{\chi_1^2 - 1}{\sqrt{2}}\right|^3\right]\right) \\ &\leq \frac{5.1}{\sqrt{2m}}. \end{aligned}$$

By combining (4.1), (4.2) and (4.4), and applying Property  $(P_4)$ , we obtain

$$\begin{aligned} d_W\left(\sqrt{X_1} - \sqrt{m}, \mathcal{N}\left(0, \frac{1}{2}\right)\right) &\leq d_W\left(\frac{X_1 - m}{2\sqrt{m}}, \mathcal{N}\left(0, \frac{1}{2}\right)\right) + \mathbb{E}[|R_1|] \\ &\leq d_W\left(\frac{X_1 - m}{2\sqrt{m}}, \mathcal{N}\left(0, \frac{1}{2}\right)\right) + \mathbb{E}[1_{\frac{X_1}{m} > \frac{3}{2}}|R_1|] \\ &\quad + \mathbb{E}[1_{\left|\frac{X_1 - m}{m}\right| \leq \frac{1}{2}}|R_1|] + \mathbb{E}\left[1_{\left|\frac{X_1}{m}\right| < \frac{1}{2}}|R_1|\right] \\ &< \frac{16}{\sqrt{m}}. \end{aligned}$$

The proof is complete.  $\square$

*Proof of Theorem 4.1.* We now adopt the same notation as in the proof of Theorem 3.1. Recall that  $T_m$  is a lower triangular matrix with independent entries, whose diagonal elements satisfy  $(T_m)_{jj} \sim \sqrt{\chi_{m-j+1}^2 - \sqrt{m}}$ ,  $\forall j = 1, 2, \dots, p$ , while the strictly lower-triangular entries follow  $\mathcal{N}(0, 1)$ . Recall also that

$$\mathbb{E}[\|T_m\|^2] \leq \frac{5}{6}p^2 + \frac{3}{2}p,$$

and

$$(L_m L_m^\top)^{\frac{1}{2}} - \sqrt{m}I = \frac{T_m + T_m^\top}{2} + 1_{\|T_m\| \leq \frac{\sqrt{m}}{5}} \sqrt{m}R' + 1_{\|T_m\| > \frac{\sqrt{m}}{5}}R''.$$

We have already obtained an estimate for  $\mathbb{E}\left[1_{\|T_m\| \leq \frac{\sqrt{m}}{5}} \sqrt{m}\|R'_{up}\|\right]$  in the proof of Theorem 3.1,

$$\mathbb{E}\left[1_{\|T_m\| \leq \frac{\sqrt{m}}{5}} \sqrt{m}\|R'_{up}\|\right] \leq \frac{11}{\sqrt{m}} \left(\frac{5}{6}p^2 + \frac{3}{2}p\right).$$

Now, we proceed to evaluate  $\mathbb{E}[1_{\|T_m\| > \frac{\sqrt{m}}{5}}\|R''_{up}\|]$ . Note that if  $\|T_m\| > \frac{\sqrt{m}}{5}$ , then

$$\begin{aligned} \|R''\| &= \left\| \sqrt{m} \left( I + \frac{T_m + T_m^\top}{\sqrt{m}} + \frac{T_m T_m^\top}{m} \right)^{\frac{1}{2}} - \sqrt{m}I - \frac{T_m + T_m^\top}{2} \right\| \\ &\leq \sqrt{mp^{\frac{1}{4}}} \left( \frac{\|T_m\|^2}{m} + \frac{2\|T_m\|}{\sqrt{m}} + \sqrt{p} \right)^{\frac{1}{2}} + \sqrt{mp} + \|T_m\| \\ &\leq p^{\frac{1}{4}}\|T_m\| + 2\sqrt{mp} + \|T_m\|. \end{aligned}$$

In the first inequality, we use the following fact: if  $G$  is a  $p \times p$  symmetric positive definite matrix, then  $\|\sqrt{G}\|^2 \leq \sqrt{p}\|G\|$ . This inequality follows directly from the Cauchy–Schwarz inequality. We then compute

$$\begin{aligned} \mathbb{E} \left[ 1_{\|T_m\| > \frac{\sqrt{m}}{5}} \|T_m\| \right] &= \int_{\frac{\sqrt{m}}{5}}^{\infty} P(\|T_m\| \geq x) dx + \frac{\sqrt{m}}{5} P\left(\|T_m\| > \frac{\sqrt{m}}{5}\right) \\ &\leq \int_{\frac{\sqrt{m}}{5}}^{\infty} \frac{\mathbb{E}[\|T_m\|^2]}{x^2} dx + \frac{5}{\sqrt{m}} \left( \frac{5}{6}p^2 + \frac{3}{2}p \right) \\ &\leq \frac{10}{\sqrt{m}} \left( \frac{5}{6}p^2 + \frac{3}{2}p \right), \\ \mathbb{E} \left[ 1_{\|T_m\| > \frac{\sqrt{m}}{5}} \sqrt{m} \right] &= \sqrt{m} P\left(\|T_m\| \geq \frac{\sqrt{m}}{5}\right) \leq \frac{25}{\sqrt{m}} \left( \frac{5}{6}p^2 + \frac{3}{2}p \right). \end{aligned}$$

Therefore,

$$\begin{aligned} \mathbb{E} \left[ 1_{\|T_m\| > \frac{\sqrt{m}}{5}} \|R''_{up}\| \right] &\leq \mathbb{E} \left[ 1_{\|T_m\| > \frac{\sqrt{m}}{5}} \|R''\| \right] \\ &\leq \frac{10}{\sqrt{m}} \left( p^{\frac{1}{4}} + 1 \right) \left( \frac{5}{6}p^2 + \frac{3}{2}p \right) + \frac{50\sqrt{p}}{\sqrt{m}} \left( \frac{5}{6}p^2 + \frac{3}{2}p \right). \end{aligned}$$

From Property  $(P_6)$ , we have

$$\begin{aligned} d_W \left( \left( \frac{T_m + T_m^\top}{2} \right)_{up}, Z' \right) &\leq \sum_{j=1}^p d_W \left( \sqrt{\chi_{m-j+1}^2} - \sqrt{m}, \mathcal{N} \left( 0, \frac{1}{2} \right) \right) \\ &\leq \sum_{j=1}^p \left( (\sqrt{m} - \sqrt{m-j+1}) + \frac{16}{\sqrt{m-j+1}} \right) \\ &< \frac{24p}{\sqrt{m}} + \frac{p^2}{2\sqrt{2m}}, \end{aligned} \tag{4.5}$$

where [Lemma 4.1](#) has been applied. Combining [\(3.10\)](#) and [\(4.5\)](#), we conclude that

$$\begin{aligned} d_W \left( \left( (L_m L_m^\top)^{\frac{1}{2}} - \sqrt{m}I \right)_{up}, Z' \right) &\leq d_W \left( \left( \frac{T_m + T_m^\top}{2} \right)_{up}, Z' \right) \\ &\quad + \mathbb{E} \left[ 1_{\|T_m\| \leq \frac{\sqrt{m}}{5}} \sqrt{m} \|R'_{up}\| + 1_{\|T_m\| > \frac{\sqrt{m}}{5}} \|R''_{up}\| \right] \\ &\leq \frac{24p}{\sqrt{m}} + \frac{p^2}{2\sqrt{2m}} + \frac{11}{\sqrt{m}} \left( \frac{5}{6}p^2 + \frac{3}{2}p \right) \\ &\quad + \frac{10}{\sqrt{m}} \left( p^{\frac{1}{4}} + 1 \right) \left( \frac{5}{6}p^2 + \frac{3}{2}p \right) + \frac{50\sqrt{p}}{\sqrt{m}} \left( \frac{5}{6}p^2 + \frac{3}{2}p \right) \\ &< \frac{1}{\sqrt{m}} \left( 42p^{\frac{5}{2}} + 9p^{\frac{9}{4}} + 18p^2 + 75p^{\frac{3}{2}} + 15p^{\frac{5}{4}} + 56p \right), \end{aligned}$$

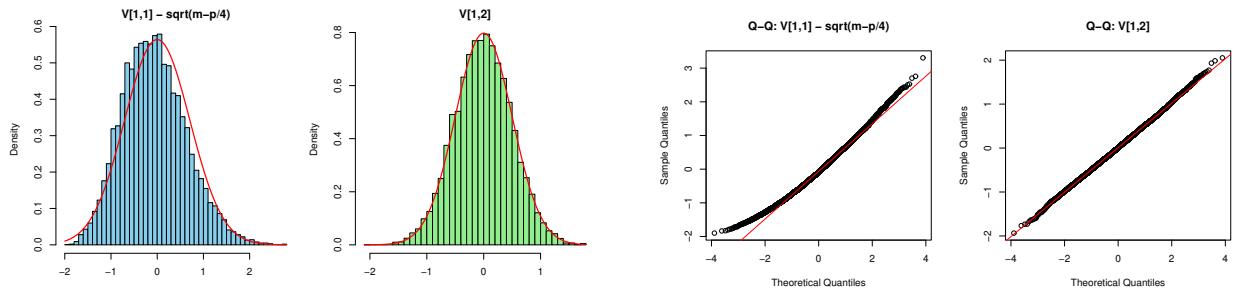
and the desired result follows from [Proposition 3.3](#).  $\square$

## 5 Numerical experiments

In this section, we present a simulation study for  $V := (W_p(m))^{\frac{1}{2}}$ . Specifically, we generate  $N = 10000$  samples from  $W_p(m)$  and compute their corresponding square root matrices via spectral decomposition. For  $p = 2$  and  $m = 5$ , the empirical distribution of  $V_{11} - \sqrt{m - \frac{p}{4}}$  closely matches a  $\mathcal{N}(0, \frac{1}{2})$  distribution, while the off-diagonal entry  $V_{12}$  aligns with  $\mathcal{N}(0, \frac{1}{4})$ , as illustrated in [Figure 1a](#). The Q-Q plots shown in [Figure 1b](#) further support the asymptotic normality of the entries.

The empirical mean values and the covariance matrix of the entries of  $V$  for  $m = 5$ ,  $p = 2$  are reported in [Table 1](#) and [2](#), respectively. The mean values of the diagonal entries are close to  $\sqrt{m - \frac{p}{4}} \approx 2.12$ , while the mean of the off-diagonal entry is close to 0. The empirical variances of the diagonal entries are approximately 0.5, and that of the off-diagonal entry is around 0.25, consistent with the theoretical limiting variances. Furthermore, the covariances of different entries are small, confirming the asymptotic independence predicted by theory.

These results indicate that the proposed approximation performs well even in low-dimensional settings, and remains accurate when  $m$  is only moderately larger than  $p$ .



(a) Comparison of empirical densities of  $V_{11} - \sqrt{m - \frac{p}{4}}$  and  $V_{12}$ , overlaid with  $\mathcal{N}(0, 1/2)$  and  $\mathcal{N}(0, 1/4)$ , respectively

(b) Q-Q plots comparing the empirical distributions of  $V_{11} - \sqrt{m - \frac{p}{4}}$  and  $V_{12}$  to their normal approximations

Figure 1: Distribution and normal Q-Q comparison for  $V_{11} - \sqrt{m - \frac{p}{4}}$  and  $V_{12}$  with  $(m, p) = (5, 2)$

Table 1: Empirical mean values of the entries of  $V = \sqrt{W}$  for  $p = 2$ ,  $m = 5$

	$V_{11}$	$V_{22}$	$V_{12}$
mean	2.069	2.066	-0.011

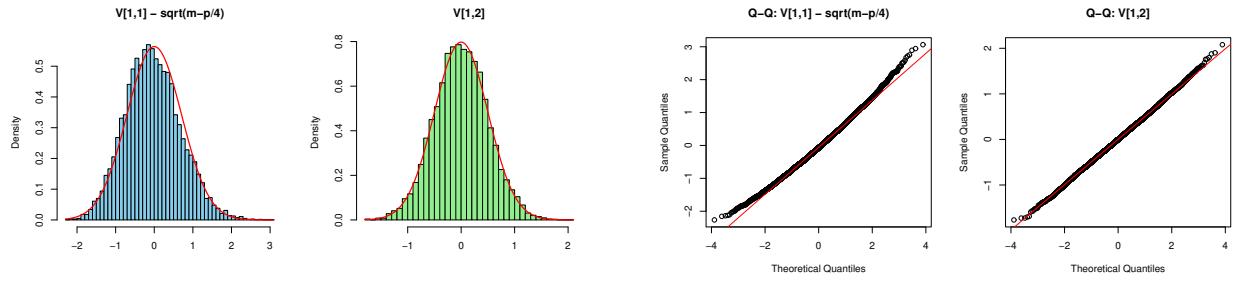
Table 2: Empirical covariance matrix of the entries of  $V = \sqrt{W}$  for  $p = 2$ ,  $m = 5$

	$V_{11}$	$V_{22}$	$V_{12}$
$V_{11}$	0.480	-0.030	0.006
$V_{22}$	-0.030	0.476	0.003
$V_{12}$	0.006	0.003	0.246

We next present numerical experiments for another parameter setting,  $(m, p)$ . [Figure 2a](#) and [Figure 2b](#) present the marginal distributions and normality diagnostics for two representative entries of the matrix  $V = \sqrt{W}$  when  $(m, p) = (12, 5)$ . [Figure 2a](#) shows the empirical density of  $V_{11} -$

$\sqrt{m - \frac{p}{4}}$  and  $V_{12}$ , overlaid with their corresponding theoretical normal density curves,  $\mathcal{N}(0, \frac{1}{2})$  and  $\mathcal{N}(0, \frac{1}{4})$ , respectively. The distribution of  $V_{11} - \sqrt{m - \frac{p}{4}}$  is approximately normal with mean 0 and variance close to 0.5, while  $V_{12}$  closely follows a normal distribution with variance approximately 0.25. The Q-Q plots in Figure 2b further confirm the near-normality of these two random variables.

Tables 3 and 4 report the empirical mean and covariance matrix of selected entries of  $V$ . The means of the diagonal entries  $V_{jj}$ ,  $\forall 1 \leq j \leq p$  are close to  $\sqrt{m - \frac{p}{4}} \approx 3.28$ , while the off-diagonal entries  $V_{1j}$ ,  $\forall 1 < j \leq p$  have means close to 0, as expected. The covariance matrix of the first nine entries supports the approximation that these components behave approximately as independent normal variables, with variances around 0.5 for diagonal elements and 0.25 for off-diagonal elements. These results indicate that the approximation continues to perform very well even as  $p$  increases.



(a) Comparison of empirical densities of  $V_{11} - \sqrt{m - \frac{p}{4}}$  and  $V_{12}$ , overlaid with  $\mathcal{N}(0, 1/2)$  and  $\mathcal{N}(0, 1/4)$ , respectively

(b) Q-Q plots comparing the empirical distributions of  $V_{11} - \sqrt{m - \frac{p}{4}}$  and  $V_{12}$  to their normal approximations

Figure 2: Distribution and normal Q-Q comparison for  $V_{11} - \sqrt{m - \frac{p}{4}}$  and  $V_{12}$  with  $(m, p) = (12, 5)$

Table 3: Empirical mean values of the first nine entries of  $V = \sqrt{W}$  for  $p = 5$ ,  $m = 12$

	$V_{11}$	$V_{22}$	$V_{33}$	$V_{44}$	$V_{55}$	$V_{12}$	$V_{13}$	$V_{14}$	$V_{15}$
mean	3.230	3.241	3.230	3.246	3.232	0.000	-0.001	0.005	-0.003

Table 4: Empirical covariance matrix of the first nine entries of  $V = \sqrt{W}$  for  $p = 5$ ,  $m = 12$

	$V_{11}$	$V_{22}$	$V_{33}$	$V_{44}$	$V_{55}$	$V_{12}$	$V_{13}$	$V_{14}$	$V_{15}$
$V_{11}$	0.501	-0.005	-0.008	-0.010	-0.016	-0.006	-0.009	0.007	-0.002
$V_{22}$	-0.005	0.493	-0.002	-0.002	-0.014	0.004	-0.010	0.002	-0.002
$V_{33}$	-0.008	-0.002	0.497	-0.003	-0.012	-0.005	0.001	-0.001	0.008
$V_{44}$	-0.010	-0.002	-0.003	0.497	-0.005	-0.003	-0.005	0.003	-0.005
$V_{55}$	-0.016	-0.014	-0.012	-0.005	0.497	-0.002	0.001	0.005	-0.002
$V_{12}$	-0.006	0.004	-0.005	-0.003	-0.002	0.260	-0.001	-0.006	-0.001
$V_{13}$	-0.009	-0.010	0.001	-0.005	0.001	-0.001	0.259	0.002	0.002
$V_{14}$	0.007	0.002	-0.001	0.003	0.005	-0.006	0.002	0.254	0.004
$V_{15}$	-0.002	-0.002	0.008	-0.005	-0.002	-0.001	0.002	0.004	0.255

We further examine the empirical covariances between selected entries for other combinations of  $(m, p)$ , as shown in Table 5. Across all cases, the correlations between distinct elements remain

close to zero, confirming that the asymptotic independence persists in higher-dimensional settings.

Table 5: Empirical covariance of certain pairs of entries for different  $(m, p)$

$(m, p)$	$\text{cov}(V_{11}, V_{11})$	$\text{cov}(V_{11}, V_{12})$	$\text{cov}(V_{12}, V_{13})$	$\text{cov}(V_{12}, V_{12})$	$\text{cov}(V_{11}, V_{22})$
(20, 10)	0.522	0.003	-0.003	0.253	-0.009
(100, 50)	0.526	-0.001	0.000	0.256	-0.007
(200, 100)	0.528	-0.009	0.004	0.260	-0.002

## 6 Conclusion

In this paper, we have investigated both the explicit density function and the asymptotic distribution of the (symmetric) square root of a Wishart matrix. We established that it converges to a multivariate normal distribution corresponding to one-half of a Gaussian Wigner matrix. Although similar asymptotic results can be obtained using the delta method, our approach has the distinct advantage of yielding explicit convergence rates, which we show to be of order  $O(p^{2.5}/\sqrt{m})$ . The simulation studies corroborate our theoretical findings, demonstrating that the empirical distributions align closely with the limiting Gaussian law even when the degrees of freedom  $m$  are only moderately larger than  $p$ . This suggests that the asymptotic regime is reached rapidly, enhancing the practical relevance of the proposed approximation. The analytical framework developed here may also be extended to other matrix functionals in high-dimensional statistics. Future research directions include deriving sharper upper bounds for the 1-Wasserstein distance and investigating convergence rates under alternative probabilistic metrics such as the total variation and Kolmogorov distances.

## Acknowledgments

The author thanks Professor Yaning Yang for valuable comments and suggestions.

## References

- Anderson, T. W. (2003). *An introduction to multivariate statistical analysis*. 3rd ed. New York: Wiley.
- Bai, Z. D. and Silverstein, J. W. (2004). *Spectral Analysis of Large Dimensional Random Matrices*. Springer.
- Bartlett, M. S. (1934). On the theory of statistical regression. *Proc. R. Soc. Edinb., Sect. A Math.*, 53, 260-283.
- Cao, X., Khare, K., and Ghosh, M. (2020). Consistent Bayesian sparsity selection for high-dimensional Gaussian DAG models with multiplicative and beta-mixture priors. *J. Multivar. Anal.*, 179, 104628.
- Couillet, R. and Debbah, M. (2011). *Random Matrix Methods for Wireless Communications*. Cambridge University Press.

- Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Stat.*, 29, 295-327.
- Jonsson, D. (1982). Some limit theorems for the eigenvalues of a sample covariance matrix. *J. Multivar. Anal.*, 12(1), 1-38.
- Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *J. Multivar. Anal.*, 88(2), 365-411.
- Lee, K. and Yoo, J. K. (2014). Bayesian Cholesky factor models in random effects covariance matrix for generalized linear mixed models. *Comput. Statist. Data Anal.*, 80, 111-116.
- Muirhead, R. J. (2005). *Aspects of Multivariate Statistical Theory*. Wiley.
- Reinert, G. (1998). Couplings for normal approximations with Stein's method. *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.*, 41, 193-207.
- Tao, T. (2023). *Topics in Random Matrix Theory*. American Mathematical Society.
- Van der Vaart, A. W. (2000). *Asymptotic statistics* (Vol. 3). Cambridge university press.