



# Moment Bounds for Large Autocovariance Matrices Under Dependence

Fang Han<sup>1</sup> · Yicheng Li<sup>1</sup>

Received: 4 October 2018 / Revised: 14 May 2019 / Published online: 13 June 2019  
© Springer Science+Business Media, LLC, part of Springer Nature 2019

## Abstract

The goal of this paper is to obtain expectation bounds for the deviation of large sample autocovariance matrices from their means under weak data dependence. While the accuracy of covariance matrix estimation corresponding to independent data has been well understood, much less is known in the case of dependent data. We make a step toward filling this gap and establish deviation bounds that depend only on the parameters controlling the “intrinsic dimension” of the data up to some logarithmic terms. Our results have immediate impacts on high-dimensional time-series analysis, and we apply them to high-dimensional linear VAR( $d$ ) model, vector-valued ARCH model, and a model used in Banna et al. (Random Matrices Theory Appl 5(2):1650006, 2016).

**Keywords** Autocovariance matrix · Effective rank · Weak dependence ·  $\tau$ -mixing

**Mathematics Subject Classification (2010)** 60E15 · 60F10

## 1 Introduction

Consider a sequence of  $p$ -dimensional mean-zero random vectors  $\{\mathbf{Y}_t\}_{t \in \mathbb{Z}}$  and a size- $n$  fraction  $\{\mathbf{Y}_i\}_{i=1}^n$  of it. This paper aims to establish moment bounds for the spectral norm deviation of lag- $m$  autocovariances of  $\{\mathbf{Y}_i\}_{i=1}^n$ ,  $\widehat{\Sigma}_m := (n - m)^{-1} \sum_{i=1}^{n-m} \mathbf{Y}_i \mathbf{Y}_{i+m}^T$ , from their mean values.

A first result at the origin of such problems concerns product measures, with  $m = 0$  and  $\{\mathbf{Y}_i\}_{i=1}^n$  independent and identically distributed (i.i.d.). For this, Rudelson [26] derived a bound on  $\mathbb{E} \|\widehat{\Sigma}_0 - \mathbb{E} \widehat{\Sigma}_0\|$ , where  $\|\cdot\|$  represents the spectral norm for matrices. The technique is based on symmetrization, and the derived maximal inequality

---

✉ Fang Han  
fanghan@uw.edu

Yicheng Li  
yl83@uw.edu

<sup>1</sup> Department of Statistics, University of Washington, Seattle, WA 98195, USA

ity is a consequence of a concentration inequality on a “symmetrized” version of  $p \times p$  symmetric and deterministic matrices,  $\mathbf{A}_1, \dots, \mathbf{A}_n$  (cf. [24]). That is, for any  $x \geq 0$ ,

$$\mathbb{P}\left(\left\|\sum_{i=1}^n \epsilon_i \mathbf{A}_i\right\| \geq x\right) \leq 2p \exp\{-x^2/(2\sigma^2)\}, \quad \sigma^2 := \left\|\sum_{i=1}^n \mathbf{A}_i^2\right\|, \quad (1.1)$$

where  $\{\epsilon_i\}_{i=1}^n$  are independent and taking values  $\{-1, 1\}$  with equal probability. The applicability of this technique then hinges on the assumption that the data are i.i.d..

Later, [8, 19, 21, 28, 30, 33], among many others, derived different types of deviation bounds for  $\widehat{\Sigma}_0$  under different distributional assumptions. For example, [19] and [8] showed that, for such  $\{\mathbf{Y}_i\}_{i=1}^n$  that are sub-Gaussian and i.i.d.,

$$\mathbb{E}\|\widehat{\Sigma}_0 - \Sigma_0\| \leq C\|\Sigma_0\| \left\{ \sqrt{\frac{r(\Sigma_0) \log(ep)}{n}} + \frac{r(\Sigma_0) \log(ep)}{n} \right\}. \quad (1.2)$$

Here,  $C > 0$  is a universal constant,  $\Sigma_0 := \mathbb{E}\mathbf{Y}_1\mathbf{Y}_1^T$ , and  $r(\Sigma_0) := \text{Tr}(\Sigma_0)/\|\Sigma_0\|$  is termed the “effective rank” [33] where  $\text{Tr}(\mathbf{X}) := \sum_{i=1}^p \mathbf{X}_{i,i}$  for any real  $p \times p$  matrix  $\mathbf{X}$ .

Statistically speaking, Eq. (1.2) is of rich implications. For example, combining (1.2) with Davis–Kahan inequality [11] suggests that the principal component analysis (PCA), a core statistical method whose aim is to recover the leading eigenvectors of  $\Sigma_0$ , could still produce consistent estimators even if the dimension  $p$  is much larger than the sample size  $n$ , as long as the “intrinsic dimension” of the data, quantified by  $r(\Sigma_0)$ , is small enough. See Sect. 1 in [14] for more discussions on the statistical performance of PCA in high dimensions.

The main goal of this paper is to give extensions of the deviation inequality (1.2) to large autocovariance matrices, where the matrices are constructed from a high-dimensional structural time series. Examples of such time series include linear vector autoregressive model of lag  $d$  (VAR( $d$ )), vector-valued autoregressive conditionally heteroscedastic (ARCH) model, and a model used in [3]. The main result is shown in Theorem 2.1 and is nonasymptotic in its nature. This result will have important consequences in a high-dimensional time-series analysis. For example, it immediately yields a new analysis for estimating large covariance matrix [10], a new proof of consistency for Brillinger’s PCA in the frequency domain (cf. Chapter 9 in [7]), and we envision that it could facilitate a new proof of consistency for the PCA procedure proposed in [9].

The rest of the paper is organized as follows. Section 2 characterizes the settings and gives the main concentration inequality for large autocovariance matrices. In Sect. 3, we present applications of our results to some specific time-series models. Proofs of the main results are given in Sect. 4, with more relegated to “Appendix.”

## 2 Main Results

We first introduce the notation that will be used in this paper. Without further specification, we use bold, italic lowercase alphabets to denote vectors, e.g.,  $\mathbf{u} = (u_1, \dots, u_p)^T$  as a  $p$ -dimensional real vector and  $\|\mathbf{u}\|_2$  as its vector  $L_2$  norm. We use bold, uppercase alphabets to denote matrices, e.g.,  $\mathbf{X} = (X_{i,j})$  as a  $p \times p$  real matrix and  $\mathbf{I}_p$  as the  $p \times p$  identity matrix. Throughout the paper, let  $c, c', C, C', C''$  be generic universal positive constants, whose actual values may vary at different locations. For any two sequences of positive numbers  $\{a_n\}, \{b_n\}$ , we denote  $a_n = O(b_n)$  if there exists an universal constant  $C$  such that  $a_n \leq Cb_n$  for all  $n$  large enough. We write  $a_n \asymp b_n$  if both  $a_n = O(b_n)$  and  $b_n = O(a_n)$  hold.

Consider a time series  $\{\mathbf{Y}_t\}_{t \in \mathbb{Z}}$  of  $p$ -dimensional real entries  $\mathbf{Y}_t \in \mathbb{R}^p$  with  $\mathbb{R}, \mathbb{Z}$  denoting the sets of real and integer numbers, respectively. In the sequel, the considered time series does not need to be stationary nor centered, and we are focused on a size- $n$  fraction of it. Without loss of generality, we denote this fraction to be  $\{\mathbf{Y}_i\}_{i=1}^n$ .

As described in “Introduction” section, the case of independent  $\{\mathbf{Y}_i\}_{i=1}^n$  has been discussed in depth in recent years. We are interested here in the time-series setting, and our main emphasis will be to describe nontrivial but easy to verify cases for which Inequality (1.2) still holds. The following four assumptions are accordingly made, with the notations that

$$\mathbb{S}^{p-1} := \{\mathbf{x} \in \mathbb{R}^p : \|\mathbf{x}\|_2 = 1\}, \quad \tilde{\mathbb{S}}^{p-1} := \{\mathbf{x} \in \mathbb{R}^p : |x_1| = \dots = |x_p| = 1\},$$

and

$$\|X\|_{L(p)} := (\mathbb{E}|X|^p)^{1/p}, \quad \|X\|_{\psi_2} := \inf\{k \in (0, \infty) : \mathbb{E}[\exp\{(|X|/k)^2\} - 1] \leq 1\}$$

for any random variable  $X$ .

(A1) Define

$$\kappa_1 := \sup_{t \in \mathbb{Z}} \sup_{\mathbf{u} \in \mathbb{S}^{p-1}} \|\mathbf{u}^T \mathbf{Y}_t\|_{\psi_2} < \infty, \quad \kappa_* := \sup_{t \in \mathbb{Z}} \sup_{\mathbf{v} \in \tilde{\mathbb{S}}^{p-1}} \|\mathbf{v}^T \mathbf{Y}_t\|_{\psi_2} < \infty.$$

Note that  $\kappa_1$  is the supremum taken over vectors in the unit hypersphere, while  $\kappa_*$  is the supremum taken over vectors in the discrete hypercube.

(A2) Assume that there exist some constants  $\gamma_1, \gamma_2, \epsilon > 0$  such that for any integer  $j$ , there exists a sequence of random vectors  $\{\tilde{\mathbf{Y}}_t\}_{t > j}$  which is independent of  $\sigma(\{\mathbf{Y}_t\}_{t \leq j})$ , identically distributed as  $\{\mathbf{Y}_t\}_{t > j}$ , and for any integer  $k \geq j + 1$ ,

$$\|\mathbf{Y}_k - \tilde{\mathbf{Y}}_k\|_2 \| \mathbf{u} \|_{L(1+\epsilon)} \leq \gamma_1 \kappa_1 \exp\{-\gamma_2(k - j - 1)\}.$$

(A3) Assume that there exist some constants  $\gamma_3, \gamma_4, \epsilon > 0$  such that for any integer  $j$ , there exists a sequence of random vectors  $\{\tilde{\mathbf{Y}}_t\}_{t > j}$  which is independent of  $\sigma(\{\mathbf{Y}_t\}_{t \leq j})$ , identically distributed as  $\{\mathbf{Y}_t\}_{t > j}$ , and for any integer  $k \geq j + 1$ ,

$$\sup_{\mathbf{u} \in \mathbb{S}^{p-1}} \|(\mathbf{Y}_k - \tilde{\mathbf{Y}}_k)^T \mathbf{u}\|_{L(1+\epsilon)} \leq \gamma_3 \kappa_1 \exp\{-\gamma_4(k - j - 1)\}.$$

(A4) Assume there exists an universal constant  $c > 0$  such that, for all  $t \in \mathbb{Z}$  and for all  $\mathbf{u} \in \mathbb{R}^p$ ,  $\|\mathbf{u}^T \mathbf{Y}_t\|_{\psi_2}^2 \leq c\mathbb{E}(\mathbf{u}^T \mathbf{Y}_t)^2$ .

Two observations are in order. We first define a generalized “effective rank” as follows:

$$r_* := \kappa_*^2 / \kappa_1^2.$$

It is easy to see the close relationship between  $r_*$  and the effective rank highlighted in (1.2). As  $\mathbf{Y}_t \sim N(\mathbf{0}, \Sigma_0)$ ,  $\kappa_1^2$  and  $\kappa_*^2$  scale at the same orders of  $\|\Sigma_0\|$  and  $\text{Tr}(\Sigma_0)$ , and the same observation applies to all sub-Gaussian distributions with the additional condition (A4), which is identical to Assumption 1 in [19]. As a matter of fact,  $r_*$  could be considered as a natural generalized version of  $r(\Sigma_0)$  without these additional assumptions and is used in our main theorem.

Secondly, we note that Assumptions (A2) and (A3) are characterizing the intrinsic coupling property of the sequence. In practice, such couples can be constructed from time to time. Consider, for example, the following causal shift model,

$$\mathbf{Y}_t = H_t(\xi_t, \xi_{t-1}, \xi_{t-2}, \dots),$$

where  $\{\xi_t\}_{t \in \mathbb{Z}}$  consists of independent elements with values in a measurable space  $\mathcal{X}$  and  $H_t : \mathcal{X}^{\mathbb{Z}^+} \rightarrow \mathbb{R}^p$  is a vector-valued function. Then, it is natural to consider

$$\tilde{\mathbf{Y}}_t = H_t(\xi_t, \dots, \xi_{j+1}, \tilde{\xi}_j, \tilde{\xi}_{j-1}, \dots)$$

for an independent copy  $\{\tilde{\xi}_t\}_{t \in \mathbb{Z}}$  of  $\{\xi_t\}_{t \in \mathbb{Z}}$ .

The following is the main result of this paper.

**Theorem 2.1** (Proof in Sect. 4.1) *Let  $\{\mathbf{Y}_t\}_{t \in \mathbb{Z}}$  be a sequence of random vectors satisfying Assumptions (A1)–(A3) and recall  $r^* = \kappa_*^2 / \kappa_1^2$ . Assume  $\gamma_1 = O(\sqrt{r^*})$  and  $\gamma_3 = O(1)$ . Then, for any integer  $n \geq 2$  and  $0 \leq m \leq n - 1$ , we have*

$$\mathbb{E}\|\hat{\Sigma}_m - \mathbb{E}\hat{\Sigma}_m\| \leq C\kappa_1^2 \left\{ \sqrt{\frac{r_* \log ep}{n-m}} + \frac{r_* \log ep (\log np)^3}{n-m} \right\} \quad (2.1)$$

for some constant  $C$  only depending on  $\epsilon, m, \gamma_2, \gamma_4$ . If in addition,  $\{\mathbf{Y}_t\}_{t \in \mathbb{Z}}$  is a second-order stationary sequence of mean-zero random vectors and Assumption (A4) holds, then

$$\mathbb{E}\|\hat{\Sigma}_m - \mathbb{E}\hat{\Sigma}_m\| \leq C'\|\Sigma_0\| \left\{ \sqrt{\frac{r(\Sigma_0) \log ep}{n-m}} + \frac{r(\Sigma_0) \log ep (\log np)^3}{n-m} \right\}$$

for some constant  $C'$  only depending on  $\epsilon, c, m, \gamma_2, \gamma_4$ .

We first comment on the temporal correlatedness conditions, Assumptions (A2) and (A3). We note that they correspond exactly to the  $\delta$ -measure of dependence introduced

in Chapter 3 of [12], for the sequence  $\{\mathbf{Y}_t\}_{t \in \mathbb{Z}}$  and  $\{\mathbf{u}^T \mathbf{Y}_t\}_{t \in \mathbb{Z}}$ , respectively. In addition, as will be seen soon, our measure of dependence is also very related to the  $\tau$ -measure introduced in [13]. In particular, ours is usually stronger than, but as  $\epsilon \rightarrow 0$  reduces to, the  $\tau$ -measure. Lastly, our conditions are also quite connected to the functional dependence measure in [34], on which many moment inequalities in real space have been established (cf. [18,35]). However, it is still unclear whether a similar matrix Bernstein inequality could be developed under Weibiao Wu's functional dependence condition.

Secondly, we note that one is ready to verify that Inequality (2.1) gives the exact control of the deviation from the mean. Actually, Inequality (2.1) is nearly a strict extension of the results in Lounici [19] and Bunea and Xiao [8] to weak data dependence up to some logarithmic terms. This extension is achieved by applying Theorem 4.3, a concentration inequality for a sequence of weakly dependent random matrices. Theorem 4.3 is an extension of the Bernstein-type inequality for real-valued weakly dependent random variables derived in [23] to dependent random matrices and is a slight extension of the Bernstein-type inequality for a sequence of  $\beta$ -mixing random matrices derived in [3]. In some applications, especially those in high dimensions, verifying the weak dependence condition in Theorem 4.3 is more straightforward than verifying the  $\beta$ -mixing condition in Theorem 1 in [3]. The details of the weak dependence condition are introduced in Sect. 4.1, and Theorem 4.3 is proved in "Appendix."

Admittedly, it is still unclear whether Inequality (2.1) could be further improved under the given conditions. Recently, in a remarkable series of papers [15–17], Koltchinskii and Lounici showed that, for sub-Gaussian independent data, the extra multiplicative  $p$  term on the right-hand side of Inequality (2.1) could be further removed. The proof rests on Talagrand's majorizing measures [29] and a corresponding maximal inequality due to Mendelson [20]. In the most general case, to the authors' knowledge, it is still unknown whether Talagrand's approach could be extent to weakly dependent data, although we conjecture that, under stronger temporal dependence (e.g., geometrically  $\phi$ -mixing) conditions, it is possible to recover Koltchinskii and Lounici's result without resorting to the matrix Bernstein inequality in the proof of Theorem 2.1.

Nevertheless, we make a first step toward eliminating these logarithmic terms via the following theorem. It shows, when assuming a Gaussian sequence is observed, one could further tighten the upper bound in Inequality (2.1) by removing all logarithm factors. The obtained bound is thus tight in view of Theorem 2 in [19] and Theorem 4 in [15].

**Theorem 2.2** (Proof in Sect. 4.2) *Let  $\{\mathbf{Y}_t\}_{t \in \mathbb{Z}}$  be a stationary mean-zero Gaussian sequence that satisfies Assumptions (A2)–(A3) with  $\gamma_1 = O(\sqrt{r(\Sigma_0)})$ ,  $\gamma_3 = O(1)$ , and  $\epsilon > 1$ . Then, for any integer  $n \geq 2$  and  $0 \leq m \leq n - 1$ ,*

$$\mathbb{E} \|\widehat{\Sigma}_m - \Sigma_m\| \leq C \|\Sigma_0\| \left( \sqrt{\frac{r(\Sigma_0)}{n-m}} + \frac{r(\Sigma_0)}{n-m} \right)$$

for some constant  $C > 0$  only depending on  $\epsilon, m, \gamma_2, \gamma_4$ .

In a related track of studies, [2,21,28,30], among many others, explored the optimal scaling requirement in approximating a large covariance matrix for heavy-tailed data. For instance, for i.i.d. data and as  $\Sigma_0$  is identity, Bai and Yin [2] showed that  $\|\hat{\Sigma}_0 - \Sigma_0\|$  will converge to zero in probability as long as  $p/n \rightarrow 0$  and fourth moments exist. Some recent developments further strengthen the moment requirement. These results cannot be compared to ours. In particular, our analysis is focused on characterizing the role of “effective rank,” a term of strong meanings in statistical implications and a feature that cannot be captured using these alternative procedures.

### 3 Applications

In this section, we examine the validity of Assumptions (A1)–(A4) in Sect. 2 under three models, a stable VAR( $d$ ) model, a model proposed by Banna et al. [3], and an ARCH-type model. One shall be aware of examples that are of VAR( $d$ ) or ARCH-type structures but are not  $\alpha$ - or  $\beta$ -mixing (cf. [1]).

We first consider such  $\{Y_t\}_{t \in \mathbb{Z}}$  that is a random sequence generated from VAR( $d$ ) model, i.e.,

$$Y_t = A_1 Y_{t-1} + \cdots + A_d Y_{t-d} + E_t,$$

where  $\{E_t\}_{t \in \mathbb{Z}}$  is a sequence of independent vectors such that for all  $t \in \mathbb{Z}$  and  $u \in \mathbb{R}^p$ ,  $\|u^T E_t\|_{\psi_2} \leq c' \|u^T E_t\|_{L(2)}$  for some universal constant  $c' > 0$ . In addition, assume  $\sup_{t \in \mathbb{Z}} \sup_{u \in \mathbb{S}^{p-1}} \|u^T E_t\|_{\psi_2} < D_1$  for some universal positive constant  $D_1 < \infty$ ,  $\|A_k\| \leq a_k < 1$  for all  $1 \leq k \leq d$ , and  $\sum_{k=1}^d a_k < 1$ , where  $\{a_k\}_{k=1}^d$ ,  $d$  are some universal constants.

Under these conditions, we have the following theorem.

**Theorem 3.1** (Proof in Sect. 4.4) *The above  $\{Y_t\}_{t \in \mathbb{Z}}$  satisfies Assumptions (A1)–(A4) with*

$$\gamma_1 = C(\kappa_*/\kappa_1)(\|\bar{A}\|/\rho_1)^K, \gamma_2 = \log(\rho_1^{-1}), \gamma_3 = C'd(\|\bar{A}\|/\rho_1)^K, \gamma_4 = \log(\rho_1^{-1}).$$

Here, we denote

$$\bar{A} := \begin{bmatrix} a_1 & a_2 & \cdots & a_{d-1} & a_d \\ 1 & 0 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix},$$

$\rho_1$  is a universal constant such that  $\rho(\bar{A}) < \rho_1 < 1$  whose existence is guaranteed by the assumption that  $\sum_{k=1}^d a_k < 1$  (cf. Lemma 4.10 in Sect. 4),  $K$  is some constant only depending on  $\rho_1$ , and  $C, C' > 0$  are some constants only depending on  $\epsilon$ .

We secondly consider the following time-series generation scheme whose corresponding matrix version has been considered by Banna et al. [3]. In detail, let  $\{Y_t\}_{t \in \mathbb{Z}}$  be a random sequence generated by

$$Y_t = W_t E_t,$$

where  $\{E_t\}_{t \in \mathbb{Z}}$  is a sequence of independent random vectors independent of  $\{W_t\}_{t \in \mathbb{Z}}$  such that for all  $t \in \mathbb{Z}$  and  $u \in \mathbb{R}^p$ ,  $\|u^T E_t\|_{\psi_2} \leq c' \|u^T E_t\|_{L(2)}$  for some universal constant  $c' > 0$ . In addition, we assume

$$\sup_{t \in \mathbb{Z}} \sup_{u \in \mathbb{S}^{p-1}} \|u^T E_t\|_{\psi_2} \leq \kappa'_1 \quad \text{and} \quad \sup_{t \in \mathbb{Z}} \sup_{v \in \tilde{\mathbb{S}}^{p-1}} \|v^T E_t\|_{\psi_2} \leq \kappa'_*$$

for some constants  $0 < \kappa'_1, \kappa'_* < \infty$ ,  $\{W_t\}_{t \in \mathbb{Z}}$  is a sequence of uniformly bounded  $\tau$ -mixing random variables such that  $\max_{t \in \mathbb{Z}} |W_t| \leq \kappa_W$ , and

$$\tau(k; \{W_t\}_{t \in \mathbb{Z}}, |\cdot|) \leq \kappa_W \gamma_5 \exp\{-\gamma_6(k-1)\}$$

for some constants  $0 < \gamma_5, \gamma_6, \kappa_W < \infty$  (see “Appendix Sect. A.1” for a detailed introduction to the  $\tau$ -mixing random variables).

**Theorem 3.2** (Proof in Sect. 4.4) *The above  $\{Y_t\}_{t \in \mathbb{Z}}$  satisfies Assumptions (A1)–(A4) with*

$$\gamma_1 = C \kappa'_* \kappa_W \gamma_5^{\frac{1}{1+\epsilon}} / \kappa_1, \gamma_2 = \gamma_6 / (1 + \epsilon), \gamma_3 = C' \kappa'_1 \kappa_W \gamma_5^{\frac{1}{1+\epsilon}} / \kappa_1, \gamma_4 = \gamma_6 / (1 + \epsilon)$$

for some constants  $C, C' > 0$  only depending on  $\epsilon$ .

Lastly, we consider an vector-valued ARCH model with  $\{Y_t\}_{t \in \mathbb{Z}}$  being a random sequence generated by

$$Y_t = \mathbf{A} Y_{t-1} + H(Y_{t-1}) E_t,$$

where  $H : \mathbb{R}^p \rightarrow \mathbb{R}^{p \times p}$  is a matrix-valued function and  $\{E_t\}_{t \in \mathbb{Z}}$  is a sequence of independent random vectors such that

$$\sup_{t \in \mathbb{Z}} \sup_{u \in \mathbb{S}^{p-1}} \|u^T E_t\|_{\psi_2} \leq \kappa'_1 \quad \text{and} \quad \sup_{t \in \mathbb{Z}} \sup_{v \in \tilde{\mathbb{S}}^{p-1}} \|v^T E_t\|_{\psi_2} \leq \kappa'_*$$

for some constants  $0 < \kappa'_1, \kappa'_* < \infty$ . Assume further that  $\|\mathbf{A}\| \leq a_1$  and the function  $H(\cdot)$  satisfies

$$\sup_{u, v \in \mathbb{R}^p} \|H(u) - H(v)\| \leq \frac{a_2}{\kappa'_*} \|u - v\|_2$$

for some universal constant  $a_1 < 1, a_2 > 0$  such that  $a_1 + a_2 < 1$ .

**Theorem 3.3** (Proof in Sect. 4.4) *If the above  $\{Y_t\}_{t \in \mathbb{Z}}$  satisfies Assumption (A1), it satisfies Assumptions (A2), (A3) with*

$$\gamma_1 = C \kappa_* / \kappa_1, \gamma_2 = -\log(a_1 + a_2), \gamma_3 = C' \max(\kappa_* \kappa'_1 / \kappa_1 \kappa'_*, 1), \gamma_4 = \log(a_1 + a_2)^{-1}$$

for some constants  $C, C' > 0$  only depending on  $\epsilon$ . If we further assume the above  $\{\mathbf{Y}_t\}_{t \in \mathbb{Z}}$  to be a stationary sequence and  $\sup_{\mathbf{u} \in \mathbb{R}^p} \|H(\mathbf{u})\| < D_2$  for some universal constant  $D_2 < \infty$ , then  $\{\mathbf{Y}_t\}_{t \in \mathbb{Z}}$  satisfies Assumption (A1).

## 4 Proofs

### 4.1 Proof of Theorem 2.1

**Proof of Theorem 2.1** The proof depends mainly on the following tail probability bound of deviation of the sample covariance from its mean.  $\square$

**Proposition 4.1** (Proof in Sect. 4.1) *Let  $\{\mathbf{Y}_t\}_{t \in \mathbb{Z}}$  be a sequence of random vectors satisfying (A1)–(A3). For any integer  $n \geq 2$ , integer  $0 \leq m \leq n - 2$  and real number  $0 < \delta \leq 1$ , define*

$$M_\delta := C \max \left\{ \left( \frac{\kappa_*}{\kappa_1} \right)^2 \log \frac{n-m}{\delta}, \left( \frac{\kappa_*}{\kappa_1} \right)^2, \frac{2\kappa_*\gamma_1}{\kappa_1} \right\}.$$

Then, for any  $x \geq 0$ ,

$$\begin{aligned} & \mathbb{P} \left[ \|\widehat{\Sigma}_m - \mathbb{E}\widehat{\Sigma}_m\| \geq \kappa_1^2 \{x + \sqrt{\delta/(n-m)}\} \right] \\ & \leq 2p \exp \left\{ - \frac{C'(n-m)^2 x^2}{A_1(n-m) + A_2 M_\delta^2 + A_3(n-m)x M_\delta} \right\} + \delta, \end{aligned}$$

with

$$\begin{aligned} A_1 &:= \frac{\{\kappa_*\gamma_1/\kappa_1 + (\kappa_*/\kappa_1)^2(\gamma_3 + 2m + 1) + 2m + 1\}}{1 - \exp\{-\min(\frac{5+\epsilon}{6\epsilon+10}\gamma_2, \gamma_4)\}}, \quad A_2 := \frac{453^2}{\gamma_2}, \\ A_3 &:= \frac{2 \log(n-m)}{\log 2} \max \left\{ 1, 8m + \frac{48 \log(n-m)p}{\gamma_2} \right\} \end{aligned}$$

for some constants  $C, C' > 0$  only depending on  $\epsilon$ .

Without loss of generality, let  $m = 0$ . Taking  $x = \sqrt{\frac{r_* \log ep}{n}} t$ ,  $\delta = x^{-\gamma}$  for some  $\gamma > 1$ ,  $\gamma_1 = O(\sqrt{r_*})$ , and  $\gamma_3 = O(1)$  in Proposition 4.1, we obtain

$$\mathbb{P} \left( \|\widehat{\Sigma}_0 - \mathbb{E}\widehat{\Sigma}_0\| \geq C_1 \kappa_1^2 \sqrt{\frac{r_* \log ep}{n}} t \right) \leq 2p \exp \left[ - \frac{C_2 (\log ep) t / \left\{ \log \left( \sqrt{\frac{r_* \log ep}{n}} t \right) \right\}^2}{1 + \frac{r_* (\log n)^2}{n} + \sqrt{\frac{r_* \log ep}{n}} t (\log np)^3} \right] + x^{-\gamma}$$

for some constants  $C_1, C_2 > 0$  only depending on  $\epsilon, \gamma_2, \gamma_4$ .



If  $1 + \frac{r_*(\log n)^2}{n} \geq \frac{r_* \log ep(\log np)^6}{n}$ , we have

$$\begin{aligned} \frac{\mathbb{E}\|\widehat{\Sigma}_0 - \mathbb{E}\widehat{\Sigma}_0\|^2}{\left(C_1\kappa_1^2\sqrt{\frac{r_* \log ep}{n}}\right)^2} &\leq 1 + \frac{r_*(\log n)^2}{n} \\ &\quad + \int_{1+\frac{r_*(\log n)^2}{n}}^{\left(1+\frac{r_*(\log n)^2}{n}\right)^2} \frac{r_* \log ep(\log np)^6}{n} 2p \exp \left[ -\frac{C_2(\log ep)t / \left\{ \log \left( \sqrt{\frac{r_* \log ep}{n}} t \right) \right\}^2}{1 + \frac{r_*(\log n)^2}{n}} \right] dt \\ &\quad + \int_{\frac{r_* \log ep(\log np)^6}{n}}^{\infty} \frac{\left(1+\frac{r_*(\log n)^2}{n}\right)^2}{r_* \log ep(\log np)^6} 2p \exp \left[ -\frac{C_2(\log ep)\sqrt{t} / \left\{ \log \left( \sqrt{\frac{r_* \log ep}{n}} t \right) \right\}^2}{\sqrt{\frac{r_* \log ep(\log np)^6}{n}}} \right] dt \\ &\leq C_3 \left( 1 + \frac{r_*(\log n)^2}{n} + \frac{r_* \log ep(\log np)^6}{n} \right). \end{aligned}$$

This gives that

$$\mathbb{E}\|\widehat{\Sigma}_0 - \mathbb{E}\widehat{\Sigma}_0\|^2 \leq C_4\kappa_1^4 \left\{ \frac{r_* \log ep}{n} + \frac{r_*^2(\log ep)^2(\log np)^6}{n^2} \right\}.$$

On the other hand, if  $1 + \frac{r_*(\log n)^2}{n} \leq \frac{r_* \log ep(\log np)^6}{n}$ ,

$$\begin{aligned} \frac{\mathbb{E}\|\widehat{\Sigma}_0 - \mathbb{E}\widehat{\Sigma}_0\|^2}{\left(C_1\kappa_1^2\sqrt{\frac{r_* \log ep}{n}}\right)^2} &\leq \frac{r_* \log ep(\log np)^6}{n} \\ &\quad + \int_{\frac{r_* \log ep(\log np)^6}{n}}^{\infty} \frac{r_* \log ep(\log np)^6}{n} 2p \exp \left[ -\frac{C_2(\log ep)\sqrt{t} / \left\{ \log \left( \sqrt{\frac{r_* \log ep}{n}} t \right) \right\}^2}{\sqrt{\frac{r_* \log ep(\log np)^6}{n}}} \right] dt \\ &\leq C_5 \frac{r_* \log ep(\log np)^6}{n}. \end{aligned}$$

This renders

$$\mathbb{E}\|\widehat{\Sigma}_0 - \mathbb{E}\widehat{\Sigma}_0\|^2 \leq C_5\kappa_1^4 \left\{ \frac{r_*^2(\log ep)^2(\log np)^6}{n^2} \right\}.$$

Combining two cases gives us the final result by using the simple fact that  $\mathbb{E}\|\widehat{\Sigma}_0 - \mathbb{E}\widehat{\Sigma}_0\| \leq (\mathbb{E}\|\widehat{\Sigma}_0 - \mathbb{E}\widehat{\Sigma}_0\|^2)^{\frac{1}{2}}$ . This completes the proof of the first part of Theorem 2.1.

Notice that under Assumptions (A1), (A4), zero-mean, and second-order stationarity, we have  $\kappa_1^2 \asymp \|\Sigma_0\|$  and  $\kappa_*^2 \asymp \text{Tr}(\Sigma_0)$ . Thus, plugging in the first part of Theorem 2.1 finishes the proof.  $\square$

Now we prove Proposition 4.1 under Assumptions (A1)–(A3). In the proof, the cases for covariance and autocovariance matrices are treated separately. In the following, we give a roadmap. The proof of Proposition 4.1 is based on combining a Bernstein-type inequality for the almost surely (a.s.) bounded matrices and a truncation method. The probability bound for the a.s. bounded part (a.k.a., the truncated part) of the random matrix is obtained by employing a Bernstein-type inequality for  $\tau$ -mixing random matrices, which is presented in Theorem 4.3, and some related lemmas (Lemmas 4.4 and 4.5), whose proofs are presented later. The tail part of the random matrix is controlled under the sub-Gaussian Assumption (A1), which uses Lemma 4.2.

In more detail, given a sequence of random vectors  $\{Y_t\}_{t \in \mathbb{Z}}$ , denote  $\mathbf{X}_t := Y_t Y_t^T$  for all  $t \in \mathbb{Z}$ . Then, for any constant  $M > 0$ , we introduce the following “truncated” version of  $\mathbf{X}_t$ :

$$\mathbf{X}_t^M := \frac{M \wedge \|\mathbf{X}_t\|}{\|\mathbf{X}_t\|} \mathbf{X}_t,$$

where  $a \wedge b := \min(a, b)$  for any two real numbers  $a, b$ .

For any integer  $m > 0$ , we denote  $\mathbf{Z}_t^{(m)} := Y_t Y_{t+m}^T$  for all  $t \in \mathbb{Z}$ . For the sake of clarification, the superscript “ $(m)$ ” is dropped when no confusion is possible. Then, the truncated version is

$$\mathbf{Z}_t^M := \frac{M \wedge \|\mathbf{Z}_t\|}{\|\mathbf{Z}_t\|} \mathbf{Z}_t$$

for any  $M > 0$ .

We further define the “variances” for  $\{\mathbf{X}_i^M\}_{i=1}^n$  and  $\{\mathbf{Z}_i^M\}_{i=1}^{n-m}$  as

$$\begin{aligned} v_{\mathbf{X}^M}^2 &:= \sup_{K \subseteq \{1, \dots, n\}} \frac{1}{\text{card}(K)} \lambda_{\max} \left\{ \mathbb{E} \left( \sum_{i \in K} \mathbf{X}_i^M - \mathbb{E} \mathbf{X}_i^M \right)^2 \right\}, \\ v_{\mathbf{Z}^M}^2 &:= \sup_{K \subseteq \{1, \dots, n-m\}} \frac{1}{\text{card}(K)} \left\| \mathbb{E} \left( \sum_{i \in K} \mathbf{Z}_i^M - \mathbb{E} \mathbf{Z}_i^M \right)^2 \right\|. \end{aligned}$$

Here,  $\lambda_{\max}(\mathbf{X})$  and  $\lambda_{\min}(\mathbf{X})$  denote the largest and smallest eigenvalues of  $\mathbf{X}$ , respectively.

**Proof of Proposition 4.1** We first assume  $\kappa_1 = 1$ . We consider two cases.

*Case I* When  $m = 0$ ,  $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$  is a sequence of symmetric random matrices. We have,

$$\begin{aligned} & \mathbb{P} \left\{ \frac{1}{n} \left\| \sum_{i=1}^n (\mathbf{X}_i - \mathbb{E} \mathbf{X}_i) \right\| \geq x \right\} \\ &= \mathbb{P} \left\{ \frac{1}{n} \left\| \sum_{i=1}^n (\mathbf{X}_i - \mathbf{X}_i^M + \mathbf{X}_i^M - \mathbb{E} \mathbf{X}_i^M + \mathbb{E} \mathbf{X}_i^M - \mathbb{E} \mathbf{X}_i) \right\| \geq x \right\} \\ &\leq \mathbb{P} \left\{ \frac{1}{n} \left\| \sum_{i=1}^n (\mathbf{X}_i^M - \mathbb{E} \mathbf{X}_i^M + \mathbb{E} \mathbf{X}_i^M - \mathbb{E} \mathbf{X}_i) \right\| + \frac{1}{n} \left\| \sum_{i=1}^n (\mathbf{X}_i - \mathbf{X}_i^M) \right\| \geq x \right\} \end{aligned}$$

$$\begin{aligned}
&\leq \mathbb{P}\left\{\left\|\sum_{i=1}^n(\mathbf{X}_i^M - \mathbb{E}\mathbf{X}_i^M + \mathbb{E}\mathbf{X}_i^M - \mathbb{E}\mathbf{X}_i)\right\| \geq nx\right\} + \mathbb{P}\left\{\left\|\sum_{i=1}^n(\mathbf{X}_i - \mathbf{X}_i^M)\right\| > 0\right\} \\
&\leq \mathbb{P}\left\{\left\|\sum_{i=1}^n(\mathbf{X}_i^M - \mathbb{E}\mathbf{X}_i^M)\right\| \geq nx - \sum_{i=1}^n \|\mathbb{E}\mathbf{X}_i^M - \mathbb{E}\mathbf{X}_i\|\right\} + \sum_{i=1}^n \mathbb{P}(\mathbf{X}_i \neq \mathbf{X}_i^M) \\
&\leq \mathbb{P}\left[\lambda_{\max}\left\{\sum_{i=1}^n(\mathbf{X}_i^M - \mathbb{E}\mathbf{X}_i^M)\right\} \geq nx - \sum_{i=1}^n \|\mathbb{E}\mathbf{X}_i^M - \mathbb{E}\mathbf{X}_i\|\right] \\
&\quad + \mathbb{P}\left[\lambda_{\min}\left\{\sum_{i=1}^n(\mathbf{X}_i^M - \mathbb{E}\mathbf{X}_i^M)\right\} \leq -nx + \sum_{i=1}^n \|\mathbb{E}\mathbf{X}_i^M - \mathbb{E}\mathbf{X}_i\|\right] + \sum_{i=1}^n \mathbb{P}(\mathbf{X}_i \neq \mathbf{X}_i^M).
\end{aligned} \tag{4.1}$$

We first show that the difference in expectation between the “truncated”  $\mathbf{X}_i^{M_\delta}$  and original one  $\mathbf{X}_i$  can be controlled with the chosen truncation level  $M_\delta$ . For this, we need the following lemma.  $\square$

**Lemma 4.2** (Proof in Sect. 4.3) *Let  $\{\mathbf{Y}_t\}_{t \in \mathbb{Z}}$  be a sequence of  $p$ -dimensional random vectors under Assumption (A1). Then, for all  $t \in \mathbb{Z}$  and for all  $x \geq 0$ ,*

$$\mathbb{P}\{\|\mathbf{Y}_t\|_2^2 \geq 2\kappa_*^2 + 8\kappa_*^2(x + \sqrt{x})\} \leq \exp(-Cx)$$

for some arbitrary constant  $C > 0$ .

By applying Lemma 4.2, we obtain that for all  $i \in \{1, \dots, n\}$ ,

$$\begin{aligned}
\|\mathbb{E}\mathbf{X}_i^{M_\delta} - \mathbb{E}\mathbf{X}_i\| &= \left\|\mathbb{E}\left(1 - \frac{M_\delta}{\|\mathbf{X}_i\|}\right)\mathbf{X}_i \mathbf{1}_{\{\|\mathbf{X}_i\| > M_\delta\}}\right\| \\
&\leq \sup_{\mathbf{u}, \mathbf{v} \in \mathbb{S}^{p-1}} \mathbb{E}|\mathbf{u}^T \mathbf{X}_i \mathbf{v}| \mathbf{1}_{\{\|\mathbf{X}_i\| > M_\delta\}} \\
&\leq \sup_{\mathbf{u}, \mathbf{v} \in \mathbb{S}^{p-1}} \{\mathbb{E}(\mathbf{u}^T \mathbf{Y}_i \mathbf{Y}_i^T \mathbf{v})^2\}^{\frac{1}{2}} \{\mathbb{P}(\|\mathbf{X}_i\| > M_\delta)\}^{\frac{1}{2}} \\
&\leq \sqrt{\delta/n},
\end{aligned}$$

where the last line is followed by Assumption (A1), Lemma 4.2, and the chosen  $M_\delta$ .

The second step heavily depends on a Bernstein-type inequality for  $\tau$ -mixing random matrices. The theorem slightly extends the main theorem of [3] in which the random matrix sequence is assumed to be  $\beta$ -mixing. Its proof is relegated to “Appendix.”

**Theorem 4.3** (Proof in Appendix) *Consider a sequence of real, mean-zero, symmetric  $p \times p$  random matrices  $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$  with  $\|\mathbf{X}_t\| \leq M$  for some positive constant  $M$ . In addition, assume that this sequence is  $\tau$ -mixing (see, “Appendix Sect. A.1” for a detailed introduction to the  $\tau$ -mixing coefficient) with geometric decay, i.e.,*

$$\tau(k; \{\mathbf{X}_t\}_{t \in \mathbb{Z}}, \|\cdot\|) \leq M\psi_1 \exp\{-\psi_2(k-1)\}$$

for some constants  $\psi_1, \psi_2 > 0$ . Denote  $\tilde{\psi}_1 := \max\{p^{-1}, \psi_1\}$ . Then, for any  $x \geq 0$  and any integer  $n \geq 2$ , we have

$$\mathbb{P}\left\{\lambda_{\max}\left(\sum_{i=1}^n \mathbf{X}_i\right) \geq x\right\} \leq p \exp\left\{-\frac{x^2}{8(15^2 n v^2 + 60^2 M^2 / \psi_2) + 2xM\tilde{\psi}(\tilde{\psi}_1, \psi_2, n, p)}\right\},$$

where

$$\begin{aligned} v^2 &:= \sup_{K \subseteq \{1, \dots, n\}} \frac{1}{\text{card}(K)} \lambda_{\max}\left\{\mathbb{E}\left(\sum_{i \in K} \mathbf{X}_i\right)^2\right\} \text{ and } \tilde{\psi}(\tilde{\psi}_1, \psi_2, n, p) \\ &:= \frac{\log n}{\log 2} \max\left\{1, \frac{8 \log(\tilde{\psi}_1 n^6 p)}{\psi_2}\right\}. \end{aligned}$$

In order to apply Theorem 4.3, we need the following two lemmas. Lemma 4.4 is to show that the sequence of “truncated” matrices  $\{\mathbf{X}_t^M\}$  under Assumptions (A1)–(A2) is a  $\tau$ -mixing random sequence with geometric decay. Lemma 4.5 calculates the upper bound for  $v^2$  term in Theorem 4.3 for  $\{\mathbf{X}_t^M\}_{t \in \mathbb{Z}}$ .

**Lemma 4.4** (Proof in Sect. 4.3) *Let  $\{\mathbf{Y}_t\}_{t \in \mathbb{Z}}$  be a sequence of random vectors under Assumptions (A1), (A2). Then,  $\{\mathbf{X}_t^M\}_{t \in \mathbb{Z}}$ ,  $\{\mathbf{X}_t^M - \mathbb{E}\mathbf{X}_t^M\}_{t \in \mathbb{Z}}$ ,  $\{\mathbf{Z}_t^M\}_{t \in \mathbb{Z}}$ , and  $\{\mathbf{Z}_t^M - \mathbb{E}\mathbf{Z}_t^M\}_{t \in \mathbb{Z}}$  are all  $\tau$ -mixing random sequences. Moreover,*

$$\begin{aligned} \tau(k; \{\mathbf{X}_t^M\}_{t \in \mathbb{Z}}, \|\cdot\|) &\leq C\gamma_1\kappa_1\kappa_* \exp\{-\gamma_2(k-1)\}, \\ \tau(k; \{\mathbf{X}_t^M - \mathbb{E}\mathbf{X}_t^M\}_{t \in \mathbb{Z}}, \|\cdot\|) &\leq C\gamma_1\kappa_1\kappa_* \exp\{-\gamma_2(k-1)\}, \\ \tau(k; \{\mathbf{Z}_t^M\}_{t \in \mathbb{Z}}, \|\cdot\|) &\leq C' \exp\{\gamma_2 \min(k, m)\} \max\{\gamma_1\kappa_1\kappa_*, \kappa_*^2\} \exp\{-\gamma_2(k-1)\}, \\ \tau(k; \{\mathbf{Z}_t^M - \mathbb{E}\mathbf{Z}_t^M\}_{t \in \mathbb{Z}}, \|\cdot\|) &\leq C' \exp\{\gamma_2 \min(k, m)\} \max\{\gamma_1\kappa_1\kappa_*, \kappa_*^2\} \exp\{-\gamma_2(k-1)\} \end{aligned}$$

for  $k \geq 1$  and some constants  $C, C' > 0$  only depending on  $\epsilon$ .

**Lemma 4.5** (Proof in Sect. 4.3) *Let  $\{\mathbf{Y}_t\}_{t \in \mathbb{Z}}$  be a sequence of random vectors under Assumptions (A1)–(A3). Take  $M \geq C\gamma_1\kappa_1\kappa_*$  for some constant  $C > 0$  only depending on  $\epsilon$ . Then, we obtain*

$$\begin{aligned} v_{\mathbf{X}^M}^2 &\leq C' \frac{\kappa_1^2 \{\kappa_1^2 + \kappa_1\kappa_*\gamma_1 + \kappa_*^2(\gamma_3 + 2)\}}{1 - \exp\{-\min(\frac{5+\epsilon}{6\epsilon+10}\gamma_2, \gamma_4)\}}, \\ v_{\mathbf{Z}^M}^2 &\leq C'' \frac{\kappa_1^2 \{(2m+1)\kappa_1^2 + \kappa_1\kappa_*\gamma_1 + \kappa_*^2(\gamma_3 + 2m+2)\}}{1 - \exp\{-\min(\frac{5+\epsilon}{6\epsilon+10}\gamma_2, \gamma_4)\}} \end{aligned}$$

for some constants  $C', C'' > 0$  only depending on  $\epsilon$ .

Therefore, by applying Theorems 4.3, Lemmas 4.4 and 4.5 with the chosen  $M_\delta$ , we obtain for any  $x > 0$ ,

$$\mathbb{P}\left[\lambda_{\max}\left\{\frac{1}{n}\sum_{i=1}^n(\mathbf{X}_i^{M_\delta} - \mathbb{E}\mathbf{X}_i^{M_\delta})\right\} \geq x + \sqrt{\delta/n}\right] \leq p \exp\left(-\frac{n^2x^2}{A_1n + A_2M_\delta^2 + A_3nxM_\delta}\right), \quad (4.2)$$

where

$$A_1 := \frac{C\{\kappa_*\gamma_1 + \kappa_*^2(\gamma_3 + 2) + 1\}}{1 - \exp\{-\min(\frac{5+\epsilon}{6\epsilon+10}\gamma_2, \gamma_4)\}}, \quad A_2 := \frac{453^2}{\gamma_2}, \quad \text{and } A_3 := \frac{2 \log n}{\log 2} \max\left\{1, \frac{48 \log(np)}{\gamma_2}\right\}$$

for some constant  $C > 0$  only depending on  $\epsilon$ .

Similarly, notice that  $\lambda_{\min}(\sum_{j=1}^n \mathbf{X}_j^{M_\delta}) = \lambda_{\max}(-\sum_{j=1}^n \mathbf{X}_j^{M_\delta})$ . Hence, the same argument renders the same upper bound

$$\mathbb{P}\left[\lambda_{\min}\left\{\frac{1}{n}\sum_{i=1}^n(\mathbf{X}_i^{M_\delta} - \mathbb{E}\mathbf{X}_i^{M_\delta})\right\} \leq -(x + \sqrt{\delta/n})\right] \leq p \exp\left(-\frac{n^2x^2}{A_1n + A_2M_\delta^2 + A_3nxM_\delta}\right) \quad (4.3)$$

with the same constants as above.

For the last term of (4.1), with the choice of  $M_\delta$  and Lemma 4.2, we obtain

$$\sum_{i=1}^n \mathbb{P}(\mathbf{X}_i \neq \mathbf{X}_i^{M_\delta}) = \sum_{i=1}^n \mathbb{P}(\|\mathbf{X}_i\| > M_\delta) \leq \delta. \quad (4.4)$$

Combining (4.2), (4.3), and (4.4), we obtain

$$\mathbb{P}(\|\widehat{\Sigma}_0 - \mathbb{E}\widehat{\Sigma}_0\| \geq x + \sqrt{\delta/n}) \leq 2p \exp\left(-\frac{n^2x^2}{A_1n + A_2M_\delta^2 + A_3nxM_\delta}\right) + \delta$$

with the constants  $A_1, A_2, A_3$  defined above.

*Case II* Now we consider the case when  $0 < m \leq n - 2$ . Since  $\mathbf{Z}_t := \mathbf{Y}_t \mathbf{Y}_{t+m}^T$  is not symmetric for all  $t \in \mathbb{Z}$ , by applying matrix dilation (see [31], Sect. 2.1.16 for more details), we define the symmetric version of  $\mathbf{Z}_t^M$  as

$$\bar{\mathbf{Z}}_t^M := \begin{bmatrix} \mathbf{0} & \mathbf{Z}_t^M \\ (\mathbf{Z}_t^M)^T & \mathbf{0} \end{bmatrix}.$$

Observe that  $\lambda_{\max}(\bar{\mathbf{Z}}_t^M) = \|\bar{\mathbf{Z}}_t^M\| = \|\mathbf{Z}_t^M\|$ . By Lemma 4.4,  $\{\bar{\mathbf{Z}}_t^M\}_{t \in \mathbb{Z}}$  and  $\{\bar{\mathbf{Z}}_t^M - \mathbb{E}\bar{\mathbf{Z}}_t^M\}_{t \in \mathbb{Z}}$  are also sequences of  $\tau$ -mixing random matrices. Define

$$v_{\bar{\mathbf{Z}}^M}^2 := \sup_{K \subseteq \{1, \dots, n-m\}} \frac{1}{\text{card}(K)} \lambda_{\max}\left\{\mathbb{E}\left(\sum_{i \in K} \bar{\mathbf{Z}}_i^M - \mathbb{E}\bar{\mathbf{Z}}_i^M\right)^2\right\}.$$

Notice that  $v_{\bar{\mathbf{Z}}^M}^2$  and  $v_{\mathbf{Z}^M}^2$  have the same upper bound since spectral norm of block diagonal matrix is less than or equal to the spectral norm of each block.

Now we apply similar arguments in Case I to  $\{\bar{\mathbf{Z}}_t\}_{t \in \mathbb{Z}}$  and  $\{\bar{\mathbf{Z}}_t^M\}_{t \in \mathbb{Z}}$ .

$$\begin{aligned} & \mathbb{P} \left\{ \frac{1}{n-m} \left\| \sum_{i=1}^{n-m} (\mathbf{Z}_i - \mathbb{E} \mathbf{Z}_i) \right\| \geq x \right\} \\ & \leq \mathbb{P} \left[ \lambda_{\max} \left\{ \sum_{i=1}^{n-m} (\bar{\mathbf{Z}}_i^M - \mathbb{E} \bar{\mathbf{Z}}_i^M) \right\} \geq (n-m)x - \sum_{i=1}^{n-m} \|\mathbb{E} \bar{\mathbf{Z}}_i - \mathbb{E} \bar{\mathbf{Z}}_i^M\| \right] + \sum_{i=1}^{n-m} \mathbb{P}(\mathbf{Z}_i \neq \mathbf{Z}_i^M). \end{aligned}$$

The rest is straightforward by using Theorem 4.3 and Lemmas 4.2–4.5, and we thus finish the rest of the proof.

Lastly, we consider  $\kappa_1 \neq 1$ . Notice that for any sequence  $\{\mathbf{Y}_t\}_{t \in \mathbb{Z}}$  satisfying Assumptions (A1)–(A3), the sequence  $\{\mathbf{Y}_t/\kappa_1\}_{t \in \mathbb{Z}}$  will satisfy Assumptions (A1) automatically and Assumptions (A2), (A3) with  $\kappa_1 = 1$ . Hence, applying the above to  $\{\mathbf{Y}_t/\kappa_1\}_{t \in \mathbb{Z}}$  renders the results. This completes the proof of Proposition 4.1.  $\square$

## 4.2 Proof of Theorem 2.2

**Proof** The proof of Theorem 2.2 consists of two cases.

*Case I* When  $m = 0$ , we first state a more general result of Gaussian process. Proposition 4.6 considers a general Gaussian process without further assumptions on the covariance and autocovariance matrices. The proof modifies that of Theorem 5.1 in [32] with dependence among observations taken into account.  $\square$

**Proposition 4.6** (Proof in Sect. 4.2) *Let  $\{\mathbf{Y}_t\}_{t \in \mathbb{Z}}$  be a stationary sequence of mean-zero Gaussian random vectors with autocovariance matrices  $\boldsymbol{\Sigma}_m$  for  $0 \leq m \leq n-1$ . Then,*

$$\begin{aligned} \mathbb{E} \|\hat{\boldsymbol{\Sigma}}_0 - \boldsymbol{\Sigma}_0\| & \leq \frac{2}{n} \left\{ 2 \left( \|\boldsymbol{\Sigma}_0\|_* + 2 \sum_{m=1}^{n-1} \|\boldsymbol{\Sigma}_m\|_* \right) + \sqrt{2n \|\boldsymbol{\Sigma}_0\| \left( \|\boldsymbol{\Sigma}_0\|_* + 2 \sum_{m=1}^{n-1} \|\boldsymbol{\Sigma}_m\|_* \right)} \right. \\ & \quad \left. + \sqrt{2n \left( \|\boldsymbol{\Sigma}_0\| + 2 \sum_{m=1}^{n-1} \|\boldsymbol{\Sigma}_m\| \right) \text{Tr}(\boldsymbol{\Sigma}_0)} \right\}, \end{aligned}$$

where  $\|\cdot\|_*$  is the matrix nuclear norm.

The rest of the proof is to show the geometric decay of spectral norm and nuclear norm of autocovariance matrices under Assumptions (A2), (A3) in order to apply Proposition 4.6. It is obvious that  $\kappa_1^2 \asymp \|\boldsymbol{\Sigma}_0\|$  and  $\kappa_*^2 \asymp \text{Tr}(\boldsymbol{\Sigma}_0)$  when the process is a centered stationary Gaussian process. We first prove the geometric decay of spectral norm of autocovariance matrices. For any  $0 \leq m \leq n-1$  and any integer  $j$ , by Assumption (A3), there exists  $\tilde{\mathbf{Y}}_{1+m}$  that is identically distributed as  $\mathbf{Y}_{1+m}$ , independent of  $\mathbf{Y}_1$ , and

$$\sup_{\mathbf{u} \in \mathbb{S}^{p-1}} \|(\mathbf{Y}_{1+m} - \tilde{\mathbf{Y}}_{1+m})^T \mathbf{u}\|_{L(1+\epsilon)} \leq \gamma_3 \sqrt{\|\Sigma_0\|} \exp\{-\gamma_4(m-1)\}.$$

Therefore,

$$\begin{aligned} \|\Sigma_m\| &= \|\mathbb{E} \mathbf{Y}_1 \mathbf{Y}_{1+m}^T\| \\ &= \|\mathbb{E} \mathbf{Y}_1 (\mathbf{Y}_{1+m} - \tilde{\mathbf{Y}}_{1+m} + \tilde{\mathbf{Y}}_{1+m})^T\| \\ &= \|\mathbb{E} \mathbf{Y}_1 (\mathbf{Y}_{1+m} - \tilde{\mathbf{Y}}_{1+m})^T\| \\ &\leq \sup_{\mathbf{u}, \mathbf{v} \in \mathbb{S}^{p-1}} |\mathbb{E} \mathbf{u}^T \mathbf{Y}_1 (\mathbf{Y}_{1+m} - \tilde{\mathbf{Y}}_{1+m})^T \mathbf{v}| \\ &\leq C \|\Sigma_0\| \exp\{-\gamma_4(m-1)\}, \end{aligned}$$

where the last inequality is followed by Assumption (A3) and  $\gamma_3 = O(1)$  for some constant  $C > 0$  only depending on  $\epsilon, \gamma_3$ .

Similarly, by Assumption (A2), there exists  $\tilde{\mathbf{Y}}_{1+m}$  that is identically distributed as  $\mathbf{Y}_{1+m}$ , independent of  $\mathbf{Y}_1$ , and

$$\|\mathbf{Y}_{1+m} - \tilde{\mathbf{Y}}_{1+m}\|_{L(1+\epsilon)} \leq \gamma_1 \sqrt{\|\Sigma_0\|} \exp\{-\gamma_2(m-1)\}.$$

Then,

$$\begin{aligned} \|\Sigma_m\|_* &= \sqrt{\text{Tr}(\Sigma_m^T \Sigma_m)} \\ &= \sqrt{\text{Tr}\{\mathbb{E}(\mathbf{Y}_{1+m} - \tilde{\mathbf{Y}}_{1+m}) \mathbf{Y}_1^T \mathbb{E} \mathbf{Y}_1 (\mathbf{Y}_{1+m} - \tilde{\mathbf{Y}}_{1+m})^T\}} \\ &\leq \sqrt{\text{Tr}\{\mathbb{E}(\mathbf{Y}_{1+m} - \tilde{\mathbf{Y}}_{1+m}) \mathbf{Y}_1^T \mathbf{Y}_1 (\mathbf{Y}_{1+m} - \tilde{\mathbf{Y}}_{1+m})^T\}} \\ &= \sqrt{\text{Tr}\{\mathbb{E} \mathbf{Y}_1^T \mathbf{Y}_1 (\mathbf{Y}_{1+m} - \tilde{\mathbf{Y}}_{1+m}) (\mathbf{Y}_{1+m} - \tilde{\mathbf{Y}}_{1+m})^T\}} \\ &= \sqrt{\mathbb{E} \|\mathbf{Y}_1\|_2^2 \|\mathbf{Y}_{1+m} - \tilde{\mathbf{Y}}_{1+m}\|_2^2} \\ &\leq \|\mathbf{Y}_1\|_2 \| \mathbf{Y}_{1+m} - \tilde{\mathbf{Y}}_{1+m} \|_{L(\frac{1+\epsilon}{\epsilon})} \\ &\leq C \text{Tr}(\Sigma_0) \exp\{-\gamma_2(m-1)\}, \end{aligned}$$

where the third line is followed by the fact that  $\mathbb{E}(\mathbf{Y}_{1+m} - \tilde{\mathbf{Y}}_{1+m}) \mathbf{Y}_1^T \mathbb{E} \mathbf{Y}_1 (\mathbf{Y}_{1+m} - \tilde{\mathbf{Y}}_{1+m})^T \leq \mathbb{E} \mathbf{Y}_1^T \mathbf{Y}_1 (\mathbf{Y}_{1+m} - \tilde{\mathbf{Y}}_{1+m}) (\mathbf{Y}_{1+m} - \tilde{\mathbf{Y}}_{1+m})^T$  (“ $\leq$ ” is the Loewner partial order of Hermitian matrices), and both matrices are positive semi-definite, and the last line by Assumption (A2) and  $\gamma_1 = O(\sqrt{r(\Sigma_0)})$ . Indeed, for any  $\mathbf{u} \in \mathbb{R}^p$ ,  $\mathbb{E}\{\mathbf{u}^T (\mathbf{Y}_{1+m} - \tilde{\mathbf{Y}}_{1+m})\}^2 (\mathbf{Y}_1^T \mathbf{Y}_1) = \sum_{j=1}^p \mathbb{E}\{\mathbf{u}^T (\mathbf{Y}_{1+m} - \tilde{\mathbf{Y}}_{1+m})\}^2 \mathbf{Y}_{1,j}^2$  and  $\mathbb{E}\{\mathbf{u}^T (\mathbf{Y}_{1+m} - \tilde{\mathbf{Y}}_{1+m})\} \mathbf{Y}_1^T \mathbb{E} \mathbf{Y}_1 (\mathbf{Y}_{1+m} - \tilde{\mathbf{Y}}_{1+m})^T \mathbf{u} = \sum_{j=1}^p [\mathbb{E}\{\mathbf{u}^T (\mathbf{Y}_{1+m} - \tilde{\mathbf{Y}}_{1+m}) \mathbf{Y}_{1,j}\}]^2$ . The result follows.

*Case II* When  $m > 0$ , we denote  $\tilde{\mathbf{Y}}_i := (\mathbf{Y}_i^T \mathbf{Y}_{i+m}^T)^T$  for  $1 \leq i \leq n-m$ . It is obvious that  $\{\tilde{\mathbf{Y}}_i\}$  is a centered stationary Gaussian process satisfying Assumptions (A2), (A3). Denote  $\tilde{\Sigma}_0 := \mathbb{E} \tilde{\mathbf{Y}}_i \tilde{\mathbf{Y}}_i^T$  and notice that  $\Sigma_m$  is the off-diagonal block submatrix of  $\tilde{\Sigma}_0$ . By Case I and the fact that spectral norm of submatrix is bounded above by that of the full matrix, we obtain

$$\mathbb{E} \|\widehat{\Sigma}_m - \Sigma_m\| \leq C \|\bar{\Sigma}_0\| \left( \sqrt{\frac{r(\Sigma_0)}{n-m}} + \frac{r(\Sigma_0)}{n-m} \right).$$

Notice that  $\|\Sigma_0\| \leq \|\bar{\Sigma}_0\| \leq \|\Sigma_0\| + \|\Sigma_m\| \leq 2\|\Sigma_0\|$  since  $\Sigma_0 - \Sigma_m$  is positive semi-definite. This completes the proof.  $\square$

**Proof of Proposition 4.6** The proof heavily depends on the following observation. Denote  $\mathbf{Y} := (Y_1 \dots Y_n)$  and let  $\tilde{\mathbf{Y}}$  be an independent copy of  $\mathbf{Y}$ . Then,

$$\mathbb{E} \|\widehat{\Sigma}_0 - \Sigma_0\| \leq \frac{2}{n} \mathbb{E} \|\mathbf{Y} \tilde{\mathbf{Y}}^T\|.$$

This is same as Lemma 5.2 in [32] by noticing that the result holds without independence assumption.

Now we state the following two core lemmas used to complete the proof.  $\square$

**Lemma 4.7** (Proof in Sect. 4.3) *We have*

$$\mathbb{E} \|\widehat{\Sigma}_0 - \Sigma_0\| \leq \frac{2\sqrt{2}}{n} \left\{ \mathbb{E} \|\mathbf{Y}\| \cdot \sqrt{\text{Tr} \left( \Sigma_0 + 2 \sum_{d=1}^{n-1} \tilde{\Sigma}_d \right)} + \sqrt{2 \left( \|\Sigma_0\| + 2 \sum_{d=1}^{n-1} \|\Sigma_d\| \right)} \cdot \sqrt{n \text{Tr}(\Sigma_0)} \right\},$$

where  $\tilde{\Sigma}_d := (\mathbf{U}_d \Lambda_d \mathbf{U}_d^T + \mathbf{V}_d \Lambda_d \mathbf{V}_d^T)/2$ . Here,  $\mathbf{U}_d, \mathbf{V}_d$  and  $\Lambda_d$  are left singular vectors, right singular vectors, and singular values of  $\Sigma_d$  for all  $1 \leq d \leq n-1$ , respectively.

**Lemma 4.8** (Proof in Sect. 4.3) *We have*

$$\mathbb{E} \|\mathbf{Y}\| \leq \sqrt{2 \text{Tr} \left( \Sigma_0 + 2 \sum_{d=1}^{n-1} \tilde{\Sigma}_d \right)} + \sqrt{2n \|\Sigma_0\|},$$

where  $\tilde{\Sigma}_d$  for all  $1 \leq d \leq n-1$  are defined in Lemma 4.7.

The proof of Proposition 4.6 completes by combining Lemmas 4.7 and 4.8.  $\square$

### 4.3 Proofs of Auxiliary Lemmas

**Proof of Lemma 4.2** By Lemma A.2 in [8], we have  $\mathbb{E} \|Y_t\|_2^{2k} \leq (2k)^k \kappa_*^{2k}$  for  $t \in \mathbb{Z}$ . Hence,

$$\|Y_t\|_2^2 - \mathbb{E} \|Y_t\|_2^2 \leq 2\|Y_t\|_2^2 \leq 4\|Y_t\|_2^2 \leq 8\kappa_*^2.$$

Thus, by property of sub-exponential random variable and Chernoff inequality, we have for any  $x \geq 0$ ,

$$\mathbb{P}(\|Y_t\|_2^2 - \mathbb{E} \|Y_t\|_2^2 \geq x) \leq \exp \left\{ -C \min \left( \frac{x^2}{64\kappa_*^4}, \frac{x}{8\kappa_*^2} \right) \right\},$$



for some arbitrary constant  $C > 0$ . Obviously, we have for all  $x \geq 0$ ,

$$\mathbb{P}\{\|\mathbf{Y}_t\|_2^2 \geq 2\kappa_*^2 + 8\kappa_*^2(x + \sqrt{x})\} \leq \exp(-Cx)$$

for some arbitrary constant  $C > 0$ . This completes the proof.  $\square$

**Proof of Lemma 4.4** We first show that  $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$  is a sequence of  $\tau$ -mixing random vectors with geometric decay. Under Assumption (A2) (without loss of generality, take  $j = 0$ ), there exists a sequence of random vectors  $\{\tilde{\mathbf{Y}}_t\}_{t > 0}$  which is independent of  $\sigma(\{\mathbf{Y}_t\}_{t \leq 0})$ , identically distributed as  $\{\mathbf{Y}_t\}_{t > 0}$ , and for any integer  $t \geq 1$ ,

$$\|\mathbf{Y}_t - \tilde{\mathbf{Y}}_t\|_2 \|_{L(1+\epsilon)} \leq \gamma_1 \kappa_1 \exp\{-\gamma_2(t-1)\}$$

for some constant  $\epsilon > 0$ . Then, for any  $m \geq 0$ ,

$$\begin{aligned} & \mathbb{E}\|\mathbf{Y}_t \mathbf{Y}_{t+m}^T - \tilde{\mathbf{Y}}_t \tilde{\mathbf{Y}}_{t+m}^T\| \\ &= \mathbb{E}\|\mathbf{Y}_t \mathbf{Y}_{t+m}^T - \mathbf{Y}_t \tilde{\mathbf{Y}}_{t+m}^T + \mathbf{Y}_t \tilde{\mathbf{Y}}_{t+m}^T - \tilde{\mathbf{Y}}_t \tilde{\mathbf{Y}}_{t+m}^T\| \\ &\leq \mathbb{E}\|\mathbf{Y}_t (\mathbf{Y}_{t+m} - \tilde{\mathbf{Y}}_{t+m})^T\| + \mathbb{E}\|(\mathbf{Y}_t - \tilde{\mathbf{Y}}_t) \tilde{\mathbf{Y}}_{t+m}^T\| \\ &\leq \|\mathbf{Y}_t\|_2 \|_{L(\frac{1+\epsilon}{\epsilon})} \|\mathbf{Y}_{t+m} - \tilde{\mathbf{Y}}_{t+m}\|_2 \|_{L(1+\epsilon)} + \|\mathbf{Y}_{t+m}\|_2 \|_{L(\frac{1+\epsilon}{\epsilon})} \|\mathbf{Y}_t - \tilde{\mathbf{Y}}_t\|_2 \|_{L(1+\epsilon)} \\ &\leq C \gamma_1 \kappa_1 \kappa_* \exp\{-\gamma_2(t-1)\}, \end{aligned}$$

where the fourth line is followed by Hölder's inequality and the fact that

$$\sup_{t \in \mathbb{Z}} \|\mathbf{Y}_t\|_2 \|_{L(\alpha)} \leq \sup_{t \in \mathbb{Z}} \sup_{\mathbf{u} \in \tilde{\mathbb{S}}^{p-1}} \|\mathbf{u}^T \mathbf{Y}_t\|_{L(\alpha)} \leq \sup_{t \in \mathbb{Z}} \sup_{\mathbf{u} \in \tilde{\mathbb{S}}^{p-1}} \sqrt{\alpha} \|\mathbf{u}^T \mathbf{Y}_t\|_{\psi_2} \leq \sqrt{\alpha} \kappa_*$$

for any  $\alpha \geq 1$ . Here,  $C > 0$  is some constant only depending on  $\epsilon$ .

Now define  $\tilde{\mathbf{X}}_t := \tilde{\mathbf{Y}}_t \tilde{\mathbf{Y}}_t^T$  for any integer  $t > 0$ . It is obvious that  $\{\tilde{\mathbf{X}}_t\}_{t > 0}$  is independent of  $\{\mathbf{X}_t\}_{t \leq 0}$  and identically distributed as  $\{\mathbf{X}_t\}_{t > 0}$ . By applying Lemma A.1, for any indices  $0 < k \leq t_1 < \dots < t_\ell$ , we obtain

$$\tau\{\sigma(\{\mathbf{X}_t\}_{t \leq 0}), (\mathbf{X}_{t_1}, \dots, \mathbf{X}_{t_\ell}); \|\cdot\|\} \leq \sum_{i=1}^{\ell} \mathbb{E}\|\mathbf{X}_{t_i} - \tilde{\mathbf{X}}_{t_i}\| \leq C \gamma_1 \kappa_1 \kappa_* \ell \exp\{-\gamma_2(k-1)\}.$$

By definition of  $\tau$ -mixing coefficient, this yields

$$\tau(k; \{\mathbf{X}_t\}_{t \in \mathbb{Z}}, \|\cdot\|) \leq C \gamma_1 \kappa_1 \kappa_* \exp\{-\gamma_2(k-1)\}$$

for some constant  $C > 0$  only depending on  $\epsilon$ .

Now we proceed to prove  $\tau$ -mixing properties for the “truncated version.” The following lemma is needed.  $\square$

**Lemma 4.9** (Proof in Sect. 4.3) *Let  $\mathbf{u}_1, \mathbf{u}_2, \mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^p$  for  $p \geq 1$  with unit length under  $\ell_2$ -norm and  $\sigma_u \geq 0$ . Then, the function*

$$f(\sigma_v) = \|\sigma_v \mathbf{v}_1 \mathbf{v}_2^T - \sigma_u \mathbf{u}_1 \mathbf{u}_2^T\|$$

*is nondecreasing in the range  $\sigma_v \in [\sigma_u, \infty]$ . In particular, for any  $M \geq 0$  such that  $M \leq \sigma_u$ ,  $M \leq \sigma_v$ , we have*

$$\|M \mathbf{v}_1 \mathbf{v}_2^T - M \mathbf{u}_1 \mathbf{u}_2^T\| \leq \|\sigma_v \mathbf{v}_1 \mathbf{v}_2^T - \sigma_u \mathbf{u}_1 \mathbf{u}_2^T\|.$$

Now consider three cases.

- (1) When  $\|\mathbf{X}_t\| \leq M$  and  $\|\tilde{\mathbf{X}}_t\| \leq M$ ,  $\|\mathbf{X}_t^M - \tilde{\mathbf{X}}_t^M\| = \|\mathbf{X}_t - \tilde{\mathbf{X}}_t\|$ .
- (2) When  $\|\mathbf{X}_t\| \leq M$  and  $\|\tilde{\mathbf{X}}_t\| > M$ , we have

$$\mathbf{X}_t^M = \mathbf{X}_t = \|\mathbf{Y}_t\|_2^2 \frac{\mathbf{Y}_t}{\|\mathbf{Y}_t\|_2} \frac{\mathbf{Y}_t^T}{\|\mathbf{Y}_t\|_2} \quad \text{and} \quad \tilde{\mathbf{X}}_t^M = M \frac{\tilde{\mathbf{Y}}_t}{\|\tilde{\mathbf{Y}}_t\|_2} \frac{\tilde{\mathbf{Y}}_t^T}{\|\tilde{\mathbf{Y}}_t\|_2}.$$

Since  $\frac{\mathbf{Y}_t}{\|\mathbf{Y}_t\|_2}, \frac{\tilde{\mathbf{Y}}_t}{\|\tilde{\mathbf{Y}}_t\|_2}$  have unit length and  $\|\mathbf{Y}_t\|_2^2 \leq M < \|\tilde{\mathbf{Y}}_t\|_2^2$ , we have  $\|\mathbf{X}_t^M - \tilde{\mathbf{X}}_t^M\| \leq \|\mathbf{X}_t - \tilde{\mathbf{X}}_t\|$  by Lemma 4.9. By symmetry, the same argument also applies to the case where  $\|\mathbf{X}_t\| > M$  and  $\|\tilde{\mathbf{X}}_t\| \leq M$ .

- (3) When  $\|\mathbf{X}_t\| > M$  and  $\|\tilde{\mathbf{X}}_t\| > M$ , we have  $\mathbf{X}_t^M = M \frac{\mathbf{Y}_t}{\|\mathbf{Y}_t\|_2} \frac{\mathbf{Y}_t^T}{\|\mathbf{Y}_t\|_2}$  and  $\tilde{\mathbf{X}}_t^M = M \frac{\tilde{\mathbf{Y}}_t}{\|\tilde{\mathbf{Y}}_t\|_2} \frac{\tilde{\mathbf{Y}}_t^T}{\|\tilde{\mathbf{Y}}_t\|_2}$ . Again by Lemma 4.9, we have  $\|\mathbf{X}_t^M - \tilde{\mathbf{X}}_t^M\| \leq \|\mathbf{X}_t - \tilde{\mathbf{X}}_t\|$ .

By combining three cases,  $\|\mathbf{X}_t^M - \tilde{\mathbf{X}}_t^M\| \leq \|\mathbf{X}_t - \tilde{\mathbf{X}}_t\|$  always holds, and hence,  $\mathbb{E}\|\mathbf{X}_t^M - \tilde{\mathbf{X}}_t^M\| \leq \mathbb{E}\|\mathbf{X}_t - \tilde{\mathbf{X}}_t\|$  for any  $t \geq 1$ . Hence, for any indices  $0 < k \leq t_1 < \dots < t_\ell$ , by Lemma A.1, we have

$$\tau\{\sigma(\{\mathbf{X}_t^M\}_{t \leq 0}), (\mathbf{X}_{t_1}^M, \dots, \mathbf{X}_{t_\ell}^M); \|\cdot\|\} \leq C \gamma_1 \kappa_1 \kappa_* \ell \exp\{-\gamma_2(k-1)\}$$

for some constant  $C > 0$  only depending on  $\epsilon$ . By definition of  $\tau$ -mixing coefficient, this yields

$$\tau(k; \{\mathbf{X}_t^M\}_{t \in \mathbb{Z}}, \|\cdot\|) \leq C \gamma_1 \kappa_1 \kappa_* \exp\{-\gamma_2(k-1)\}$$

for some constant  $C > 0$  only depending on  $\epsilon$ . Notice that  $\mathbb{E}\|\mathbf{X}_t^M - \mathbb{E}\mathbf{X}_t^M - (\tilde{\mathbf{X}}_t^M - \mathbb{E}\tilde{\mathbf{X}}_t^M)\| = \mathbb{E}\|\mathbf{X}_t^M - \tilde{\mathbf{X}}_t^M\|$  since  $\mathbb{E}\tilde{\mathbf{X}}_t^M = \mathbb{E}\mathbf{X}_t^M$  for any  $t \geq 1$ . The  $\tau$ -mixing property stated above applies to  $\{\mathbf{X}_t^M - \mathbb{E}\mathbf{X}_t^M\}$  directly.

Similar arguments apply to  $\{\mathbf{Z}_t^M\}_{t \in \mathbb{Z}}$  and  $\{\mathbf{Z}_t^M - \mathbb{E}\mathbf{Z}_t^M\}_{t \in \mathbb{Z}}$ , so we omit the details. This completes the proof.  $\square$

**Proof of Lemma 4.5** The proof consists of two steps.

**Step I** We first provide an upper bound for  $v_{\mathbf{X}}^2$ . Without loss of generality, we only consider  $\|\mathbb{E}(\mathbf{X}_0 - \mathbb{E}\mathbf{X}_0)(\mathbf{X}_k - \mathbb{E}\mathbf{X}_k)\|$  for  $k \geq 0$ . Under Assumptions (A2)–(A3), there exists  $\tilde{\mathbf{Y}}_k$  where  $\tilde{\mathbf{Y}}_k$  is independent of  $\sigma(\{\mathbf{Y}_t\}_{t \leq 0})$ , identically distributed as  $\mathbf{Y}_k$ , and

$$\begin{aligned}\|\mathbf{Y}_k - \tilde{\mathbf{Y}}_k\|_2 &\leq \gamma_1 \kappa_1 \exp\{-\gamma_2(k-1)\}, \\ \|(\mathbf{Y}_k - \tilde{\mathbf{Y}}_k)^T \mathbf{u}\|_{L(1+\epsilon)} &\leq \gamma_3 \kappa_1 \exp\{-\gamma_4(k-1)\}\end{aligned}$$

for constants  $\gamma_1, \gamma_2, \gamma_3, \gamma_4 > 0$  in Assumptions (A2)–(A3).

For  $k = 0$ , we have

$$\|\mathbb{E}\mathbf{X}_0\mathbf{X}_0 - \mathbb{E}\mathbf{X}_0\mathbb{E}\mathbf{X}_0\| \leq C(\kappa_1^4 + \kappa_1^2 \kappa_*^2)$$

by Assumption (A1) for some universal constant  $C > 0$ . For  $k > 0$ , we obtain

$$\begin{aligned}\|\mathbb{E}\mathbf{X}_0\mathbf{X}_k - \mathbb{E}\mathbf{X}_0\mathbb{E}\mathbf{X}_k\| &= \|\mathbb{E}\mathbf{X}_0\mathbf{X}_k - \mathbb{E}\mathbf{X}_0\tilde{\mathbf{X}}_k\| \\ &= \|\mathbb{E}\mathbf{X}_0(\mathbf{X}_k - \tilde{\mathbf{X}}_k)\| \\ &= \sup_{\mathbf{u}, \mathbf{v} \in \mathbb{S}^{p-1}} |\mathbb{E}[\mathbf{u}^T \mathbf{Y}_0 \mathbf{Y}_0^T (\mathbf{Y}_k \mathbf{Y}_k^T - \tilde{\mathbf{Y}}_k \tilde{\mathbf{Y}}_k^T) \mathbf{v}]| \\ &\leq \sup_{\mathbf{u}, \mathbf{v} \in \mathbb{S}^{p-1}} |\mathbb{E}[\mathbf{u}^T \mathbf{Y}_0 \mathbf{Y}_0^T \mathbf{Y}_k (\mathbf{Y}_k^T - \tilde{\mathbf{Y}}_k^T) \mathbf{v} + \mathbf{u}^T \mathbf{Y}_0 \mathbf{Y}_0^T (\mathbf{Y}_k - \tilde{\mathbf{Y}}_k) \tilde{\mathbf{Y}}_k^T \mathbf{v}]| \\ &\leq \sup_{\mathbf{u}, \mathbf{v} \in \mathbb{S}^{p-1}} \left\{ \mathbb{E}|\mathbf{Y}_0^T \mathbf{Y}_k|^{\frac{3(1+\epsilon)}{2\epsilon}} \right\}^{\frac{2\epsilon}{3(1+\epsilon)}} \|\mathbf{u}^T \mathbf{Y}_0\|_{L(\frac{3(1+\epsilon)}{\epsilon})} \|(\mathbf{Y}_k - \tilde{\mathbf{Y}}_k)^T \mathbf{v}\|_{L(1+\epsilon)} \\ &\quad + \{\mathbb{E}|\mathbf{u}^T \mathbf{Y}_0 \tilde{\mathbf{Y}}_k^T \mathbf{v}|^{\frac{3(1+\epsilon)}{2\epsilon}}\}^{\frac{2\epsilon}{3(1+\epsilon)}} \|\mathbf{Y}_0\|_2 \| \tilde{\mathbf{Y}}_k \|_{L(\frac{3(1+\epsilon)}{\epsilon})} \|\mathbf{Y}_k - \tilde{\mathbf{Y}}_k\|_2 \| \mathbf{v} \|_{L(1+\epsilon)} \\ &\leq C \kappa_1^2 \kappa_* (\kappa_* \gamma_3 + \kappa_1 \gamma_1) \exp\{-\min(\gamma_2, \gamma_4)(k-1)\},\end{aligned}$$

where the first line is followed by  $\mathbb{E}\mathbf{X}_k = \mathbb{E}\tilde{\mathbf{X}}_k$ , fifth line by Hölder's inequality, and sixth line by Assumptions (A1)–(A3) for some constant  $C > 0$  only depending on  $\epsilon$ .

Hence, for any  $K \subseteq \{1, \dots, n\}$ ,

$$\begin{aligned}&\frac{1}{\text{card}(K)} \lambda_{\max} \left\{ \mathbb{E} \left( \sum_{i \in K} \mathbf{X}_i - \mathbb{E}\mathbf{X}_i \right)^2 \right\} \\ &\leq \frac{1}{\text{card}(K)} \left\| \sum_{i, j \in K} \mathbb{E}(\mathbf{X}_i - \mathbb{E}\mathbf{X}_i)(\mathbf{X}_j - \mathbb{E}\mathbf{X}_j) \right\| \\ &\leq \frac{1}{\text{card}(K)} \sum_{i, j \in K} \|\mathbb{E}(\mathbf{X}_i - \mathbb{E}\mathbf{X}_i)(\mathbf{X}_j - \mathbb{E}\mathbf{X}_j)\| \\ &\leq C \left[ \kappa_1^4 + \kappa_1^2 \kappa_*^2 + \frac{\kappa_1^2 \kappa_* (\kappa_* \gamma_3 + \kappa_1 \gamma_1)}{\text{card}(K)} \sum_{i, j \in K, i \neq j} \exp\{-\min(\gamma_2, \gamma_4)(|i-j|-1)\} \right] \\ &\leq C \left[ \frac{\kappa_1^2 \{\kappa_1^2 + \kappa_1 \kappa_* \gamma_1 + \kappa_*^2 (\gamma_3 + 1)\}}{1 - \exp(-\min\{\gamma_2, \gamma_4\})} \right].\end{aligned}$$

**Step II** We first bound  $v_{\mathbf{X}^M}^2$ . By definition, we have

$$\begin{aligned} \left\| \mathbb{E} \left( \sum_{i \in K} \mathbf{X}_i^M - \mathbb{E} \mathbf{X}_i^M \right)^2 \right\| &= \left\| \sum_{i, j \in K} \mathbb{E} (\mathbf{X}_i^M - \mathbb{E} \mathbf{X}_i^M) (\mathbf{X}_j^M - \mathbb{E} \mathbf{X}_j^M) \right\| \\ &= \left\| \sum_{i, j \in K} (\mathbb{E} \mathbf{X}_i^M \mathbf{X}_j^M - \mathbb{E} \mathbf{X}_i^M \mathbb{E} \mathbf{X}_j^M) \right\|. \end{aligned}$$

Without loss of generality, we consider  $\|\mathbb{E} \mathbf{X}_0^M \mathbf{X}_k^M - \mathbb{E} \mathbf{X}_0^M \mathbb{E} \mathbf{X}_k^M\|$  for  $k \geq 0$ . Let  $\tilde{\mathbf{X}}_k^M$  be defined as in the proof of Lemma 4.4. Then,  $\tilde{\mathbf{X}}_k^M$  is independent of  $\tilde{\mathbf{X}}_0^M$  and distributed as  $\mathbf{X}_k^M$ . Hence,

$$\|\mathbb{E} \mathbf{X}_0^M \mathbf{X}_k^M - \mathbb{E} \mathbf{X}_0^M \mathbb{E} \mathbf{X}_k^M\| = \|\mathbb{E} \mathbf{X}_0^M \mathbf{X}_k^M - \mathbb{E} \mathbf{X}_0^M \mathbb{E} \tilde{\mathbf{X}}_k^M\|.$$

Then, we could rewrite

$$\begin{aligned} \|\mathbb{E} \mathbf{X}_0^M \mathbf{X}_k^M - \mathbb{E} \mathbf{X}_0^M \mathbb{E} \tilde{\mathbf{X}}_k^M\| &= \|\mathbb{E} \mathbf{X}_0 \mathbf{X}_k \zeta_0 \zeta_k - \mathbb{E} \mathbf{X}_0 \tilde{\mathbf{X}}_k \zeta_0 \tilde{\zeta}_k\| \\ &= \|\mathbb{E} \mathbf{X}_0 (\mathbf{X}_k - \tilde{\mathbf{X}}_k) \zeta_0 \zeta_k + \mathbb{E} \mathbf{X}_0 \tilde{\mathbf{X}}_k \zeta_0 (\zeta_k - \tilde{\zeta}_k)\|, \end{aligned}$$

where  $\zeta_i = \frac{M \wedge \|\mathbf{X}_i\|}{\|\mathbf{X}_i\|}$ ,  $\tilde{\zeta}_i = \frac{M \wedge \|\tilde{\mathbf{X}}_i\|}{\|\tilde{\mathbf{X}}_i\|}$ . Since  $\zeta_0, \zeta_k$  are bounded by 1, we have

$$\begin{aligned} \|\mathbb{E} \mathbf{X}_0 (\mathbf{X}_k - \tilde{\mathbf{X}}_k) \zeta_0 \zeta_k\| &= \sup_{\mathbf{u}, \mathbf{v} \in \mathbb{S}^{p-1}} \mathbb{E} |\mathbf{u}^T \mathbf{X}_0 (\mathbf{X}_k - \tilde{\mathbf{X}}_k) \zeta_0 \zeta_k \mathbf{v}| \\ &\leq \sup_{\mathbf{u}, \mathbf{v} \in \mathbb{S}^{p-1}} \mathbb{E} |\mathbf{u}^T \mathbf{X}_0 (\mathbf{X}_k - \tilde{\mathbf{X}}_k) \mathbf{v}| \\ &= \|\mathbb{E} \mathbf{X}_0 (\mathbf{X}_k - \tilde{\mathbf{X}}_k)\| \\ &\leq C \kappa_1^2 (\kappa_1 \kappa_* \gamma_1 + \kappa_*^2 \gamma_3) \exp\{-\min(\gamma_2, \gamma_4)(k-1)\}, \end{aligned}$$

where the last inequality is from result in Step I for some constant  $C > 0$  only depending on  $\epsilon$ .

On the other hand, by applying Hölder's inequality, we have

$$\begin{aligned} \|\mathbb{E} \mathbf{X}_0 \tilde{\mathbf{X}}_k \zeta_0 (\zeta_k - \tilde{\zeta}_k)\| &= \sup_{\mathbf{u}, \mathbf{v} \in \mathbb{S}^{p-1}} \mathbb{E} |\mathbf{u}^T \mathbf{X}_0 \tilde{\mathbf{X}}_k \mathbf{v}| |\zeta_k - \tilde{\zeta}_k| \\ &\leq \sup_{\mathbf{u}, \mathbf{v} \in \mathbb{S}^{p-1}} \{\mathbb{E} |\mathbf{u}^T \mathbf{Y}_0 \mathbf{Y}_0^T \tilde{\mathbf{Y}}_k \tilde{\mathbf{Y}}_k^T \mathbf{v}|^{\frac{5(1+\epsilon)}{4\epsilon}}\}^{\frac{4\epsilon}{5(1+\epsilon)}} \{\mathbb{E} |\zeta_k - \tilde{\zeta}_k|^{\frac{5(1+\epsilon)}{5+\epsilon}}\}^{\frac{5+\epsilon}{5(1+\epsilon)}}. \end{aligned}$$

Hence, for any  $\mathbf{u}, \mathbf{v} \in \mathbb{S}^{p-1}$ ,

$$\begin{aligned} &\{\mathbb{E} |\mathbf{u}^T \mathbf{Y}_0 \mathbf{Y}_0^T \tilde{\mathbf{Y}}_k \tilde{\mathbf{Y}}_k^T \mathbf{v}|^{\frac{5(1+\epsilon)}{4\epsilon}}\}^{\frac{4\epsilon}{5(1+\epsilon)}} \\ &\leq \|\mathbf{u}^T \mathbf{Y}_0\|_{L(\frac{5(1+\epsilon)}{\epsilon})} \|\mathbf{u}^T \tilde{\mathbf{Y}}_k\|_{L(\frac{5(1+\epsilon)}{\epsilon})} \|\mathbf{Y}_0\|_2 \| \tilde{\mathbf{Y}}_k \|_2 \| \tilde{\mathbf{Y}}_k \|_2 \| \tilde{\mathbf{Y}}_k \|_2 \\ &\leq C \kappa_1^2 \kappa_*^2, \end{aligned}$$

where the first line is followed by Hölder's inequality and the last line by Assumption (A1) for some constant  $C > 0$  only depending on  $\epsilon$ .

Next, we need to bound  $\|\zeta_k - \tilde{\zeta}_k\|_{L(\frac{5(1+\epsilon)}{5+\epsilon})}$ . For the sake of presentation clearness, we denote  $a_k := \|\mathbf{X}_k\|$  and  $\tilde{a}_k := \|\tilde{\mathbf{X}}_k\|$  and rewrite

$$\begin{aligned} & \|\zeta_k - \tilde{\zeta}_k\|_{L(\frac{5(1+\epsilon)}{5+\epsilon})} \\ &= \left\| M \left| \frac{1}{a_k} - \frac{1}{\tilde{a}_k} \right| \mathbf{1}_{\{a_k > M, \tilde{a}_k > M\}} + \left(1 - \frac{M}{a_k}\right) \mathbf{1}_{\{a_k > M, \tilde{a}_k \leq M\}} + \left(1 - \frac{M}{\tilde{a}_k}\right) \mathbf{1}_{\{a_k \leq M, \tilde{a}_k > M\}} \right\|_{L(\frac{5(1+\epsilon)}{5+\epsilon})} \\ &\leq \left\| M \left| \frac{1}{a_k} - \frac{1}{\tilde{a}_k} \right| \mathbf{1}_{\{a_k > M, \tilde{a}_k > M\}} \right\|_{L(\frac{5(1+\epsilon)}{5+\epsilon})} + \left\| \left(1 - \frac{M}{a_k}\right) \mathbf{1}_{\{a_k > M, \tilde{a}_k \leq M\}} \right\|_{L(\frac{5(1+\epsilon)}{5+\epsilon})} \\ &\quad + \left\| \left(1 - \frac{M}{\tilde{a}_k}\right) \mathbf{1}_{\{a_k \leq M, \tilde{a}_k > M\}} \right\|_{L(\frac{5(1+\epsilon)}{5+\epsilon})}, \end{aligned} \quad (4.5)$$

where the last inequality is followed by the fact that  $\|\cdot\|_{L(\frac{5(1+\epsilon)}{5+\epsilon})}$  is a norm for  $\epsilon > 0$ .

For the first term, we have

$$\begin{aligned} & \left\| M \left| \frac{1}{a_k} - \frac{1}{\tilde{a}_k} \right| \mathbf{1}_{\{a_k > M, \tilde{a}_k > M\}} \right\|_{L(\frac{5(1+\epsilon)}{5+\epsilon})} = \left\| M \left| \frac{\tilde{a}_k - a_k}{a_k \tilde{a}_k} \right| \mathbf{1}_{\{a_k > M, \tilde{a}_k > M\}} \right\|_{L(\frac{5(1+\epsilon)}{5+\epsilon})} \\ &\leq \frac{1}{M} \{\mathbb{E}|\tilde{a}_k - a_k|^{\frac{5(1+\epsilon)}{5+\epsilon}}\}^{\frac{5+\epsilon}{5(1+\epsilon)}} \\ &\leq \frac{1}{M} \{\mathbb{E}\|\mathbf{X}_k - \tilde{\mathbf{X}}_k\|^{\frac{5(1+\epsilon)}{5+\epsilon}}\}^{\frac{5+\epsilon}{5(1+\epsilon)}} \\ &\leq C\gamma_1\kappa_1\kappa_* \exp\{-\gamma_2(k-1)\}/M, \end{aligned}$$

where the last inequality is followed by Lemma 4.4 for some constant  $C > 0$  only depending on  $\epsilon$ . With the chosen  $M \geq C\gamma_1\kappa_1\kappa_*$ , we have

$$\left\| M \left| \frac{1}{a_k} - \frac{1}{\tilde{a}_k} \right| \mathbf{1}_{\{a_k > M, \tilde{a}_k > M\}} \right\|_{L(\frac{5(1+\epsilon)}{5+\epsilon})} \leq \exp\{-\gamma_2(k-1)\}.$$

For the second term, taking any  $\epsilon_k > 0$ , we have

$$\begin{aligned} & \left\| \left(1 - \frac{M}{a_k}\right) \mathbf{1}_{\{a_k > M, \tilde{a}_k \leq M\}} \right\|_{L(\frac{5(1+\epsilon)}{5+\epsilon})} \\ &= \left\| \left(1 - \frac{M}{M + \epsilon_k}\right) \mathbf{1}_{\{M < a_k \leq M + \epsilon_k, \tilde{a}_k \leq M\}} \right\|_{L(\frac{5(1+\epsilon)}{5+\epsilon})} + \left\| \left(1 - \frac{M}{a_k}\right) \mathbf{1}_{\{a_k > M + \epsilon_k, \tilde{a}_k \leq M\}} \right\|_{L(\frac{5(1+\epsilon)}{5+\epsilon})} \\ &\leq \frac{\epsilon_k}{M} + \left\| \mathbf{1}_{\{a_k > M + \epsilon_k, \tilde{a}_k \leq M\}} \right\|_{L(\frac{5(1+\epsilon)}{5+\epsilon})} \leq \frac{\epsilon_k}{M} + \{\mathbb{P}(|a_k - \tilde{a}_k| > \epsilon_k)\}^{\frac{5+\epsilon}{5(1+\epsilon)}}. \end{aligned}$$

By Markov inequality and Lemma 4.4, we have

$$\mathbb{P}(|a_k - \tilde{a}_k| > \epsilon_k) \leq \frac{\mathbb{E}\|\mathbf{X}_k - \tilde{\mathbf{X}}_k\|}{\epsilon_k} \leq \frac{C\gamma_1\kappa_1\kappa_* \exp\{-\gamma_2(k-1)\}}{\epsilon_k}$$

for some constant  $C > 0$  only depending on  $\epsilon$ . Taking  $\epsilon_k = C\gamma_1\kappa_1\kappa_* \exp\{-\frac{5+\epsilon}{6\epsilon+10}\gamma_2(k-1)\}$ , we obtain

$$\left\| \left(1 - \frac{M}{a_k}\right) \mathbf{1}_{\{a_k > M, \tilde{a}_k \leq M\}} \right\|_{L(\frac{5(1+\epsilon)}{5+\epsilon})} \leq 2 \exp \left\{ -\frac{5+\epsilon}{6\epsilon+10}\gamma_2(k-1) \right\}.$$

The third term follows by symmetry. Putting together, we have for  $k > 0$ ,

$$\begin{aligned} \|\zeta_k - \tilde{\zeta}_k\|_{L(\frac{5(1+\epsilon)}{5+\epsilon})} &\leq C \exp \left\{ -\frac{5+\epsilon}{6\epsilon+10}\gamma_2(k-1) \right\}, \\ \|\mathbb{E}\mathbf{X}_0\tilde{\mathbf{X}}_k\zeta_0(\zeta_k - \tilde{\zeta}_k)\| &\leq C\kappa_1^2\kappa_*^2 \exp \left\{ -\frac{5+\epsilon}{6\epsilon+10}\gamma_2(k-1) \right\}, \\ \|\mathbb{E}\mathbf{X}_0^M\mathbf{X}_k^M - \mathbb{E}\mathbf{X}_0^M\mathbb{E}\tilde{\mathbf{X}}_k^M\| &\leq C\kappa_1^2\{\kappa_1\kappa_*\gamma_1 + \kappa_*^2(\gamma_3 + 1)\} \exp \left\{ -\min\left(\frac{5+\epsilon}{6\epsilon+10}\gamma_2, \gamma_4\right)(k-1) \right\} \end{aligned}$$

for some constant  $C > 0$  only depending on  $\epsilon$ . Hence, for any  $K \subseteq \{1, \dots, n\}$ ,

$$\begin{aligned} &\frac{1}{\text{card}(K)} \lambda_{\max} \left\{ \mathbb{E} \left( \sum_{i \in K} \mathbf{X}_i^M - \mathbb{E}\mathbf{X}_i^M \right)^2 \right\} \\ &\leq \frac{1}{\text{card}(K)} \left\| \sum_{i,j \in K} \mathbb{E}(\mathbf{X}_i^M - \mathbb{E}\mathbf{X}_i^M)(\mathbf{X}_j^M - \mathbb{E}\mathbf{X}_j^M) \right\| \\ &\leq \frac{1}{\text{card}(K)} \sum_{i,j \in K} \|\mathbb{E}(\mathbf{X}_i^M - \mathbb{E}\mathbf{X}_i^M)(\mathbf{X}_j^M - \mathbb{E}\mathbf{X}_j^M)\| \\ &\leq C \left[ \kappa_1^4 + \kappa_*^2\kappa_1^2 + \frac{\kappa_1^2\{\kappa_1\kappa_*\gamma_1 + \kappa_*^2(\gamma_3 + 1)\}}{\text{card}(K)} \right. \\ &\quad \left. \sum_{i,j \in K, i \neq j} \exp \left\{ -\min\left(\frac{5+\epsilon}{6\epsilon+10}\gamma_2, \gamma_4\right)(|i-j|-1) \right\} \right] \\ &\leq C \frac{\kappa_1^2\{\kappa_1^2 + \kappa_1\kappa_*\gamma_1 + \kappa_*^2(\gamma_3 + 2)\}}{1 - \exp\{-\min(\frac{5+\epsilon}{6\epsilon+10}\gamma_2, \gamma_4)\}} \end{aligned}$$

for some constant  $C > 0$  only depending on  $\epsilon$ .

Similar arguments apply to  $v_{Z,M}^2$ , so we omit the details. This completes the proof.  $\square$

**Proof of Lemma 4.9** Fix  $\mathbf{u}_1, \mathbf{u}_2, \mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^p$  with unit length and  $\sigma_u \geq 0$ . For any  $\sigma_v \geq \sigma_u$ , we perform singular value decomposition for matrix  $\mathbf{X}(\sigma_v) := \sigma_u \mathbf{u}_1 \mathbf{u}_2^T - \sigma_v \mathbf{v}_1 \mathbf{v}_2^T$ .

According to Eq. (8) in [6], the nonzero singular values of  $\mathbf{X}(\sigma_v)$  are identical to those of

$$\mathbf{S}(\sigma_v) = \begin{bmatrix} \sigma_u - \sigma_v \mathbf{u}_1^T \mathbf{v}_1 \mathbf{v}_2^T \mathbf{u}_2 & -\sigma_v \mathbf{u}_1^T \mathbf{v}_1 \|\mathbf{v}_2 - \mathbf{u}_2 \mathbf{u}_2^T \mathbf{v}_2\|_2 \\ \sigma_v \mathbf{u}_2^T \mathbf{v}_2 \|\mathbf{v}_1 - \mathbf{u}_1 \mathbf{u}_1^T \mathbf{v}_1\|_2 & \sigma_v^2 \|\mathbf{v}_1 - \mathbf{u}_1 \mathbf{u}_1^T \mathbf{v}_1\|_2 \|\mathbf{v}_2 - \mathbf{u}_2 \mathbf{u}_2^T \mathbf{v}_2\|_2 \end{bmatrix}.$$

For simplicity, denote  $w = \mathbf{u}_1^T \mathbf{v}_1 \mathbf{v}_2^T \mathbf{u}_2$ ,  $\tilde{\mathbf{v}}_1 = \mathbf{v}_1 - \mathbf{u}_1 \mathbf{u}_1^T \mathbf{v}_1$ ,  $\tilde{\mathbf{u}}_1 = \mathbf{v}_2 - \mathbf{u}_2 \mathbf{u}_2^T \mathbf{v}_2$ . Hence,  $\mathbf{S}(\sigma_v)$  could be rewritten as

$$\mathbf{S}(\sigma_v) = \begin{bmatrix} \sigma_u - \sigma_v w & -\sigma_v \mathbf{u}_1^T \mathbf{v}_1 \|\tilde{\mathbf{v}}_2\|_2 \\ \sigma_v \mathbf{u}_2^T \mathbf{v}_2 \|\tilde{\mathbf{v}}_1\|_2 & \sigma_v^2 \|\tilde{\mathbf{v}}_1\|_2 \|\tilde{\mathbf{v}}_2\|_2 \end{bmatrix}.$$

Using the calculation on Page 86 in [5],  $\|\mathbf{S}(\sigma_v)\| = Q(\sigma_v) + R(\sigma_v)$ , where

$$Q(\sigma_v) := \sqrt{(\sigma_u - \sigma_v w + \sigma_v \|\tilde{\mathbf{v}}_1\|_2 \|\tilde{\mathbf{v}}_2\|_2)^2 + \sigma_v^2 (\mathbf{u}_1^T \mathbf{v}_1 \|\tilde{\mathbf{v}}_2\|_2 + \mathbf{u}_2^T \mathbf{v}_2 \|\tilde{\mathbf{v}}_1\|_2)^2 / 2},$$

$$R(\sigma_v) := \sqrt{(\sigma_u - \sigma_v w - \sigma_v \|\tilde{\mathbf{v}}_1\|_2 \|\tilde{\mathbf{v}}_2\|_2)^2 + \sigma_v^2 (\mathbf{u}_1^T \mathbf{v}_1 \|\tilde{\mathbf{v}}_2\|_2 - \mathbf{u}_2^T \mathbf{v}_2 \|\tilde{\mathbf{v}}_1\|_2)^2 / 2}.$$

We are left to show that both  $Q$  and  $R$  are nondecreasing functions of  $\sigma_v \in [\sigma_u, \infty]$ . By differentiating  $Q$ ,  $R$  with respect to  $\sigma_v$ , we obtain

$$\begin{aligned} \frac{dQ}{d\sigma_v} &= c_Q(\sigma_v) [\sigma_u (\|\tilde{\mathbf{v}}_1\|_2 \|\tilde{\mathbf{v}}_2\|_2 - w) \\ &\quad + \sigma_v \{w^2 + \|\tilde{\mathbf{v}}_1\|_2^2 \|\tilde{\mathbf{v}}_2\|_2^2 + (\mathbf{u}_1^T \mathbf{v}_1)^2 \|\tilde{\mathbf{v}}_2\|_2^2 + (\mathbf{u}_2^T \mathbf{v}_2)^2 \|\tilde{\mathbf{v}}_1\|_2^2\}], \\ \frac{dR}{d\sigma_v} &= c_Q(\sigma_v) [-\sigma_u (\|\tilde{\mathbf{v}}_1\|_2 \|\tilde{\mathbf{v}}_2\|_2 + w) \\ &\quad + \sigma_v \{w^2 + \|\tilde{\mathbf{v}}_1\|_2^2 \|\tilde{\mathbf{v}}_2\|_2^2 + (\mathbf{u}_1^T \mathbf{v}_1)^2 \|\tilde{\mathbf{v}}_2\|_2^2 + (\mathbf{u}_2^T \mathbf{v}_2)^2 \|\tilde{\mathbf{v}}_1\|_2^2\}] \end{aligned}$$

for some nonnegative constants  $c_Q(\sigma_v)$ ,  $c_R(\sigma_v)$ .

By simple algebra, we have  $w^2 + \|\tilde{\mathbf{v}}_1\|_2^2 \|\tilde{\mathbf{v}}_2\|_2^2 + (\mathbf{u}_1^T \mathbf{v}_1)^2 \|\tilde{\mathbf{v}}_2\|_2^2 + (\mathbf{u}_2^T \mathbf{v}_2)^2 \|\tilde{\mathbf{v}}_1\|_2^2 = 1$  so that

$$\frac{dQ}{d\sigma_v} = c_Q(\sigma_v) [\sigma_u (\|\tilde{\mathbf{v}}_1\|_2 \|\tilde{\mathbf{v}}_2\|_2 - w) + \sigma_v].$$

Moreover, since  $\mathbf{u}_1, \mathbf{u}_2, \mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^p$  are all length 1, we have  $|w| \leq 1$  by Cauchy–Schwartz. Hence, by the fact that  $\sigma_v \geq \sigma_u \geq 0$ , we have  $\frac{dQ}{d\sigma_v} \geq 0$ . On the other hand, denote  $a := \mathbf{u}_1^T \mathbf{v}_1$  and  $b := \mathbf{u}_2^T \mathbf{v}_2$  and again by Cauchy–Schwartz we have  $|a| \leq 1$ ,  $|b| \leq 1$ . In addition, we have

$$\begin{aligned} \|\tilde{\mathbf{v}}_1\|_2 &= \sqrt{(\mathbf{v}_1 - \mathbf{u}_1 \mathbf{u}_1^T \mathbf{v}_1)^T (\mathbf{v}_1 - \mathbf{u}_1 \mathbf{u}_1^T \mathbf{v}_1)} \\ &= \sqrt{\mathbf{v}_1^T \mathbf{v}_1 - \mathbf{v}_1^T \mathbf{u}_1 \mathbf{u}_1^T \mathbf{v}_1 - \mathbf{v}_1^T \mathbf{u}_1 \mathbf{u}_1^T \mathbf{v}_1 + \mathbf{v}_1^T \mathbf{u}_1 \mathbf{u}_1^T \mathbf{u}_1 \mathbf{u}_1^T \mathbf{v}_1} \\ &= \sqrt{1 - a^2}. \end{aligned}$$

Similarly, we have  $\|\tilde{\mathbf{v}}_2\|_2 = \sqrt{1 - b^2}$ . Then

$$\begin{aligned} \frac{dR}{d\sigma_v} &= c_Q(\sigma_v)\{\sigma_v - \sigma_u(\|\tilde{\mathbf{v}}_1\|_2\|\tilde{\mathbf{v}}_2\|_2 + w)\} \\ &\geq c_Q(\sigma_v)\sigma_u(1 - \|\tilde{\mathbf{v}}_1\|_2\|\tilde{\mathbf{v}}_2\|_2 - w) \\ &\geq c_Q(\sigma_v)\sigma_u(1 - \sqrt{(1 - a^2)(1 - b^2)} - ab). \end{aligned}$$

Since  $(1 - ab)^2 \geq (1 - a^2)(1 - b^2)$  and  $|ab| \leq 1$ , we obtain  $\frac{dR}{d\sigma_v} \geq 0$ . Therefore, we have shown that  $\|\mathbf{S}(\sigma_v)\| = Q(\sigma_v) + R(\sigma_v)$  is a nondecreasing function with respect to  $\sigma_v$ .

Obviously  $\|M\mathbf{v}_1\mathbf{v}_2^T - M\mathbf{u}_1\mathbf{u}_2^T\| \leq \|\sigma_u\mathbf{v}_1\mathbf{v}_2^T - \sigma_u\mathbf{u}_1\mathbf{u}_2^T\|$  since  $0 < M \leq \sigma_u$ . Applying the monotonicity property proved above, we have  $\|\sigma_u\mathbf{v}_1\mathbf{v}_2^T - \sigma_u\mathbf{u}_1\mathbf{u}_2^T\| \leq \|\sigma_u\mathbf{v}_1\mathbf{v}_2^T - \sigma_v\mathbf{u}_1\mathbf{u}_2^T\|$ . This completes the proof.  $\square$

**Proof of Lemma 4.7** By the observation in the proof of Proposition 4.6, we have

$$\mathbb{E}\|\hat{\Sigma}_0 - \Sigma_0\| \leq \frac{2}{n}\mathbb{E}\|\mathbf{Y}\tilde{\mathbf{Y}}^T\| = \frac{2}{n}\mathbb{E}\left(\sup_{\mathbf{u}, \mathbf{v} \in \mathbb{S}^{p-1}} \sum_{k=1}^n \mathbf{u}^T \mathbf{Y}_k \tilde{\mathbf{Y}}_k^T \mathbf{v}\right) := \frac{2}{n}\mathbb{E}\left(\sup_{\mathbf{u}, \mathbf{v} \in \mathbb{S}^{p-1}} W_{\mathbf{u}, \mathbf{v}}\right).$$

Now consider

$$\begin{aligned} (W_{\mathbf{u}, \mathbf{v}} - W_{\mathbf{u}', \mathbf{v}'})^2 &= \left(\sum_{k=1}^n \mathbf{u}^T \mathbf{Y}_k \tilde{\mathbf{Y}}_k^T \mathbf{v} - \sum_{k=1}^n \mathbf{u}'^T \mathbf{Y}_k \tilde{\mathbf{Y}}_k^T \mathbf{v}'\right)^2 \\ &= \left(\sum_{k=1}^n \mathbf{u}^T \mathbf{Y}_k \tilde{\mathbf{Y}}_k^T \mathbf{v} - \sum_{k=1}^n \mathbf{u}'^T \mathbf{Y}_k \tilde{\mathbf{Y}}_k^T \mathbf{v} + \sum_{k=1}^n \mathbf{u}'^T \mathbf{Y}_k \tilde{\mathbf{Y}}_k^T \mathbf{v} - \sum_{k=1}^n \mathbf{u}'^T \mathbf{Y}_k \tilde{\mathbf{Y}}_k^T \mathbf{v}'\right)^2 \\ &= \left(\sum_{k=1}^n (\mathbf{u} - \mathbf{u}')^T \mathbf{Y}_k \tilde{\mathbf{Y}}_k^T \mathbf{v} + \sum_{k=1}^n \mathbf{u}'^T \mathbf{Y}_k \tilde{\mathbf{Y}}_k^T (\mathbf{v} - \mathbf{v}')\right)^2 \\ &\leq 2\left(\sum_{k=1}^n (\mathbf{u} - \mathbf{u}')^T \mathbf{Y}_k \tilde{\mathbf{Y}}_k^T \mathbf{v}\right)^2 + 2\left(\sum_{k=1}^n \mathbf{u}'^T \mathbf{Y}_k \tilde{\mathbf{Y}}_k^T (\mathbf{v} - \mathbf{v}')\right)^2 \\ &= 2\sum_{d=0}^{n-1} \sum_{|j-k|=d} (\mathbf{u} - \mathbf{u}')^T \mathbf{Y}_j \cdot (\mathbf{u} - \mathbf{u}')^T \mathbf{Y}_k \cdot \mathbf{v}^T \tilde{\mathbf{Y}}_j \cdot \mathbf{v}^T \tilde{\mathbf{Y}}_k \\ &\quad + 2\sum_{d=0}^{n-1} \sum_{|j-k|=d} \mathbf{u}'^T \mathbf{Y}_j \cdot \mathbf{u}'^T \mathbf{Y}_k \cdot (\mathbf{v} - \mathbf{v}')^T \tilde{\mathbf{Y}}_j \cdot (\mathbf{v} - \mathbf{v}')^T \tilde{\mathbf{Y}}_k. \end{aligned}$$

Now denote the conditional expectation  $\mathbb{E}_{\tilde{\mathbf{Y}}} := \mathbb{E}(\cdot | \tilde{\mathbf{Y}})$ . Then,

$$\begin{aligned} \mathbb{E}_{\tilde{\mathbf{Y}}}(W_{\mathbf{u}, \mathbf{v}} - W_{\mathbf{u}', \mathbf{v}'})^2 \\ \leq 2(\mathbf{u} - \mathbf{u}')^T \Sigma_0 (\mathbf{u} - \mathbf{u}') \sum_{j=1}^n \mathbf{v}^T \tilde{\mathbf{Y}}_j \tilde{\mathbf{Y}}_j^T \mathbf{v} \end{aligned}$$



$$\begin{aligned}
& + 2 \sum_{d=1}^{n-1} (\mathbf{u} - \mathbf{u}')^T (\boldsymbol{\Sigma}_d + \boldsymbol{\Sigma}_d^T) (\mathbf{u} - \mathbf{u}') \sum_{(j-k)=d} \mathbf{v}^T \tilde{\mathbf{Y}}_j \cdot \mathbf{v}^T \tilde{\mathbf{Y}}_k \\
& + 2 \sum_{j,k=1}^n \mathbf{u}'^T \boldsymbol{\Sigma}_{|j-k|} \mathbf{u}' \cdot (\mathbf{v} - \mathbf{v}')^T \tilde{\mathbf{Y}}_j \cdot (\mathbf{v} - \mathbf{v}')^T \tilde{\mathbf{Y}}_k \\
& \leq 2(\mathbf{u} - \mathbf{u}')^T \left( \boldsymbol{\Sigma}_0 + 2 \sum_{d=1}^{n-1} \tilde{\boldsymbol{\Sigma}}_d \right) (\mathbf{u} - \mathbf{u}') \sum_{j=1}^n \mathbf{v}^T \tilde{\mathbf{Y}}_j \tilde{\mathbf{Y}}_j^T \mathbf{v} \\
& + 2 \left( \|\boldsymbol{\Sigma}_0\| + 2 \sum_{d=1}^{n-1} \|\boldsymbol{\Sigma}_d\| \right) \sum_{j=1}^n (\mathbf{v} - \mathbf{v}')^T \tilde{\mathbf{Y}}_j \tilde{\mathbf{Y}}_j^T (\mathbf{v} - \mathbf{v}') \\
& \leq 2 \left\| \left( \boldsymbol{\Sigma}_0 + 2 \sum_{d=1}^{n-1} \tilde{\boldsymbol{\Sigma}}_d \right)^{\frac{1}{2}} (\mathbf{u} - \mathbf{u}') \right\|^2 \|\tilde{\mathbf{Y}}\|^2 + 2 \left( \|\boldsymbol{\Sigma}_0\| + 2 \sum_{d=1}^{n-1} \|\boldsymbol{\Sigma}_d\| \right) \|(\mathbf{v} - \mathbf{v}')^T \tilde{\mathbf{Y}}\|^2,
\end{aligned}$$

where the second inequality is followed by defining  $\tilde{\boldsymbol{\Sigma}}_d := (\mathbf{U}_d \boldsymbol{\Lambda}_d \mathbf{U}_d^T + \mathbf{V}_d \boldsymbol{\Lambda}_d \mathbf{V}_d^T)/2$ . Here  $\mathbf{U}_d$ ,  $\mathbf{V}_d$  and  $\boldsymbol{\Lambda}_d$  are left singular vectors, right singular vectors, and singular values of  $\boldsymbol{\Sigma}_d$  for all  $1 \leq d \leq n-1$ , respectively. Note that  $\tilde{\boldsymbol{\Sigma}}_d$  are symmetric and positive semi-definite for all  $d$  and hence so are  $\boldsymbol{\Sigma}_0 + 2 \sum_{d=1}^{n-1} \tilde{\boldsymbol{\Sigma}}_d$ .

Define the following Gaussian process:

$$Y_{\mathbf{u}, \mathbf{v}} := \sqrt{2} \|\tilde{\mathbf{Y}}\| \mathbf{u}^T \left( \boldsymbol{\Sigma}_0 + 2 \sum_{d=1}^{n-1} \tilde{\boldsymbol{\Sigma}}_d \right)^{\frac{1}{2}} \mathbf{g} + \sqrt{2} \left( \|\boldsymbol{\Sigma}_0\| + 2 \sum_{d=1}^{n-1} \|\boldsymbol{\Sigma}_d\| \right)^{\frac{1}{2}} \mathbf{v}^T \tilde{\mathbf{Y}} \mathbf{g}',$$

where  $\mathbf{g}$  and  $\mathbf{g}'$  are independent standard Gaussian random vectors in  $\mathbb{R}^p$  and  $\mathbb{R}^n$ , respectively. Thus, by previous inequality, we have

$$\mathbb{E}_{\tilde{\mathbf{Y}}} (W_{\mathbf{u}, \mathbf{v}} - W_{\mathbf{u}', \mathbf{v}'})^2 \leq \mathbb{E}_{\tilde{\mathbf{Y}}} (Y_{\mathbf{u}, \mathbf{v}} - Y_{\mathbf{u}', \mathbf{v}'})^2.$$

Hence, by Slepian–Fernique inequality [27], we have

$$\begin{aligned}
& \mathbb{E}_{\tilde{\mathbf{Y}}} \sup_{\mathbf{u}, \mathbf{v} \in \mathbb{S}^{p-1}} W_{\mathbf{u}, \mathbf{v}} \\
& \leq \mathbb{E}_{\tilde{\mathbf{Y}}} \sup_{\mathbf{u}, \mathbf{v} \in \mathbb{S}^{p-1}} Y_{\mathbf{u}, \mathbf{v}} \\
& = \sqrt{2} \|\tilde{\mathbf{Y}}\| \cdot \mathbb{E} \sup_{\mathbf{u} \in \mathbb{S}^{p-1}} \mathbf{u}^T \left( \boldsymbol{\Sigma}_0 + 2 \sum_{d=1}^{n-1} \tilde{\boldsymbol{\Sigma}}_d \right)^{\frac{1}{2}} \mathbf{g} + \sqrt{2} \left( \|\boldsymbol{\Sigma}_0\| + 2 \sum_{d=1}^{n-1} \|\boldsymbol{\Sigma}_d\| \right)^{\frac{1}{2}} \cdot \mathbb{E}_{\tilde{\mathbf{Y}}} \sup_{\mathbf{v} \in \mathbb{S}^{p-1}} \mathbf{v}^T \tilde{\mathbf{Y}} \mathbf{g}' \\
& \leq \sqrt{2} \|\tilde{\mathbf{Y}}\| \cdot \mathbb{E} \left\| \left( \boldsymbol{\Sigma}_0 + 2 \sum_{d=1}^{n-1} \tilde{\boldsymbol{\Sigma}}_d \right)^{\frac{1}{2}} \mathbf{g} \right\| + \sqrt{2} \left( \|\boldsymbol{\Sigma}_0\| + 2 \sum_{d=1}^{n-1} \|\boldsymbol{\Sigma}_d\| \right)^{\frac{1}{2}} \cdot \mathbb{E}_{\tilde{\mathbf{Y}}} \|\tilde{\mathbf{Y}} \mathbf{g}'\| \\
& \leq \sqrt{2} \|\tilde{\mathbf{Y}}\| \cdot \sqrt{\text{Tr} \left( \boldsymbol{\Sigma}_0 + 2 \sum_{d=1}^{n-1} \tilde{\boldsymbol{\Sigma}}_d \right)} + \sqrt{2} \left( \|\boldsymbol{\Sigma}_0\| + 2 \sum_{d=1}^{n-1} \|\boldsymbol{\Sigma}_d\| \right)^{\frac{1}{2}} \cdot \sqrt{\text{Tr}(\tilde{\mathbf{Y}} \tilde{\mathbf{Y}}^T)}.
\end{aligned}$$

Taking expectation with respect to  $\tilde{\mathbf{Y}}$  and using the fact that  $\tilde{\mathbf{Y}}$  is an independent copy of  $\mathbf{Y}$ , we obtain

$$\mathbb{E} \sup_{\mathbf{u}, \mathbf{v} \in \mathbb{S}^{p-1}} W_{\mathbf{u}, \mathbf{v}} \leq \sqrt{2} \mathbb{E} \|\mathbf{Y}\| \cdot \sqrt{\text{Tr} \left( \mathbf{\Sigma}_0 + 2 \sum_{d=1}^{n-1} \tilde{\mathbf{\Sigma}}_d \right)} + \sqrt{2} \sqrt{\|\mathbf{\Sigma}_0\| + 2 \sum_{d=1}^{n-1} \|\mathbf{\Sigma}_d\|} \cdot \sqrt{n \text{Tr}(\mathbf{\Sigma}_0)}.$$

This completes the proof of Lemma 4.7.  $\square$

**Proof of Lemma 4.8** Define  $W_{\mathbf{u}, \mathbf{v}} := \mathbf{u}^T \mathbf{Y} \mathbf{v}$ . Then,

$$\begin{aligned} \mathbb{E}(W_{\mathbf{u}, \mathbf{v}} - W_{\mathbf{u}', \mathbf{v}'})^2 &= \mathbb{E}(\mathbf{u}^T \mathbf{Y} \mathbf{v} - \mathbf{u}'^T \mathbf{Y} \mathbf{v}')^2 \\ &\leq 2\mathbb{E}((\mathbf{u} - \mathbf{u}')^T \mathbf{Y} \mathbf{v})^2 + 2\mathbb{E}(\mathbf{u}'^T \mathbf{Y}(\mathbf{v} - \mathbf{v}'))^2 \\ &= 2 \sum_{i,j} (\mathbf{u} - \mathbf{u}')^T \mathbf{\Sigma}_{|i-j|} (\mathbf{u} - \mathbf{u}') v_i v_j + 2 \sum_{i,j} \mathbf{u}'^T \mathbf{\Sigma}_{|i-j|} \mathbf{u}' (v_i - v'_i)(v_j - v'_j). \end{aligned}$$

In addition, define

$$\begin{aligned} \mathbf{\Sigma}_L &:= \begin{bmatrix} \mathbf{\Sigma}_0 & \mathbf{\Sigma}_1 & \cdots & \mathbf{\Sigma}_{n-1} \\ \mathbf{\Sigma}_1^T & \mathbf{\Sigma}_0 & \cdots & \mathbf{\Sigma}_{n-2} \\ \cdots & \cdots & \cdots & \cdots \\ \mathbf{\Sigma}_{n-1}^T & \mathbf{\Sigma}_{n-2}^T & \cdots & \mathbf{\Sigma}_0 \end{bmatrix}, \quad \mathbf{\Sigma}_{L, \mathbf{u}} := \begin{bmatrix} \mathbf{u}^T & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{u}^T & \cdots & \mathbf{0} \\ \cdots & \cdots & \cdots & \cdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{u}^T \end{bmatrix} \mathbf{\Sigma}_L \begin{bmatrix} \mathbf{u} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{u} & \cdots & \mathbf{0} \\ \cdots & \cdots & \cdots & \cdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{u} \end{bmatrix}, \\ \mathbf{\Sigma}_u^\circ &:= (\mathbf{u}^T \mathbf{\Sigma}_0 \mathbf{u}) \mathbf{1}_n \mathbf{1}_n^T, \quad \mathbf{\Sigma}^\circ := \|\mathbf{\Sigma}_0\| \mathbf{1}_n \mathbf{1}_n^T. \end{aligned}$$

Since  $\mathbf{\Sigma}_L$  is a positive semi-definite matrix, we have

$$\mathbf{\Sigma}_{L, \mathbf{u}} \preceq \mathbf{\Sigma}_u^\circ \preceq \mathbf{\Sigma}^\circ$$

for all  $\mathbf{u} \in \mathbb{S}^{p-1}$ , where “ $\preceq$ ” is the Loewner partial order of Hermitian matrices. Hence,

$$\mathbb{E}(W_{\mathbf{u}, \mathbf{v}} - W_{\mathbf{u}', \mathbf{v}'})^2 \leq 2\left\| \left( \mathbf{\Sigma}_0 + 2 \sum_{d=1}^{n-1} \tilde{\mathbf{\Sigma}}_d \right)^{\frac{1}{2}} (\mathbf{u} - \mathbf{u}') \right\|^2 + 2\|\mathbf{\Sigma}_0\| (\mathbf{v} - \mathbf{v}')^T \mathbf{1}_n \mathbf{1}_n^T (\mathbf{v} - \mathbf{v}').$$

Then, define the following Gaussian process:

$$Y_{\mathbf{u}, \mathbf{v}} := \sqrt{2} \mathbf{u}^T \left( \mathbf{\Sigma}_0 + 2 \sum_{d=1}^{n-1} \tilde{\mathbf{\Sigma}}_d \right)^{\frac{1}{2}} \mathbf{g} + \sqrt{2} \|\mathbf{\Sigma}_0\|^{\frac{1}{2}} \mathbf{v}^T \mathbf{g}',$$

where  $\mathbf{g} \in \mathbb{R}^p$ ,  $\mathbf{g}' \in \mathbb{R}^n$  are independent Gaussian random vectors with mean  $\mathbf{0}$  and covariance matrices  $\mathbf{I}_p$  and  $\mathbf{1}_n \mathbf{1}_n^T$ , respectively. Thus, by previous inequality, we have

$$\mathbb{E}(W_{\mathbf{u}, \mathbf{v}} - W_{\mathbf{u}', \mathbf{v}'})^2 \leq \mathbb{E}(Y_{\mathbf{u}, \mathbf{v}} - Y_{\mathbf{u}', \mathbf{v}'})^2.$$

Hence, by Slepian–Fernique inequality, we have

$$\begin{aligned}
 \mathbb{E} \sup_{\mathbf{u}, \mathbf{v} \in \mathbb{S}^{p-1}} W_{\mathbf{u}, \mathbf{v}} &\leq \mathbb{E} \sup_{\mathbf{u}, \mathbf{v} \in \mathbb{S}^{p-1}} Y_{\mathbf{u}, \mathbf{v}} \\
 &= \sqrt{2} \mathbb{E} \sup_{\mathbf{u} \in \mathbb{S}^{p-1}} \mathbf{u}^T \left( \boldsymbol{\Sigma}_0 + 2 \sum_{d=1}^{n-1} \tilde{\boldsymbol{\Sigma}}_d \right)^{\frac{1}{2}} \mathbf{g} + \sqrt{2} \|\boldsymbol{\Sigma}_0\|^{\frac{1}{2}} \cdot \mathbb{E} \sup_{\mathbf{v} \in \mathbb{S}^{p-1}} \mathbf{v}^T \mathbf{g}' \\
 &\leq \sqrt{2} \mathbb{E} \left\| \left( \boldsymbol{\Sigma}_0 + 2 \sum_{d=1}^{n-1} \tilde{\boldsymbol{\Sigma}}_d \right)^{\frac{1}{2}} \mathbf{g} \right\| + \sqrt{2} \|\boldsymbol{\Sigma}_0\|^{\frac{1}{2}} \cdot \mathbb{E} \|\mathbf{g}'\| \\
 &\leq \sqrt{2} \sqrt{\text{Tr} \left( \boldsymbol{\Sigma}_0 + 2 \sum_{d=1}^{n-1} \tilde{\boldsymbol{\Sigma}}_d \right)} + \sqrt{2} \|\boldsymbol{\Sigma}_0\|^{\frac{1}{2}} \cdot \sqrt{n}.
 \end{aligned}$$

This completes the proof of Lemma 4.8.  $\square$

#### 4.4 Proof of Results in Section 3

**Proof of Theorem 3.1** We first examine Assumptions (A1) and (A4). First of all, we will study VAR(1) model, i.e.,  $\mathbf{Y}_t = \mathbf{A}\mathbf{Y}_{t-1} + \mathbf{E}_t$ . Notice that for VAR(1), we could rewrite the original sequence as a moving-average model, i.e.,  $\mathbf{Y}_t = \sum_{j=0}^{\infty} \mathbf{A}^j \mathbf{E}_{t-j}$ . For any  $\mathbf{u} \in \mathbb{R}^p$ , we have

$$\begin{aligned}
 \|\mathbf{u}^T \mathbf{Y}_t\|_{\psi_2} &= \left\| \sum_{j=0}^{\infty} \mathbf{u}^T \mathbf{A}^j \mathbf{E}_{t-j} \right\|_{\psi_2} \\
 &\leq C \left( \sum_{j=0}^{\infty} \|\mathbf{u}^T \mathbf{A}^j \mathbf{E}_{t-j}\|_{\psi_2}^2 \right)^{\frac{1}{2}} \\
 &\leq C c' \left( \sum_{j=0}^{\infty} \|\mathbf{u}^T \mathbf{A}^j \mathbf{E}_{t-j}\|_{L(2)}^2 \right)^{\frac{1}{2}} = C c' \|\mathbf{u}^T \mathbf{Y}_t\|_{L(2)}
 \end{aligned}$$

for some universal constant  $C > 0$ . Here, the second line and last equality are followed by the fact that  $\{\mathbf{E}_t\}_{t \in \mathbb{Z}}$  is a sequence of independent random vector and the third line by the moment assumption on  $\{\mathbf{E}_t\}_{t \in \mathbb{Z}}$ . Since  $\mathbf{Y}_{t-1}$  is a stable process when  $\|\mathbf{A}\| < 1$ ,  $\|\mathbf{u}^T \mathbf{Y}_t\|_{\psi_2} \leq C c' \|\mathbf{u}^T \mathbf{Y}_t\|_{L(2)} < \infty$  for all  $\mathbf{u} \in \mathbb{R}^p$ .

Denote  $\tilde{\mathbf{Y}}_t := (\mathbf{Y}_t^T \dots \mathbf{Y}_{t-d}^T)^T$  and  $\tilde{\mathbf{E}}_t := (\mathbf{E}_t^T \mathbf{0}^T \dots \mathbf{0}^T)^T$ . For  $\{\mathbf{Y}_t\}_{t \in \mathbb{Z}}$  generated from a VAR( $d$ ) model,  $\{\tilde{\mathbf{Y}}_t\}_{t \in \mathbb{Z}}$  is a VAR(1) process, i.e.,  $\tilde{\mathbf{Y}}_t = \tilde{\mathbf{A}} \cdot \tilde{\mathbf{Y}}_{t-1} + \tilde{\mathbf{E}}_t$ . Thus, by previous argument, taking any  $\mathbf{v} \in \mathbb{R}^{p(d+1)}$  where only the first  $p$  digits are nonzero and denoting  $\mathbf{v}' \in \mathbb{R}^p$  to be first- $p$  part of  $\mathbf{v}$ , we have  $\|\mathbf{v}'^T \mathbf{Y}_t\|_{\psi_2} = \|\mathbf{v}^T \tilde{\mathbf{Y}}_t\|_{\psi_2} \leq C \|\mathbf{v}^T \tilde{\mathbf{Y}}_t\|_{L(2)} = C \|\mathbf{v}'^T \tilde{\mathbf{Y}}_t\|_{L(2)} < \infty$  for some constant  $C > 0$  only depending on  $c'$  where the last inequality is followed by the fact that  $\{\mathbf{Y}_t\}$  is a stable process (see Lemma 4.10). Assumptions (A1) and (A4) are verified.

Then, we examine Assumption (A2). Without loss of generality, take  $j = 0$  in Assumption (A2). Let  $\{\tilde{Y}_t\}_{t=1-d}^0$  be a sequence of random vectors independent of  $\{Y_t\}_{t \leq 0}$  and identically distributed as  $\{Y_t\}_{t=1-d}^0$ . Define  $\tilde{Y}_t = A_1 \tilde{Y}_{t-1} + \cdots + A_d \tilde{Y}_{t-d} + E_t$  for every  $t > 0$ . It is obvious that  $\{\tilde{Y}_t\}_{t > 0}$  is independent of  $\{\tilde{Y}_t\}_{t \leq 0}$  and identically distributed as  $\{Y_t\}_{t > 0}$ . Moreover, for any  $t \geq 1$ , we have

$$\begin{aligned} \|Y_t - \tilde{Y}_t\|_{2, L(1+\epsilon)} &= \{\mathbb{E}\|A_1 Y_{t-1} + \cdots + A_d Y_{t-d} + E_t \\ &\quad - (A_1 \tilde{Y}_{t-1} + \cdots + A_d \tilde{Y}_{t-d} + E_t)\|_2^{1+\epsilon}\}^{\frac{1}{1+\epsilon}} \\ &\leq \{\mathbb{E}\|A_1 (Y_{t-1} - \tilde{Y}_{t-1}) + \cdots + A_d (Y_{t-d} - \tilde{Y}_{t-d})\|_2^{1+\epsilon}\}^{\frac{1}{1+\epsilon}} \\ &\leq \sum_{k=1}^d a_k \{\mathbb{E}\|Y_{t-k} - \tilde{Y}_{t-k}\|_2^{1+\epsilon}\}^{\frac{1}{1+\epsilon}}, \end{aligned}$$

where the third line is followed by  $\|\cdot\|_{L(1+\epsilon)}$  is a norm for  $\epsilon > 0$ . Denoting  $\phi_t = \|Y_t - \tilde{Y}_t\|_{2, L(1+\epsilon)}$ , we have  $\phi_t \leq \sum_{k=1}^d a_k \phi_{t-k}$ . Let  $v$  be the unit vector with 1 at first position and 0 elsewhere. Then, by iteration, we have

$$v^T (\phi_t, \dots, \phi_{t-d+1})^T \leq v^T \bar{A}^t (\phi_0, \dots, \phi_{1-d})^T \leq \|\bar{A}^t\| \|(\phi_0, \dots, \phi_{1-d})^T\|_2.$$

Note that  $\phi_t = C\kappa_*$  for  $t \leq 0$  by Assumption (A1) for some constant  $C > 0$  only depending on  $\epsilon$ . By the following lemma which provides sufficient and necessary conditions for matrix  $\bar{A}$  to have spectral radius strictly less than 1, we could choose some arbitrary  $\rho_1$  such that  $\rho(\bar{A}) < \rho_1 < 1$ .  $\square$

**Lemma 4.10** For  $\bar{A}$  defined above,  $\rho(\bar{A}) < 1$  if and only if  $\sum_{k=1}^d a_k < 1$ , where  $\rho(\bar{A})$  is the spectral radius of  $\bar{A}$ .

**Proof of Lemma 4.10** The result is well known and here we include a proof merely for completeness. First of all, we prove the sufficient condition. A key observation is that the characteristic equation  $\det(\bar{A} - \lambda I_d) = 0$  for matrix  $\bar{A}$  is

$$f(\lambda) = \lambda^d - a_1 \lambda^{d-1} - \cdots - a_{d-1} \lambda - a_d = 0.$$

Assume  $\sum_{j=1}^d a_j \geq 1$ . We obtain  $f(1) = 1 - \sum_{j=1}^d a_j \leq 0$  and  $f(\infty) = \infty$ . By continuity of  $f(\lambda)$ , there exists at least one root whose modulus is greater than or equal to 1. This contradicts with the fact that  $\rho(\bar{A})$  is strictly less than 1.

Secondly, we prove the necessary condition. Suppose there exists a root  $z \in \mathbb{C}$  (the set of complex numbers) of  $f(\lambda)$  such that  $|z| \geq 1$ . Here  $|z|$  is the modulus of  $z$ . Then

$$|z|^d = |a_1 z^{d-1} + \cdots + a_{d-1} z + a_d| \leq a_1 |z|^{d-1} + \cdots + a_{d-1} |z| + a_d.$$

Since  $|z| \geq 1$ , we have  $|z|^k \leq |z|^d$  for  $0 \leq k \leq d-1$ . Hence,  $|z|^d \leq (a_1 + \cdots + a_d) |z|^d$  implies  $a_1 + \cdots + a_d \geq 1$ . This contradicts the fact that  $\sum_{j=1}^d a_j$  is strictly less than 1. This completes the proof.  $\square$

By Gelfand's formula, there exists a  $K > 0$ , such that for all  $t \geq K$ ,  $\|\bar{\mathbf{A}}^t\| < \rho_1^t$ . For  $t < K$ , we have

$$\phi_t \leq 2d\kappa_* \left( \frac{\|\bar{\mathbf{A}}\|}{\rho_1} \right)^K \rho_1^t.$$

For  $t \geq K$ , we have  $\phi_t \leq Cd\kappa_*\rho_1^t$  for some constant  $C > 0$  only depending on  $\epsilon$ . Taking  $\gamma_1 = Cd(\kappa_*/\kappa_1)(\|\bar{\mathbf{A}}\|/\rho_1)^K$  for some constant  $C > 0$  only depending on  $\epsilon$  and  $\gamma_2 = \log(\rho_1^{-1})$  verifies Assumption (A2).

Lastly, we verify Assumption (A3). Following the same construction as in verifying Assumption (A2), we have for any  $\mathbf{u} \in \mathbb{S}^{p-1}$ ,

$$\begin{aligned} & \|(\mathbf{Y}_t - \tilde{\mathbf{Y}}_t)^T \mathbf{u}\|_{L(1+\epsilon)} \\ &= (\mathbb{E}|\{\mathbf{A}_1 \mathbf{Y}_{t-1} + \cdots + \mathbf{A}_d \mathbf{Y}_{t-d} + \mathbf{E}_t - (\mathbf{A}_1 \tilde{\mathbf{Y}}_{t-1} + \cdots + \mathbf{A}_d \tilde{\mathbf{Y}}_{t-d} + \mathbf{E}_t)\}^T \mathbf{u}|^{1+\epsilon})^{\frac{1}{1+\epsilon}} \\ &\leq (\mathbb{E}|\{\mathbf{A}_1 \mathbf{Y}_{t-1} + \cdots + \mathbf{A}_d \mathbf{Y}_{t-d} - (\mathbf{A}_1 \tilde{\mathbf{Y}}_{t-1} + \cdots + \mathbf{A}_d \tilde{\mathbf{Y}}_{t-d})\}^T \mathbf{u}|^{1+\epsilon})^{\frac{1}{1+\epsilon}} \\ &\leq \sum_{k=1}^d a_k \{\mathbb{E}|(\mathbf{Y}_{t-k} - \tilde{\mathbf{Y}}_{t-k})^T \mathbf{u}_k|^{1+\epsilon}\}^{\frac{1}{1+\epsilon}}, \end{aligned}$$

for  $\mathbf{u}_k := \mathbf{A}_k \mathbf{u} / \|\mathbf{A}_k \mathbf{u}\|_2$ ,  $k \in \{1, \dots, d\}$ . The result follows as we follow the same arguments to verify Assumption (A2). This completes the proof of Theorem 3.1.  $\square$

**Proof of Theorem 3.2** First of all, we verify Assumptions (A1) and (A4). It is trivial that Assumptions (A1) and (A4) are satisfied if  $W_t = 0$  almost surely for all  $t \in \mathbb{Z}$ . If  $W_t \neq 0$  almost surely, then for all  $\mathbf{u} \in \mathbb{R}^p$ ,  $\|\mathbf{u}^T \mathbf{Y}_t\|_{\psi_2} \leq \|W_t\|_{L(\infty)} \|\mathbf{u}^T \mathbf{E}_t\|_{\psi_2} \leq c' \kappa_W \|\mathbf{u}^T \mathbf{E}_t\|_{L(2)} \leq c' \frac{\kappa_W}{\inf_{t \in \mathbb{Z}} \|W_t\|_{L(2)}} \|\mathbf{u}^T \mathbf{Y}_t\|_{L(2)} < \infty$ . This verifies Assumptions (A1) and (A4).

For Assumption (A2), without loss of generality, take  $j = 0$ . Since  $\{W_t\}_{t \in \mathbb{Z}}$  is a sequence of uniformly bounded  $\tau$ -mixing random variables, we may find  $\{\tilde{W}_t\}_{t > 0}$  which is independent of  $\{W_t\}_{t \leq 0}$ , identically distributed as  $\{W_t\}_{t > 0}$ , and for any  $t \geq 1$ ,

$$\mathbb{E}|\tilde{W}_t - W_t| \leq \kappa_W \gamma_5 \exp\{-\gamma_6(t-1)\}.$$

Define  $\tilde{\mathbf{Y}}_t := \tilde{W}_t \mathbf{E}_t$  for all  $t \geq 1$ . It is obvious that  $\{\tilde{\mathbf{Y}}_t\}_{t > 0}$  is independent of  $\{\mathbf{Y}_t\}_{t \leq 0}$  and identically distributed as  $\{\mathbf{Y}_t\}_{t > 0}$ . Moreover, for any integer  $t \geq 1$ ,

$$\begin{aligned} \|\mathbf{Y}_t - \tilde{\mathbf{Y}}_t\|_{L(1+\epsilon)} &\leq (\mathbb{E}\|W_t \mathbf{E}_t - \tilde{W}_t \mathbf{E}_t\|_2^{1+\epsilon})^{\frac{1}{1+\epsilon}} \\ &\leq (\mathbb{E}|W_t - \tilde{W}_t| \cdot |W_t - \tilde{W}_t|^{1+\epsilon})^{1+\epsilon} (\mathbb{E}\|\mathbf{E}_t\|_2^{1+\epsilon})^{\frac{1}{1+\epsilon}} \\ &\leq C\kappa'_* \kappa_W \gamma_5^{\frac{1}{1+\epsilon}} \exp\left\{-\frac{1}{1+\epsilon} \gamma_6(t-1)\right\} \end{aligned}$$

for some constant  $C > 0$  only depending on  $\epsilon$ . Taking  $\gamma_1 = C\kappa'_* \kappa_W \gamma_5^{\frac{1}{1+\epsilon}} / \kappa_1$  and  $\gamma_2 = \frac{1}{1+\epsilon} \gamma_6$  verifies Assumption (A2).

For Assumption (A3), without loss of generality, take  $j = 0$ . Let  $\{\tilde{Y}_t\}_{t>0}$  be the same construction as above. For any integer  $t \geq 1$ ,

$$\begin{aligned} \sup_{u \in \mathbb{S}^{p-1}} \|(Y_t - \tilde{Y}_t)^T u\|_{L(1+\epsilon)} &= \sup_{u \in \mathbb{S}^{p-1}} \{\mathbb{E}|(W_t E_t - \tilde{W}_t E_t)^T u|^{1+\epsilon}\}^{\frac{1}{1+\epsilon}} \\ &= (\mathbb{E}|W_t - \tilde{W}_t|^{1+\epsilon})^{\frac{1}{1+\epsilon}} \sup_{u \in \mathbb{S}^{p-1}} (\mathbb{E}|E_t^T u|^{1+\epsilon})^{\frac{1}{1+\epsilon}} \\ &\leq C\kappa'_1 \kappa_W \gamma_5^{\frac{1}{\epsilon}} \exp\left\{-\frac{1}{1+\epsilon} \gamma_6(t-1)\right\} \end{aligned}$$

for some constant  $C > 0$  only depending on  $\epsilon$ . Taking  $\gamma_3 = C\kappa'_1 \kappa_W \gamma_5^{\frac{1}{\epsilon}}/\kappa_1$  and  $\gamma_4 = \frac{1}{1+\epsilon} \gamma_6$  verifies Assumption (A2). This completes the proof of Theorem 3.2.  $\square$

**Proof of Theorem 3.3** We first verify Assumptions (A2) and (A3). Without loss of generality, take  $j = 0$  in Assumption (A2). Let  $\tilde{Y}_0$  be a random vector independent of  $\{Y_t\}_{t \leq 0}$  and identically distributed as  $Y_0$ . Define  $\tilde{Y}_t = \mathbf{A}\tilde{Y}_{t-1} + H(\tilde{Y}_{t-1})E_t$  for every  $t \geq 1$ . It is obvious that  $\{\tilde{Y}_t\}_{t>0}$  is independent of  $\{Y_t\}_{t \leq 0}$  and identically distributed as  $\{Y_t\}_{t>0}$ . We obtain for any  $t \geq 1$ ,

$$\begin{aligned} \|\|Y_t - \tilde{Y}_t\|_2\|_{L(1+\epsilon)} &= [\mathbb{E}\|\mathbf{A}Y_{t-1} + H(Y_{t-1})E_t - \{\mathbf{A}\tilde{Y}_{t-1} + H(\tilde{Y}_{t-1})E_t\}\|_2^{1+\epsilon}]^{\frac{1}{1+\epsilon}} \\ &\leq [\mathbb{E}\|\mathbf{A}Y_{t-1} - \mathbf{A}\tilde{Y}_{t-1} + \{H(Y_{t-1}) - H(\tilde{Y}_{t-1})\}E_t\|_2^{1+\epsilon}]^{\frac{1}{1+\epsilon}} \\ &\leq (a_1 + a_2)\|\|Y_{t-1} - \tilde{Y}_{t-1}\|_2\|_{L(1+\epsilon)}. \end{aligned}$$

By iteration, we obtain

$$\|\|Y_t - \tilde{Y}_t\|_2\|_{L(1+\epsilon)} \leq (a_1 + a_2)^t (\mathbb{E}\|Y_0 - \tilde{Y}_0\|_2^{1+\epsilon})^{\frac{1}{1+\epsilon}} \leq C\kappa_*(a_1 + a_2)^t$$

for some constant  $C > 0$  only depending on  $\epsilon$ . Taking  $\gamma_1 = C\kappa_*/\kappa_1$  and  $\gamma_2 = -\log(a_1 + a_2)$  verifies Assumption (A2).

For Assumption (A3), following the construction above, we have for any  $u \in \mathbb{S}^{p-1}$  and  $t \geq 1$ ,

$$\begin{aligned} \|(Y_t - \tilde{Y}_t)^T u\|_{L(1+\epsilon)} &= [\mathbb{E}\{|\mathbf{A}Y_{t-1} + H(Y_{t-1})E_t - (\mathbf{A}\tilde{Y}_{t-1} + H(\tilde{Y}_{t-1})E_t)|^T u|^{1+\epsilon}\}]^{\frac{1}{1+\epsilon}} \\ &\leq [\mathbb{E}\{|\mathbf{A}Y_{t-1} - \mathbf{A}\tilde{Y}_{t-1} + (H(Y_{t-1}) - H(\tilde{Y}_{t-1}))E_t|^T u|^{1+\epsilon}\}]^{\frac{1}{1+\epsilon}} \\ &\leq a_1 \|(Y_{t-1} - \tilde{Y}_{t-1})^T v\|_{L(1+\epsilon)} + a_2 \frac{\kappa'_1}{\kappa_*} \|\|Y_{t-1} - \tilde{Y}_{t-1}\|_2\|_{L(1+\epsilon)}, \end{aligned}$$

where  $v := \mathbf{A}u/\|\mathbf{A}u\|_2 \in \mathbb{S}^{p-1}$ . By iteration, we obtain

$$\|(Y_t - \tilde{Y}_t)^T u\|_{L(1+\epsilon)} \leq C \left\{ \kappa_1 a_1^t + 2\kappa_* \frac{\kappa'_1}{\kappa_*} a_2 \sum_{\ell=0}^{t-1} a_1^\ell (a_1 + a_2)^{t-1-\ell} \right\}$$

$$\leq C(a_1 + a_2)^t \max\left(\kappa_* \frac{\kappa'_1}{\kappa'_*}, \kappa_1\right)$$

for some constant  $C > 0$  only depending on  $\epsilon$ . Taking  $\gamma_3 = C \max(\frac{\kappa_* \kappa'_1}{\kappa'_1 \kappa'_*}, 1)$  and  $\gamma_4 = -\log(a_1 + a_2)$  verifies Assumption (A3).

By further assuming that  $\{\mathbf{Y}_t\}$  is a stationary process and  $H(\cdot)$  is uniformly bounded, we have that for all  $t \in \mathbb{Z}$ ,  $\sup_{\mathbf{u} \in \mathbb{S}^{p-1}} \|\mathbf{u}^T \mathbf{Y}_t\|_{\psi_2} \leq \|\mathbf{A}\| \sup_{\mathbf{u} \in \mathbb{S}^{p-1}} \|\mathbf{u}^T \mathbf{Y}_{t-1}\|_{\psi_2} + D_2 \sup_{\mathbf{v} \in \mathbb{S}^{p-1}} \|\mathbf{v}^T \mathbf{E}_t\|$ . By stationarity, this renders  $\kappa_1 = \sup_{\mathbf{u} \in \mathbb{S}^{p-1}} \|\mathbf{u}^T \mathbf{Y}_t\|_{\psi_2} \leq \frac{1}{1 - \|\mathbf{A}\|} D_2 \kappa'_1 < \infty$ . Similar argument applies to  $\kappa_*$ . This verifies Assumption (A1) under additional assumptions and completes the proof of Theorem 3.3.  $\square$

## Appendix

### A Proof of Theorem 4.3

In this “Appendix,” we present the proof of Theorem 4.3, which slightly extends the Bernstein-type inequality proven by Banna et al. [3] in which the random matrix sequence is assumed to be  $\beta$ -mixing. The proof is largely identical to theirs, and we include it here mainly for completeness.

In the following,  $\tau_k$  is abbreviation of  $\tau(k)$  for  $k \geq 1$ . If a matrix  $\mathbf{X}$  is positive semi-definite, denote it as  $\mathbf{X} \geq 0$ . For any  $x > 0$ , we define  $h(x) = x^{-2}(e^x - x - 1)$ . Denote the floor, ceiling, and integer parts of a real number  $x$  by  $\lfloor x \rfloor$ ,  $\lceil x \rceil$ , and  $[x]$ . For any two real numbers  $a, b$ , denote  $a \vee b := \max\{a, b\}$ . Denote the exponential of matrix  $\mathbf{X}$  as  $\exp(\mathbf{X}) = \mathbf{I}_p + \sum_{q=1}^{\infty} \mathbf{X}^q / q!$ . Letting  $\sigma_1$  and  $\sigma_2$  be two sigma fields, denote  $\sigma_1 \vee \sigma_2$  to be the smallest sigma field that contains  $\sigma_1$  and  $\sigma_2$  as sub-sigma fields.

A roadmap of this “Appendix” is as follows. Section A.1 formally introduces the concept of  $\tau$ -mixing coefficient. Section A.2 previews the proof of Theorem 4.3 and indicates some major differences from the proofs in [3]. Section A.3 contains the construction of Cantor-like set which is essential for decoupling dependent matrices. Section A.4 develops a major decoupling lemma for  $\tau$ -mixing random matrices and will be used in Sect. A.6 to prove Lemma A.4. Then Sect. A.5 finishes the proof of Theorem 4.3.

#### A.1 Introduction to $\tau$ -Mixing Random Sequence

This section introduces the  $\tau$ -mixing coefficient. Consider  $(\Omega, \mathcal{F}, \mathbb{P})$  to be a probability space,  $X$  an  $L_1$ -integrable random variable taking value in a Polish space  $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$ , and  $\mathcal{A}$  a sigma algebra of  $\mathcal{F}$ . The  $\tau$ -measure of dependence between  $X$  and  $\mathcal{A}$  is defined to be

$$\tau(\mathcal{A}, X; \|\cdot\|_{\mathcal{X}}) = \left\| \sup_{g \in \Lambda(\|\cdot\|_{\mathcal{X}})} \left\{ \int g(x) \mathbb{P}_{X|\mathcal{A}}(dx) - \int g(x) \mathbb{P}_X(dx) \right\} \right\|_{L(1)},$$

where  $\mathbb{P}_X$  is the distribution of  $X$ ,  $\mathbb{P}_{X|\mathcal{A}}$  is the conditional distribution of  $X$  given  $\mathcal{A}$ , and  $\Lambda(\|\cdot\|_{\mathcal{X}})$  stands for the set of 1-Lipschitz functions from  $\mathcal{X}$  to  $\mathbb{R}$  with respect to the norm  $\|\cdot\|_{\mathcal{X}}$ .

The following two lemmas from [13] and [12] characterize the intrinsic “coupling property” of  $\tau$ -measure of dependence, which will be heavily exploited in the derivation of our results.

**Lemma A.1** (Lemma 3 in [13]) *Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space,  $X$  be an integrable random variable with values in a Banach space  $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$  and  $\mathcal{A}$  a sigma algebra of  $\mathcal{F}$ . If  $Y$  is a random variable distributed as  $X$  and independent of  $\mathcal{A}$ , then*

$$\tau(\mathcal{A}, X; \|\cdot\|_{\mathcal{X}}) \leq \mathbb{E}\|X - Y\|_{\mathcal{X}}.$$

**Lemma A.2** (Lemma 5.3 in [12]) *Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space,  $\mathcal{A}$  be a sigma algebra of  $\mathcal{F}$ , and  $X$  be a random variable with values in a Polish space  $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$ . Assume that  $\int \|x - x_0\|_{\mathcal{X}} \mathbb{P}_X(dx)$  is finite for any  $x_0 \in \mathcal{X}$ . Assume that there exists a random variable  $U$  uniformly distributed over  $[0, 1]$ , independent of the sigma algebra generated by  $X$  and  $\mathcal{A}$ . Then, there exists a random variable  $\tilde{X}$ , measurable with respect to  $\mathcal{A} \vee \sigma(X) \vee \sigma(U)$ , independent of  $\mathcal{A}$  and distributed as  $X$ , such that*

$$\tau(\mathcal{A}, X; \|\cdot\|_{\mathcal{X}}) = \mathbb{E}\|X - \tilde{X}\|_{\mathcal{X}}.$$

Let  $\{X_j\}_{j \in J}$  be a set of  $\mathcal{X}$ -valued random variables with index set  $J$  of finite cardinality. Then, define

$$\tau(\mathcal{A}, \{X_j\}_{j \in J}; \|\cdot\|_{\mathcal{X}}) = \left\| \sup_{g \in \Lambda(\|\cdot\|'_{\mathcal{X}})} \left\{ \int g(x) \mathbb{P}_{\{X_j\}_{j \in J}|\mathcal{A}}(dx) - \int g(x) \mathbb{P}_{\{X_j\}_{j \in J}}(dx) \right\} \right\|_{L(1)},$$

where  $\mathbb{P}_{\{X_j\}_{j \in J}}$  is the distribution of  $\{X_j\}_{j \in J}$ ,  $\mathbb{P}_{\{X_j\}_{j \in J}|\mathcal{A}}$  is the conditional distribution of  $\{X_j\}_{j \in J}$  given  $\mathcal{A}$ , and  $\Lambda(\|\cdot\|'_{\mathcal{X}})$  stands for the set of 1-Lipschitz functions from  $\underbrace{\mathcal{X} \times \cdots \times \mathcal{X}}_{\text{card}(J)}$  to  $\mathbb{R}$  with respect to the norm  $\|x\|'_{\mathcal{X}} := \sum_{j \in J} \|x_j\|_{\mathcal{X}}$  induced by  $\|\cdot\|_{\mathcal{X}}$

for any  $x = (x_1, \dots, x_J) \in \mathcal{X}^{\text{card}(J)}$ .

Using these concepts, for a sequence of temporally dependent data  $\{X_t\}_{t \in \mathbb{Z}}$ , we are ready to define measure of temporal correlation strength as follows,

$$\begin{aligned} \tau(k; \{X_t\}_{t \in \mathbb{Z}}, \|\cdot\|_{\mathcal{X}}) \\ := \sup_{i > 0} \max_{1 \leq \ell \leq i} \frac{1}{\ell} \sup \{ \tau(\sigma(X_{-\infty}^a), \{X_{j_1}, \dots, X_{j_\ell}\}; \|\cdot\|_{\mathcal{X}}), a + k \leq j_1 < \cdots < j_\ell \}, \end{aligned}$$

where the inner supremum is taken over all  $a \in \mathbb{Z}$  and all  $\ell$ -tuples  $(j_1, \dots, j_\ell)$ .  $\{X_t\}_{t \in \mathbb{Z}}$  is said to be  $\tau$ -mixing if  $\tau(k; \{X_t\}_{t \in \mathbb{Z}}, \|\cdot\|_{\mathcal{X}})$  converges to zero as  $k \rightarrow \infty$ . In [12], the authors gave numerous examples of random sequences that are  $\tau$ -mixing.



## A.2 Overview of Proof of Theorem 4.3

The proof of Theorem 4.3 follows largely the proof of Theorem 1 in [3]. Section A.3 reviews the Cantor-like set construction developed and used in [22] and [3]. Lemma A.3 is a slight extension of Lemma 8 in [3]. The major difference is that the 0–1 function used to quantify the distance between two random matrices under  $\beta$ -mixing by Berbee's decoupling lemma [4] is replaced by an absolute distance function, which is used under  $\tau$ -mixing by Lemma A.1 [13]. Proofs of Lemma A.4 and the rest of Theorem 4.3 follow largely the proofs of Proposition 7 and Theorem 1 in [3] respectively, though with more algebras involved.

## A.3 Construction of Cantor-Like Set

We follow [3] to construct the Cantor-like set  $K_B$  for  $\{1, \dots, B\}$ . Let  $\delta = \frac{\log 2}{2 \log B}$  and  $\ell_B = \sup\{k \in \mathbb{Z}^+ : \frac{B\delta(1-\delta)^{k-1}}{2^k} \geq 2\}$ . We abbreviate  $\ell := \ell_B$ . Let  $n_0 = B$  and for  $j \in \{1, \dots, \ell\}$ ,

$$n_j = \left\lceil \frac{B(1-\delta)^j}{2^j} \right\rceil \quad \text{and} \quad d_{j-1} = n_{j-1} - 2n_j.$$

We start from the set  $\{1, \dots, B\}$  and divide the set into three disjoint subsets  $I_1^1, J_0^1, I_1^2$  so that  $\text{card}(I_1^1) = \text{card}(I_1^2) = n_1$  and  $\text{card}(J_0^1) = d_0$ . Specifically,

$$I_1^1 = \{1, \dots, n_1\}, \quad J_0^1 = \{n_1 + 1, \dots, n_1 + d_0\}, \quad I_1^2 = \{n_1 + d_0 + 1, \dots, 2n_1 + d_0\},$$

where  $B = 2n_1 + d_0$ . Then, we divide  $I_1^1, I_1^2$  with  $J_0^1$  unchanged.  $I_1^1$  is divided into three disjoint subsets  $I_2^1, J_1^1, I_2^2$  in the same way as the previous step with  $\text{card}(I_2^1) = \text{card}(I_2^2) = n_2$  and  $\text{card}(J_1^1) = d_1$ . We obtain

$$I_1^1 = \{1, \dots, n_2\}, \quad J_1^1 = \{n_2 + 1, \dots, n_2 + d_1\}, \quad I_2^2 = \{n_2 + d_1 + 1, \dots, 2n_2 + d_1\},$$

where  $n_1 = 2n_2 + d_1$ . Similarly,  $I_1^2$  is divided into  $I_2^3, J_1^2, I_2^4$  with  $\text{card}(I_2^3) = \text{card}(I_2^4) = n_2$  and  $\text{card}(J_1^2) = d_1$ . We obtain

$$I_2^3 = \{2n_2 + d_0 + d_1 + 1, \dots, 3n_2 + d_0 + d_1\}, \quad J_1^2 = \{3n_2 + d_0 + d_1 + 1, \dots, 3n_2 + d_0 + 2d_1\}, \\ I_2^4 = \{3n_2 + d_0 + 2d_1 + 1, \dots, 4n_2 + d_0 + 2d_1\},$$

where  $B = 4n_2 + d_0 + 2d_1$ .

Suppose we iterate this process for  $k$  times ( $k \in \{1, \dots, \ell\}$ ) with intervals  $I_k^i, i \in \{1, \dots, 2^k\}$ . For each  $I_k^i$ , we divide it into three disjoint subsets  $I_{k+1}^{2i-1}, J_k^i, I_{k+1}^{2i}$  so that  $\text{card}(I_{k+1}^{2i-1}) = \text{card}(I_{k+1}^{2i}) = n_{k+1}$  and  $\text{card}(J_k^i) = d_k$ . More specifically, if  $I_k^i = \{a_k^i, \dots, b_k^i\}$ , then

$$I_{k+1}^{2i-1} = \{a_k^i, \dots, a_k^i + n_{k+1} - 1\}, \quad J_k^i = \{a_k^i + n_{k+1}, \dots, a_k^i + n_{k+1} + d_k - 1\}, \\ I_{k+1}^{2i} = \{a_k^i + n_{k+1} + d_k, \dots, a_k^i + 2n_{k+1} + d_k - 1\}.$$

After  $\ell$  steps, we obtain  $2^\ell$  disjoint subsets  $I_\ell^i, i \in \{1, \dots, 2^\ell\}$  with  $\text{card}(I_\ell^i) = n_\ell$ . Then, the Cantor-like set is defined as

$$K_B = \bigcup_{i=1}^{2^\ell} I_\ell^i,$$

and for each level  $k \in \{0, \dots, \ell\}$  and each  $j \in \{1, \dots, 2^k\}$ , define

$$K_k^j = \bigcup_{i=(j-1)2^{\ell-k}+1}^{j2^{\ell-k}} I_\ell^i.$$

Some properties derived from this construction are given by Banna et al. [3]:

1.  $\delta \leq \frac{1}{2}$  and  $\ell \leq \frac{\log B}{\log 2}$ ;
2.  $d_j \geq \frac{B\delta(1-\delta)^j}{2^{j+1}}$  and  $n_\ell \leq \frac{B(1-\delta)^\ell}{2^{\ell-1}}$ ;
3. Each  $I_\ell^i, i \in \{1, \dots, 2^\ell\}$  contains  $n_\ell$  consecutive integers, and for any  $i \in \{1, \dots, 2^{\ell-1}\}$ ,  $I_\ell^{2i-1}$  and  $I_\ell^{2i}$  are spaced by  $d_{\ell-1}$  integers;
4.  $\text{card}(K_B) \geq \frac{B}{2}$ ;
5. For each  $k \in \{0, \dots, \ell\}$  and each  $j \in \{1, \dots, 2^k\}$ ,  $\text{card}(K_k^j) = 2^{\ell-k}n_\ell$ . For each  $j \in \{1, \dots, 2^{k-1}\}$ ,  $K_k^{2j-1}$  and  $K_k^{2j}$  are spaced by  $d_{k-1}$  integers;
6.  $K_0^1 = K_B$  and  $K_\ell^j = I_\ell^j$  for  $j \in \{1, \dots, 2^\ell\}$ .

#### A.4 A Decoupling Lemma for $\tau$ -Mixing Random Matrices

This section introduces the key tool to decouple  $\tau$ -mixing random matrices using Cantor-like set constructed in Sect. A.3. With some abuse of notation, within this section let's use  $\{\mathbf{X}_j\}_{j \in \{1, \dots, n\}}$  to denote a generic sequence of  $p \times p$  symmetric random matrices. Assume  $\mathbb{E}(\mathbf{X}_j) = \mathbf{0}$  and  $\|\mathbf{X}_j\| \leq M$  for some positive constant  $M$  and for all  $j \geq 1$ . For a collection of index sets  $H_1^k, k \in \{1, \dots, d\}$ , we assume that their cardinalities are equal and even. Denote  $\{\mathbf{X}_j\}_{j \in H_1^k}$  to be the set of matrices whose indices are in  $H_1^k$ . Assume  $\{\mathbf{X}_j\}_{j \in H_1^1}, \dots, \{\mathbf{X}_j\}_{j \in H_1^d}$  are mutually independent, while within each block  $H_1^k$  the matrices are possibly dependent. For each  $k$ , decompose  $H_1^k$  into two disjoint sets  $H_2^{2k-1}$  and  $H_2^{2k}$  with equal size, containing the first and second half of  $H_1^k$ , respectively. In addition, we denote  $\tau_0 := \tau\{\sigma(\{\mathbf{X}_j\}_{j \in H_2^{2k-1}}), \{\mathbf{X}_j\}_{j \in H_2^{2k}}; \|\cdot\|\}$  for some constant  $\tau_0 \geq 0$  and for all  $k \in \{1, \dots, d\}$ . For a given  $\epsilon > 0$ , we achieve the following decoupling lemma.

**Lemma A.3** We obtain for any  $\epsilon > 0$ ,

$$\begin{aligned} & \mathbb{E} \operatorname{Tr} \exp \left( t \sum_{k=1}^d \sum_{j \in H_1^k} \mathbf{X}_j \right) \\ & \leq \sum_{i=0}^d \binom{d}{i} (1 + L_1 + L_2)^{d-i} (L_1)^i \mathbb{E} \operatorname{Tr} \exp \left\{ (-1)^i t \left( \sum_{k=1}^{2d} \sum_{j \in H_2^k} \tilde{\mathbf{X}}_j \right) \right\}, \\ & \mathbb{E} \operatorname{Tr} \exp \left( -t \sum_{k=1}^d \sum_{j \in H_1^k} \mathbf{X}_j \right) \\ & \leq \sum_{i=0}^d \binom{d}{i} (1 + L_1 + L_2)^{d-i} (L_1)^i \mathbb{E} \operatorname{Tr} \exp \left\{ (-1)^{i+1} t \left( \sum_{k=1}^{2d} \sum_{j \in H_2^k} \tilde{\mathbf{X}}_j \right) \right\}, \end{aligned}$$

where

$$L_1 := pt\epsilon \exp(t\epsilon), \quad L_2 := \exp\{\operatorname{card}(H_1^1)tM\}\tau_0/\epsilon,$$

and  $\{\tilde{\mathbf{X}}_j\}_{j \in H_2^k}$ ,  $k \in \{1, \dots, 2d\}$ , are mutually independent and have the same distributions as  $\{\mathbf{X}_j\}_{j \in H_2^k}$ ,  $k \in \{1, \dots, 2d\}$ .

**Proof** We prove this lemma by induction. For any  $k \in \{1, \dots, d\}$ , we have  $H_1^k = H_2^{2k-1} \cup H_2^{2k}$  and hence  $\sum_{j \in H_1^k} \mathbf{X}_j = \sum_{j \in H_2^{2k-1}} \mathbf{X}_j + \sum_{j \in H_2^{2k}} \mathbf{X}_j$ .

By Lemma A.2, for each  $k \in \{1, \dots, d\}$ , we could find a sequence of random matrices  $\{\tilde{\mathbf{X}}_j\}_{j \in H_2^{2k}}$  and an independent uniformly distributed random variable  $U_k$  on  $[0, 1]$  such that

1.  $\{\tilde{\mathbf{X}}_j\}_{j \in H_2^{2k}}$  is measurable with respect to the sigma field  $\sigma(\{\mathbf{X}_j\}_{j \in H_2^{2k-1}}) \vee \sigma(\{\mathbf{X}_j\}_{j \in H_2^{2k}}) \vee \sigma(U_k)$ ;
2.  $\{\tilde{\mathbf{X}}_j\}_{j \in H_2^{2k}}$  is independent of  $\sigma(\{\mathbf{X}_j\}_{j \in H_2^{2k-1}})$ ;
3.  $\{\tilde{\mathbf{X}}_j\}_{j \in H_2^{2k}}$  has the same distribution as  $\{\mathbf{X}_j\}_{j \in H_2^{2k}}$ ;
4.  $\mathbb{P}(\|\sum_{j \in H_2^{2k}} \mathbf{X}_j - \sum_{j \in H_2^{2k}} \tilde{\mathbf{X}}_j\| > \epsilon_k) \leq \mathbb{E}(\|\sum_{j \in H_2^{2k}} \mathbf{X}_j - \sum_{j \in H_2^{2k}} \tilde{\mathbf{X}}_j\|)/\epsilon_k \leq \tau_0/\epsilon_k$  by Markov's inequality and the fact that  $\tau_0 = \sum_{j \in H_2^{2k}} \mathbb{E}(\|\mathbf{X}_j - \tilde{\mathbf{X}}_j\|)$ .

To make notation easier to follow, we set equal value to  $\epsilon_k$  for  $k \in \{1, \dots, d\}$  and denote it as  $\epsilon$ . Moreover, we denote the event  $\Gamma_k = \{\|\sum_{j \in H_2^{2k}} \tilde{\mathbf{X}}_j - \sum_{j \in H_2^{2k}} \mathbf{X}_j\| \leq \epsilon\}$  for  $k \in \{1, \dots, d\}$ .

For the base case,  $k = 1$ .

$$\mathbb{E} \operatorname{Tr} \exp \left( t \sum_{k=1}^d \sum_{j \in H_1^k} \mathbf{X}_j \right) = \underbrace{\mathbb{E} \left\{ \mathbb{1}_{\Gamma_1} \operatorname{Tr} \exp \left( t \sum_{k=1}^d \sum_{j \in H_1^k} \mathbf{X}_j \right) \right\}}_I + \underbrace{\mathbb{E} \left\{ \mathbb{1}_{(\Gamma_1)^c} \operatorname{Tr} \exp \left( t \sum_{k=1}^d \sum_{j \in H_1^k} \mathbf{X}_j \right) \right\}}_{II}.$$

Notice the definitions of terms  $I$  and  $II$  therein.

We have

$$\begin{aligned} I &= \mathbb{E} \left[ \mathbb{1}_{\Gamma_1} \operatorname{Tr} \exp \left\{ t \left( \sum_{j \in H_2^1} \mathbf{X}_j + \sum_{j \in H_2^2} \mathbf{X}_j + \sum_{k=2}^d \sum_{j \in H_1^k} \mathbf{X}_j \right) \right\} \right] \\ &\leq \mathbb{E} \operatorname{Tr} \exp \left\{ t \left( \sum_{j \in H_2^1} \mathbf{X}_j + \sum_{j \in H_2^2} \tilde{\mathbf{X}}_j + \sum_{k=2}^d \sum_{j \in H_1^k} \mathbf{X}_j \right) \right\} \\ &\quad + \mathbb{E} \left( \mathbb{1}_{\Gamma_1} \left[ \operatorname{Tr} \exp \left\{ t \left( \sum_{j \in H_2^1} \mathbf{X}_j + \sum_{j \in H_2^2} \mathbf{X}_j + \sum_{k=2}^d \sum_{j \in H_1^k} \mathbf{X}_j \right) \right\} \right. \right. \\ &\quad \left. \left. - \operatorname{Tr} \exp \left\{ t \left( \sum_{j \in H_2^1} \mathbf{X}_j + \sum_{j \in H_2^2} \tilde{\mathbf{X}}_j + \sum_{k=2}^d \sum_{j \in H_1^k} \mathbf{X}_j \right) \right\} \right] \right). \end{aligned}$$

By linearity of expectation and the facts that  $\operatorname{Tr}(\mathbf{X}) \leq p \|\mathbf{X}\|$  and  $\|\exp(\mathbf{X}) - \exp(\mathbf{Y})\| \leq \|\mathbf{X} - \mathbf{Y}\| \exp(\|\mathbf{X}\|) \exp(\|\mathbf{Y}\|)$ , we obtain

$$\begin{aligned} &\mathbb{E} \left( \mathbb{1}_{\Gamma_1} \left[ \operatorname{Tr} \exp \left\{ t \left( \sum_{j \in H_2^1} \mathbf{X}_j + \sum_{j \in H_2^2} \mathbf{X}_j + \sum_{k=2}^d \sum_{j \in H_1^k} \mathbf{X}_j \right) \right\} \right. \right. \\ &\quad \left. \left. - \operatorname{Tr} \exp \left\{ t \left( \sum_{j \in H_2^1} \mathbf{X}_j + \sum_{j \in H_2^2} \tilde{\mathbf{X}}_j + \sum_{k=2}^d \sum_{j \in H_1^k} \mathbf{X}_j \right) \right\} \right] \right) \\ &\leq \mathbb{E} \left[ \mathbb{1}_{\Gamma_1} p \left\| \exp \left\{ t \left( \sum_{j \in H_2^1} \mathbf{X}_j + \sum_{j \in H_2^2} \mathbf{X}_j + \sum_{k=2}^d \sum_{j \in H_1^k} \mathbf{X}_j \right) \right\} \right. \right. \\ &\quad \left. \left. - \exp \left\{ t \left( \sum_{j \in H_2^1} \mathbf{X}_j + \sum_{j \in H_2^2} \tilde{\mathbf{X}}_j + \sum_{k=2}^d \sum_{j \in H_1^k} \mathbf{X}_j \right) \right\} \right\| \right] \end{aligned}$$

$$\leq \mathbb{E} \left[ \mathbb{1}_{\Gamma_1} p \left\| t \sum_{j \in H_2^2} (\mathbf{X}_j - \tilde{\mathbf{X}}_j) \right\| \exp \left\{ \left\| t \sum_{j \in H_2^2} (\mathbf{X}_j - \tilde{\mathbf{X}}_j) \right\| \right\} \exp \left\{ \left\| t \left( \sum_{j \in H_2^1} \mathbf{X}_j + \sum_{j \in H_2^2} \tilde{\mathbf{X}}_j + \sum_{k=2}^d \sum_{j \in H_1^k} \mathbf{X}_j \right) \right\| \right\} \right].$$

By spectral mapping theorem, for a symmetric matrix  $\mathbf{X}$  with  $\|\mathbf{X}\| \leq M$ , we have  $\exp(\|\mathbf{X}\|) \leq \|\exp(\mathbf{X})\| \vee \|\exp(-\mathbf{X})\| \leq \|\exp(\mathbf{X})\| + \|\exp(-\mathbf{X})\|$ . Moreover, since  $\exp(\mathbf{X})$  is always positive definite for any matrix  $\mathbf{X}$  and  $\|\mathbf{X}\| \leq \text{Tr}(\mathbf{X})$  for any positive definite symmetric matrix  $\mathbf{X}$ , we obtain  $\|\exp(\mathbf{X})\| \leq \text{Tr} \exp(\mathbf{X})$  and  $\|\exp(-\mathbf{X})\| \leq \text{Tr} \exp(-\mathbf{X})$ . In addition, since we have  $\|\sum_{j \in H_2^2} (\mathbf{X}_j - \tilde{\mathbf{X}}_j)\| \leq \epsilon$  on  $\Gamma_1$ , we could further bound the inequality above by

$$\begin{aligned} & \mathbb{E} \left[ \mathbb{1}_{\Gamma_1} p t \epsilon \exp(t \epsilon) \left\| \exp \left\{ t \left( \sum_{j \in H_2^1} \mathbf{X}_j + \sum_{j \in H_2^2} \tilde{\mathbf{X}}_j + \sum_{k=2}^d \sum_{j \in H_1^k} \mathbf{X}_j \right) \right\} \right\| \right] \\ & \leq p t \epsilon \exp(t \epsilon) \left[ \mathbb{E} \text{Tr} \exp \left\{ t \left( \sum_{j \in H_2^1} \mathbf{X}_j + \sum_{j \in H_2^2} \tilde{\mathbf{X}}_j + \sum_{k=2}^d \sum_{j \in H_1^k} \mathbf{X}_j \right) \right\} \right. \\ & \quad \left. + \mathbb{E} \text{Tr} \exp \left\{ -t \left( \sum_{j \in H_2^1} \mathbf{X}_j + \sum_{j \in H_2^2} \tilde{\mathbf{X}}_j + \sum_{k=2}^d \sum_{j \in H_1^k} \mathbf{X}_j \right) \right\} \right]. \end{aligned}$$

Putting together, we reach

$$\begin{aligned} I & \leq \{1 + p t \epsilon \exp(t \epsilon)\} \mathbb{E} \text{Tr} \exp \left\{ t \left( \sum_{j \in H_2^1} \mathbf{X}_j + \sum_{j \in H_2^2} \tilde{\mathbf{X}}_j + \sum_{k=2}^d \sum_{j \in H_1^k} \mathbf{X}_j \right) \right\} \\ & \quad + p t \epsilon \exp(t \epsilon) \mathbb{E} \text{Tr} \exp \left\{ -t \left( \sum_{j \in H_2^1} \mathbf{X}_j + \sum_{j \in H_2^2} \tilde{\mathbf{X}}_j + \sum_{k=2}^d \sum_{j \in H_1^k} \mathbf{X}_j \right) \right\}. \quad (\text{A.1}) \end{aligned}$$

We then aim at *II*. For this, the proof largely follows the same argument as in [3]. Omitting the details, we obtain

$$II \leq \exp\{\text{card}(H_1^1) t M\} (\tau_0 / \epsilon) \mathbb{E} \text{Tr} \exp \left\{ t \left( \sum_{j \in H_2^1} \mathbf{X}_j + \sum_{j \in H_2^2} \tilde{\mathbf{X}}_j + \sum_{k=2}^d \sum_{j \in H_1^k} \mathbf{X}_j \right) \right\}. \quad (\text{A.2})$$

Denote  $L_1 := p t \epsilon \exp(t \epsilon)$  and  $L_2 := \exp\{\text{card}(H_1^1) t M\} \tau_0 / \epsilon$ . Combining (A.1) and (A.2) yields

$$\begin{aligned} & \mathbb{E} \text{Tr exp} \left( t \sum_{k=1}^d \sum_{j \in H_1^k} \mathbf{X}_j \right) \\ & \leq (1 + L_1 + L_2) \mathbb{E} \text{Tr exp} \left\{ t \left( \sum_{j \in H_2^1} \mathbf{X}_j + \sum_{j \in H_2^2} \tilde{\mathbf{X}}_j + \sum_{k=2}^d \sum_{j \in H_1^k} \mathbf{X}_j \right) \right\} \\ & \quad + L_1 \mathbb{E} \text{Tr exp} \left\{ -t \left( \sum_{j \in H_2^1} \mathbf{X}_j + \sum_{j \in H_2^2} \tilde{\mathbf{X}}_j + \sum_{k=2}^d \sum_{j \in H_1^k} \mathbf{X}_j \right) \right\} \\ & = \sum_{i=0}^1 \binom{1}{i} (1 + L_1 + L_2)^{1-i} (L_1)^i \mathbb{E} \text{Tr exp} \left\{ (-1)^i t \left( \sum_{j \in H_2^1} \mathbf{X}_j + \sum_{j \in H_2^2} \tilde{\mathbf{X}}_j + \sum_{k=2}^d \sum_{j \in H_1^k} \mathbf{X}_j \right) \right\}. \end{aligned}$$

This finishes the base case.

The induction steps are followed similarly and we omit the details. By iterating  $d$  times, we arrive at the following inequality:

$$\begin{aligned} & \mathbb{E} \text{Tr exp} \left( t \sum_{k=1}^d \sum_{j \in H_1^k} \mathbf{X}_j \right) \\ & \leq \sum_{i=0}^d \binom{d}{i} (1 + L_1 + L_2)^{d-i} (L_1)^i \mathbb{E} \text{Tr exp} \left\{ (-1)^i t \left( \sum_{k=1}^d \sum_{j \in H_2^{2k-1}} \mathbf{X}_j + \sum_{k=1}^d \sum_{j \in H_2^{2k}} \tilde{\mathbf{X}}_j \right) \right\}, \end{aligned} \quad (\text{A.3})$$

where  $\{\mathbf{X}_j\}_{j \in H_2^{2k-1}}$ ,  $k \in \{1, \dots, d\}$  and  $\{\tilde{\mathbf{X}}_j\}_{j \in H_2^{2k}}$ ,  $k \in \{1, \dots, d\}$  are mutually independent. In addition, they have the same distributions as  $\{\mathbf{X}_j\}_{j \in H_2^{2k-1}}$ ,  $k \in \{1, \dots, d\}$  and  $\{\mathbf{X}_j\}_{j \in H_2^{2k}}$ ,  $k \in \{1, \dots, d\}$ , respectively. For the sake of simplicity and clarity, we add an upper tilde to the matrices with indices in  $H_2^{2k-1}$ ,  $k \in \{1, \dots, d\}$ , i.e.,  $\{\tilde{\mathbf{X}}_j\}_{j \in H_2^{2k-1}}$  is identically distributed as  $\{\mathbf{X}_j\}_{j \in H_2^{2k-1}}$  for  $k \in \{1, \dots, d\}$ . Hence, (A.3) could be rewritten as

$$\mathbb{E} \text{Tr exp} \left( t \sum_{k=1}^d \sum_{j \in H_1^k} \mathbf{X}_j \right) \leq \sum_{i=0}^d \binom{d}{i} (1 + L_1 + L_2)^{d-i} (L_1)^i \mathbb{E} \text{Tr exp} \left\{ (-1)^i t \left( \sum_{k=1}^{2d} \sum_{j \in H_2^k} \tilde{\mathbf{X}}_j \right) \right\},$$

where  $\{\tilde{\mathbf{X}}_j\}_{j \in H_2^k}$ ,  $k \in \{1, \dots, 2d\}$  are mutually independent and their distributions are the same as  $\{\mathbf{X}_j\}_{j \in H_2^k}$ ,  $k \in \{1, \dots, 2d\}$ .

By changing  $\mathbf{X}$  to  $-\mathbf{X}$ , we immediately get the following bound:

$$\begin{aligned} & \mathbb{E} \operatorname{Tr} \exp \left( -t \sum_{k=1}^d \sum_{j \in H_1^k} \mathbf{X}_j \right) \\ & \leq \sum_{i=0}^d \binom{d}{i} (1 + L_1 + L_2)^{d-i} (L_1)^i \mathbb{E} \operatorname{Tr} \exp \left\{ (-1)^{i+1} t \left( \sum_{k=1}^{2d} \sum_{j \in H_2^k} \tilde{\mathbf{X}}_j \right) \right\}. \end{aligned}$$

This completes the proof of Lemma A.3.  $\square$

### A.5 Proof of Theorem 4.3

**Proof** Without loss of generality, let  $\psi_1 = \tilde{\psi}_1$ .

*Case I* First of all, we consider  $M = 1$ .

**Step I (Summation decomposition)** Let  $B_0 = n$  and  $\mathbf{U}_j^{(0)} = \mathbf{X}_j$  for  $j \in \{1, \dots, n\}$ . Let  $K_{B_0}$  be the Cantor-like set from  $\{1, \dots, B_0\}$  by construction of Sect. A.3,  $K_{B_0}^c = \{1, \dots, B_0\} \setminus K_{B_0}$ , and  $B_1 = \operatorname{card}(K_{B_0}^c)$ . Then, define

$$\mathbf{U}_j^{(1)} = \mathbf{X}_{i_j}, \text{ where } i_j \in K_{B_0}^c = \{i_1, \dots, i_{B_1}\}.$$

For each  $i \geq 1$ , let  $K_{B_i}$  be constructed from  $\{1, \dots, B_i\}$  by the same Cantor-like set construction. Denote  $K_{B_i}^c = \{1, \dots, B_i\} \setminus K_{B_i}$  and  $B_{i+1} = \operatorname{card}(K_{B_i}^c)$ . Then

$$\mathbf{U}_j^{(i+1)} = \mathbf{U}_{k_j}^{(i)}, \text{ where } k_j \in K_{B_i}^c = \{k_1, \dots, k_{B_{i+1}}\}.$$

We stop the process when there is a smallest  $L$  such that  $B_L \leq 2$ . Then, we have for  $i \leq L-1$ ,  $B_i \leq n2^{-i}$  because each Cantor-like set  $K_{B_{i+1}}$  has cardinality greater than  $B_i/2$ . Also notice that  $L \leq \lceil \log n / \log 2 \rceil$ .

For  $i \in \{0, \dots, L-1\}$ , denote

$$\mathbf{S}_i = \sum_{j \in K_{B_i}} \mathbf{U}_j^{(i)} \text{ and } \mathbf{S}_L = \sum_{j \in K_{B_{L-1}}^c} \mathbf{U}_j^{(L)}.$$

Then, we observe

$$\sum_{j=1}^n \mathbf{X}_j = \sum_{i=0}^L \mathbf{S}_i.$$

**Step II (Bounding Laplacian transform)** This step hinges on the following lemma, which provides an upper bound for the Laplace transform of sum of a sequence of random matrices which are  $\tau$ -mixing with geometric decay, i.e.,  $\tau(k) \leq \psi_1 \exp\{-\psi_2(k-1)\}$  for all  $k \geq 1$  for some constants  $\psi_1, \psi_2 > 0$ .  $\square$

**Lemma A.4** (Proof in Sect. A.6) *For a sequence of  $p \times p$  matrices  $\{\mathbf{X}_i\}$ ,  $i \in \{1, \dots, B\}$  satisfying conditions in Theorem 4.3 with  $M = 1$  and  $\psi_1 \geq p^{-1}$ , there exists a subset  $K_B \subseteq \{1, \dots, B\}$  such that for  $0 < t \leq \min\{1, \frac{\psi_2}{8 \log(\psi_1 B^6 p)}\}$ ,*

$$\begin{aligned} \log \mathbb{E} \operatorname{Tr} \exp \left( t \sum_{j \in K_B} \mathbf{X}_j \right) &\leq \log p + 4h(4)Bt^2v^2 \\ &+ 151 \left[ 1 + \exp \left\{ \frac{1}{\sqrt{p}} \exp \left( -\frac{\psi_2}{64t} \right) \right\} \right] \frac{t^2}{\psi_2} \exp \left( -\frac{\psi_2}{64t} \right). \end{aligned}$$

For each  $\mathbf{S}_i$ ,  $i \in \{0, \dots, L-1\}$ , by applying Lemma A.4 with  $B = B_i$ , we have for any positive  $t$  satisfying  $0 < t \leq \min\{1, \frac{\psi_2}{8 \log(\psi_1 (n2^{-i})^6 p)}\}$ ,

$$\log \mathbb{E} \operatorname{Tr} \exp(t\mathbf{S}_i) \leq \log p + t^2(C_1 2^{-i}n + C_{2,i}),$$

where  $C_1 := 4h(4)v^2$ ,  $C_{2,i} := 302 \cdot 2^{\frac{6i}{8}}/\psi_2 n^{\frac{6}{8}}$ .

Denote

$$\tilde{f}(\psi_1, \psi_2, i) := \min \left\{ 1, \frac{\psi_2}{8 \log\{\psi_1 (n2^{-i})^6 p\}} \right\}.$$

For any  $0 < t \leq \tilde{f}(\psi_1, \psi_2, i)$ , we obtain

$$\log \mathbb{E} \operatorname{Tr} \exp(t\mathbf{S}_i) \leq \log p + \frac{t^2(C_1 2^{-i}n + C_{2,i})}{1 - t/\tilde{f}(\psi_1, \psi_2, i)} \leq \log p + \frac{t^2\{C_1^{\frac{1}{2}}(2^{-i}n)^{\frac{1}{2}} + C_{2,i}^{\frac{1}{2}}\}^2}{1 - t/\tilde{f}(\psi_1, \psi_2, i)}.$$

For  $\mathbf{S}_L$ , since  $B_L \leq 2$ , for  $0 < t \leq 1$ ,

$$\log \mathbb{E} \operatorname{Tr} \exp(t\mathbf{S}_L) \leq \log p + t^2 h(2t) \lambda_{\max}\{\mathbb{E}(\mathbf{S}_L^2)\} \leq \log p + \frac{2t^2 v^2}{1-t}.$$

Denote  $\sigma_i := C_1^{\frac{1}{2}}(2^{-i}n)^{\frac{1}{2}} + C_{2,i}^{\frac{1}{2}}$ ,  $\sigma_L := \sqrt{2}v$ ,  $\kappa_i := 1/\tilde{f}(\psi_1, \psi_2, i)$ , and  $\kappa_L := 1$ .

Summing up, we have

$$\begin{aligned} \sum_{i=0}^L \sigma_i &= \sum_{i=0}^{L-1} \{C_1^{\frac{1}{2}}(2^{-i}n)^{\frac{1}{2}} + C_{2,i}^{\frac{1}{2}}\} + \sqrt{2}v \leq 15\sqrt{n}v + 60\sqrt{1/\psi_2}, \\ \sum_{i=0}^L \kappa_i &\leq \frac{\log n}{\log 2} \max \left\{ 1, \frac{8 \log(\psi_1 n^6 p)}{\psi_2} \right\} := \tilde{\psi}(\psi_1, \psi_2, n, p). \end{aligned}$$



Hence, by Lemma 3 in [22], for  $0 < t \leq \{\tilde{\psi}(\psi_1, \psi_2, n, p)\}^{-1}$ , we have

$$\log \mathbb{E} \operatorname{Tr} \exp \left( t \sum_{j=1}^n \mathbf{X}_j \right) \leq \log p + \frac{t^2 (15\sqrt{n}v + 60\sqrt{1/\psi_2})^2}{1 - t\tilde{\psi}(\psi_1, \psi_2, n, p)}.$$

**Step III (Matrix Chernoff bound)** Lastly by matrix Chernoff bound, we obtain

$$\mathbb{P} \left\{ \lambda_{\max} \left( \sum_{j=1}^n \mathbf{X}_j \right) \geq x \right\} \leq p \exp \left\{ - \frac{x^2}{8(15^2 n v^2 + 60^2 / \psi_2) + 2x \tilde{\psi}(\psi_1, \psi_2, n, p)} \right\}.$$

*Case II* We consider general  $M > 0$ . It is obvious that if  $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$  is a sequence of  $\tau$ -mixing random matrices such that  $\tau(k; \{\mathbf{X}_t\}_{t \in \mathbb{Z}}, \|\cdot\|) \leq M\psi_1 \exp\{-\psi_2(k-1)\}$ , then  $\{\mathbf{X}_i/M\}_{i \in \mathbb{Z}}$  is also a sequence of  $\tau$ -mixing random matrices such that  $\tau(k; \{\mathbf{X}_i/M\}_{i \in \mathbb{Z}}, \|\cdot\|) \leq \psi_1 \exp\{-\psi_2(k-1)\}$  and  $\|\mathbf{X}_i/M\| \leq 1$ . Then applying the result of Case I to  $\{\mathbf{X}_i/M\}_{i \in \mathbb{Z}}$ , we obtain

$$\mathbb{P} \left\{ \lambda_{\max} \left( \sum_{j=1}^n \mathbf{X}_j / M \right) \geq x \right\} \leq p \exp \left\{ - \frac{x^2}{8(15^2 n v_M^2 + 60^2 / \psi_2) + 2x \tilde{\psi}(\psi_1, \psi_2, n, p)} \right\},$$

where  $v_M^2 := \sup_{K \subseteq \{1, \dots, n\}} \frac{1}{\operatorname{card}(K)} \lambda_{\max} \left\{ \mathbb{E} \left( \sum_{i \in K} \mathbf{X}_i / M \right)^2 \right\} = v^2 / M^2$  for  $v^2$  defined in Theorem 4.3. Thus,

$$\mathbb{P} \left\{ \lambda_{\max} \left( \sum_{j=1}^n \mathbf{X}_j \right) \geq x \right\} \leq p \exp \left\{ - \frac{x^2}{8(15^2 n v^2 + 60^2 M^2 / \psi_2) + 2x M \tilde{\psi}(\psi_1, \psi_2, n, p)} \right\}.$$

This completes the proof of Theorem 4.3.  $\square$

## A.6 The Proof of Lemma A.4

**Proof** Let  $K_B$  be constructed as in Sect. A.3 for any arbitrary  $B \geq 2$  and  $M = 1$ .

*Case I* If  $0 < t \leq 4/B$ , by Lemma 4 in [3], we have

$$\mathbb{E} \operatorname{Tr} \exp \left( t \sum_{i \in K_B} \mathbf{X}_i \right) \leq p \exp \left[ t^2 h \left\{ t \lambda_{\max} \left( \sum_{i \in K_B} \mathbf{X}_i \right) \right\} \lambda_{\max} \left\{ \mathbb{E} \left( \sum_{i \in K_B} \mathbf{X}_i \right)^2 \right\} \right].$$

By Weyl's inequality,  $\lambda_{\max}(\sum_{i \in K_B} \mathbf{X}_i) \leq B$  since  $\operatorname{card}(K_B) \leq B$ , and by definition of  $v^2$  in Theorem 4.3, we have  $\lambda_{\max}\{\mathbb{E}(\sum_{i \in K_B} \mathbf{X}_i)^2\} \leq Bv^2$ . Therefore, we obtain  $h\{t \lambda_{\max}(\sum_{i \in K_B} \mathbf{X}_i)\} \leq h(tB) \leq h(4)$  and

$$\mathbb{E} \operatorname{Tr} \exp \left( t \sum_{i \in K_B} \mathbf{X}_i \right) \leq p \exp \{ t^2 h(4) B v^2 \}. \quad (\text{A.4})$$

*Case II* Now we consider the case where  $4/B < t \leq \min\{1, \frac{\psi_2}{8 \log(\psi_1 B^6 p)}\}$ .

**Step I** Let  $J$  be a chosen integer from  $\{0, \dots, \ell_B\}$  whose actual value will be determined later. We will use the same notation to denote Cantor-like sets as in Sect. A.3. By Lemma A.3 and similar induction argument as in [3], we obtain

$$\mathbb{E} \operatorname{Tr} \exp \left( t \sum_{j \in K_0^1} \mathbf{X}_j \right) \leq \sum_{i_1=0}^{2^0} \cdots \sum_{i_J=0}^{2^{J-1}} \left[ \left( \prod_{k=1}^J A_{k, i_k} \right) \mathbb{E} \operatorname{Tr} \exp \left\{ (-1)^{\sum_{k=1}^J i_k} t \left( \sum_{i'=1}^{2^J} \sum_{j \in K_{i'}^J} \tilde{\mathbf{X}}_j \right) \right\} \right], \quad (\text{A.5})$$

where  $\{\tilde{\mathbf{X}}_j\}_{j \in K_{i'}^J}$  for  $i' \in \{1, \dots, 2^J\}$  are mutually independent and have the same distributions as  $\{\mathbf{X}_j\}_{j \in K_{i'}^J}$  for  $i' \in \{1, \dots, 2^J\}$ , and

$$\begin{aligned} A_{k, i_k} &:= \binom{2^{k-1}}{i_k} (1 + L_{k,1} + L_{k,2})^{2^{k-1}-i_k} (L_{k,1})^{i_k}, \\ \epsilon_k &:= (2pt)^{-\frac{1}{2}} \{2^{\ell-k} n_\ell \exp(t2^{\ell-k+1} n_\ell) \tau_{d_{k-1}+1}\}^{\frac{1}{2}}, \\ L_{k,1} &:= (pt/2)^{\frac{1}{2}} \exp(t\epsilon_k) \{2^{\ell-k} n_\ell \exp(t2^{\ell-k+1} n_\ell) \tau_{d_{k-1}+1}\}^{\frac{1}{2}}, \\ L_{k,2} &:= (2pt)^{\frac{1}{2}} \exp(t\epsilon_k) \{2^{\ell-k} n_\ell \exp(t2^{\ell-k+1} n_\ell) \tau_{d_{k-1}+1}\}^{\frac{1}{2}}, \end{aligned}$$

**Step II** Now we choose  $J$  as follows:

$$J = \inf \left\{ k \in \{0, \dots, \ell\} : \frac{B(1-\delta)^k}{2^k} \leq \min \left\{ \frac{\psi_2}{8t^2}, B \right\} \right\}.$$

We first bound  $\mathbb{E} \operatorname{Tr} \exp \{ t (\sum_{i'=1}^{2^J} \sum_{j \in K_{i'}^J} \tilde{\mathbf{X}}_j) \}$  and  $\mathbb{E} \operatorname{Tr} \exp \{ -t (\sum_{i'=1}^{2^J} \sum_{j \in K_{i'}^J} \tilde{\mathbf{X}}_j) \}$ .

From (A.5), we obtain  $2^J$  sets of  $\{\tilde{\mathbf{X}}_j\}$  that are mutually independent. To make notation less cluttered, we will remove the upper tilde from  $\tilde{\mathbf{X}}_j$  for all  $j$ . Denote the number of matrices in each set  $K_{i'}^J$  to be  $q := 2^{\ell-J} n_\ell$ . For each set  $K_{i'}^J$ ,  $i \in \{1, \dots, 2^J\}$ , we divide it into consecutive sets with cardinality  $\tilde{q}$  and potentially a residual term if  $q$  is not divisible by  $\tilde{q}$ . More specifically, we have  $2\tilde{q} \leq q$  and  $m_{q,\tilde{q}} := \lceil q/2\tilde{q} \rceil$ . The value  $\tilde{q}$  will be determined later.

Then, each set  $K_{i'}^J$  contains  $2m_{q,\tilde{q}}$  numbers of sets with cardinality  $\tilde{q}$  and one set with cardinality less than  $2\tilde{q}$ . For each  $K_{i'}^J$ ,  $i \in \{1, \dots, 2^J\}$ , denote these consecutive sets described above by  $\mathcal{Q}_k^i$ ,  $k \in \{1, \dots, 2m_{q,\tilde{q}} + 1\}$ . Given these notation, we could rewrite the bound in the following:

$$\begin{aligned} & \mathbb{E} \operatorname{Tr} \exp \left( t \sum_{i=1}^{2^J} \sum_{j \in K_j^i} \mathbf{X}_j \right) \\ &= \mathbb{E} \operatorname{Tr} \exp \left( t \sum_{i=1}^{2^J} \sum_{k=1}^{2m_{q,\tilde{q}}+1} \sum_{j \in Q_k^i} \mathbf{X}_j \right) = \mathbb{E} \operatorname{Tr} \exp \left( t \sum_{i=1}^{2^J} \sum_{k=1}^{m_{q,\tilde{q}}} \sum_{j \in Q_{2k}^i} \mathbf{X}_j + t \sum_{i=1}^{2^J} \sum_{k=1}^{m_{q,\tilde{q}}+1} \sum_{j \in Q_{2k-1}^i} \mathbf{X}_j \right). \end{aligned}$$

Since  $\operatorname{Tr} \exp(\cdot)$  is convex (cf. Proposition 2 in [25]), by Jensen's inequality, we have

$$\begin{aligned} \mathbb{E} \operatorname{Tr} \exp \left( t \sum_{i=1}^{2^J} \sum_{j \in K_j^i} \mathbf{X}_j \right) &\leq \frac{1}{2} \mathbb{E} \operatorname{Tr} \exp \left( 2t \sum_{i=1}^{2^J} \sum_{k=1}^{m_{q,\tilde{q}}} \sum_{j \in Q_{2k}^i} \mathbf{X}_j \right) \\ &\quad + \frac{1}{2} \mathbb{E} \operatorname{Tr} \exp \left( 2t \sum_{i=1}^{2^J} \sum_{k=1}^{m_{q,\tilde{q}}+1} \sum_{j \in Q_{2k-1}^i} \mathbf{X}_j \right). \end{aligned}$$

Since the number of odd index sets is always equal to or one more than that of the even index sets, the upper bound of  $\frac{1}{2} \mathbb{E} \operatorname{Tr} \exp \left( 2t \sum_{i=1}^{2^J} \sum_{k=1}^{m_{q,\tilde{q}}} \sum_{j \in Q_{2k}^i} \mathbf{X}_j \right)$  will always be less than or equal to that of  $\frac{1}{2} \mathbb{E} \operatorname{Tr} \exp \left( 2t \sum_{i=1}^{2^J} \sum_{k=1}^{m_{q,\tilde{q}}+1} \sum_{j \in Q_{2k-1}^i} \mathbf{X}_j \right)$ . Hence, we only need to provide an upper bound for  $\mathbb{E} \operatorname{Tr} \exp \left( 2t \sum_{i=1}^{2^J} \sum_{k=1}^{m_{q,\tilde{q}}+1} \sum_{j \in Q_{2k-1}^i} \mathbf{X}_j \right)$ . Our goal is then to replace all  $\{\mathbf{X}_j\}_{j \in Q_{2k-1}^i}$  in the last inequality by mutually independent copies  $\{\tilde{\mathbf{X}}_j\}_{j \in Q_{2k-1}^i}$  with same distributions for  $k \in \{1, \dots, 2m_{q,\tilde{q}}+1\}$ ,  $i \in \{1, \dots, 2^J\}$ . Again we will proceed by induction. We first show

$$\begin{aligned} & \mathbb{E} \operatorname{Tr} \exp \left( 2t \sum_{i=1}^{2^J} \sum_{k=1}^{m_{q,\tilde{q}}+1} \sum_{j \in Q_{2k-1}^i} \mathbf{X}_j \right) \\ &\leq \sum_{i_1=0}^1 \tilde{A}_{i_1} \mathbb{E} \operatorname{Tr} \exp \left\{ (-1)^{i_1} 2t \left( \sum_{k=1}^{m_{q,\tilde{q}}+1} \sum_{j \in Q_{2k-1}^1} \tilde{\mathbf{X}}_j + \sum_{i=2}^{2^J} \sum_{k=1}^{m_{q,\tilde{q}}+1} \sum_{j \in Q_{2k-1}^i} \mathbf{X}_j \right) \right\}, \end{aligned}$$

where the constants  $\tilde{A}_{i_1}$  will be specified later. For each  $\{\mathbf{X}_j\}_{j \in Q_{2k-1}^1}$ ,  $k \in \{1, \dots, m_{q,\tilde{q}}+1\}$ , we could find a sequence of  $\{\tilde{\mathbf{X}}_j\}_{j \in Q_{2k-1}^1}$ ,  $k \in \{1, \dots, m_{q,\tilde{q}}+1\}$  that are mutually independent with each other. More specifically, let  $\{\tilde{\mathbf{X}}_j\}_{j \in Q_1^1} = \{\mathbf{X}_j\}_{j \in Q_1^1}$ . By applying Lemma A.2 on  $\{\tilde{\mathbf{X}}_j\}_{j \in Q_1^1}$  and  $\{\mathbf{X}_j\}_{j \in Q_3^1}$  with a chosen  $\tilde{\epsilon} > 0$ , we may find a sequence of random matrices  $\{\tilde{\mathbf{X}}_j\}_{j \in Q_3^1}$  such that for each  $j_0 \in Q_3^1$ , we have

1.  $\tilde{\mathbf{X}}_{j_0}$  is measurable with respect to  $\sigma(\{\tilde{\mathbf{X}}_j\}_{j \in Q_1^1}) \vee \sigma(\mathbf{X}_{j_0}) \vee \sigma(\tilde{U}_{j_0}^1)$ ;
2.  $\tilde{\mathbf{X}}_{j_0}$  is independent of  $\sigma(\{\tilde{\mathbf{X}}_j\}_{j \in Q_1^1})$ ;
3.  $\tilde{\mathbf{X}}_{j_0}$  has the same distribution as  $\mathbf{X}_{j_0}$ ;
4.  $\mathbb{P}(\|\tilde{\mathbf{X}}_{j_0} - \mathbf{X}_{j_0}\| \geq \tilde{\epsilon}) \leq \mathbb{E}(\|\tilde{\mathbf{X}}_{j_0} - \mathbf{X}_{j_0}\|)/\tilde{\epsilon} \leq \tau_{\tilde{q}+1}/\tilde{\epsilon}$  by Markov's inequality.

For each  $j_0 \in Q_3^1$ ,  $\tilde{U}_{j_0}^1$  is independent with  $\{\tilde{\mathbf{X}}_j\}_{j \in Q_1^1}$  and  $\mathbf{X}_{j_0}$ . In addition, since there are at least  $\tilde{q}$  number of matrices between  $\{\tilde{\mathbf{X}}_j\}_{j \in Q_1^1}$  and  $\mathbf{X}_{j_0}$  by our construction, we have  $\tau\{\sigma(\{\tilde{\mathbf{X}}_j\}_{j \in Q_1^1}), \mathbf{X}_{j_0}; \|\cdot\|\} \leq \tau_{\tilde{q}+1}$ . Note that  $\{\tilde{\mathbf{X}}_j\}_{j \in Q_3^1}$  is independent with  $\{\tilde{\mathbf{X}}_j\}_{j \in Q_1^1}$  but not mutually independent within the set  $Q_3^1$ .

Following the induction steps similar to the previous step and without redundancy, we obtain

$$\begin{aligned} & \mathbb{E} \operatorname{Tr} \exp \left( 2t \sum_{i=1}^{2^J} \sum_{k=1}^{m_{q,\tilde{q}}+1} \sum_{j \in Q_{2k-1}^i} \mathbf{X}_j \right) \\ & \leq \sum_{i_1=0}^1 \tilde{A}_{i_1} \mathbb{E} \operatorname{Tr} \exp \left\{ (-1)^{i_1} 2t \left( \sum_{k=1}^{m_{q,\tilde{q}}+1} \sum_{j \in Q_{2k-1}^1} \tilde{\mathbf{X}}_j + \sum_{i=2}^{2^J} \sum_{k=1}^{m_{q,\tilde{q}}+1} \sum_{j \in Q_{2k-1}^i} \mathbf{X}_j \right) \right\}, \end{aligned}$$

where

$$\begin{aligned} \tilde{\epsilon} &:= (4pt)^{-\frac{1}{2}} \{\exp(2tq)\tau_{\tilde{q}+1}\}^{\frac{1}{2}}, \\ \tilde{L}_1 &:= \frac{1}{2} (4pt)^{\frac{1}{2}} q \exp(2tq\tilde{\epsilon}) \{\exp(2tq)\tau_{\tilde{q}+1}\}^{\frac{1}{2}}, \\ \tilde{L}_2 &:= (4pt)^{\frac{1}{2}} q \{\exp(2tq)\tau_{\tilde{q}+1}\}^{\frac{1}{2}}, \\ \tilde{A}_{i_1} &:= \binom{1}{i_1} (1 + \tilde{L}_1 + \tilde{L}_2)^{1-i_1} (\tilde{L}_1)^{i_1}. \end{aligned}$$

This completes the base case.

Iterating the above calculation, we arrive at the following bound:

$$\begin{aligned} & \mathbb{E} \operatorname{Tr} \exp \left( 2t \sum_{i=1}^{2^J} \sum_{k=1}^{m_{q,\tilde{q}}+1} \sum_{j \in Q_{2k-1}^i} \mathbf{X}_j \right) \\ & \leq \sum_{i_1=0}^1 \cdots \sum_{i_{2^J}=0}^1 \left( \prod_{r=1}^{2^J} \tilde{A}_{i_r} \right) \mathbb{E} \operatorname{Tr} \exp \left\{ (-1)^{\sum_{r=1}^{2^J} i_r} 2t \sum_{i=1}^{2^J} \sum_{k=1}^{m_{q,\tilde{q}}+1} \sum_{j \in Q_{2k-1}^i} \tilde{\mathbf{X}}_j \right\}, \end{aligned} \quad (\text{A.6})$$

where  $\{\tilde{\mathbf{X}}_j\}_{j \in Q_{2k-1}^i}$  for  $(i, k) \in \{1, \dots, 2^J\} \times \{1, \dots, m_{q,\tilde{q}} + 1\}$  are mutually independent and identically distributed as  $\{\mathbf{X}_j\}_{j \in Q_{2k-1}^i}$  for  $(i, k) \in \{1, \dots, 2^J\} \times \{1, \dots, m_{q,\tilde{q}} + 1\}$ , and

$$\begin{aligned}\tilde{\epsilon} &:= (4pt)^{-\frac{1}{2}} \{\exp(2tq)\tau_{\tilde{q}+1}\}^{\frac{1}{2}}, \\ \tilde{L}_1 &:= \frac{1}{2}(4pt)^{\frac{1}{2}} q \exp(2tq\tilde{\epsilon}) \{\exp(2tq)\tau_{\tilde{q}+1}\}^{\frac{1}{2}}, \\ \tilde{L}_2 &:= (4pt)^{\frac{1}{2}} q \{\exp(2tq)\tau_{\tilde{q}+1}\}^{\frac{1}{2}}, \\ \tilde{A}_{i_r} &:= \binom{1}{i_r} (1 + \tilde{L}_1 + \tilde{L}_2)^{1-i_r} (\tilde{L}_1)^{i_r}.\end{aligned}$$

Let  $\tilde{q} := [2/t] \wedge [q/2]$ .  $\{\tilde{\mathbf{X}}_j\}_{j \in Q_{2k-1}^i}$  for  $(i, k) \in \{1, \dots, 2^J\} \times \{1, \dots, m_{q, \tilde{q}} + 1\}$  are mutually independent with mean  $\mathbf{0}$  and  $2^J \sum_{k=1}^{m_{q, \tilde{q}}+1} \text{card}(Q_{2k-1}^i) \leq B$ . Moreover, by Weyl's inequality, for  $(i, k) \in \{1, \dots, 2^J\} \times \{1, \dots, m_{q, \tilde{q}} + 1\}$ , we have

$$2\lambda_{\max} \left( \sum_{j \in Q_{2k-1}^i} \tilde{\mathbf{X}}_j \right) \leq 2\tilde{q} \leq \frac{4}{t}.$$

By Lemma 4 in [3], we obtain

$$\mathbb{E} \text{Tr} \exp \left( 2t \sum_{i=1}^{2^J} \sum_{k=1}^{m_{q, \tilde{q}}+1} \sum_{j \in Q_{2k-1}^i} \tilde{\mathbf{X}}_j \right) \leq p \exp\{4h(4)Bt^2v^2\}, \quad (\text{A.7})$$

$$\mathbb{E} \text{Tr} \exp \left( -2t \sum_{i=1}^{2^J} \sum_{k=1}^{m_{q, \tilde{q}}+1} \sum_{j \in Q_{2k-1}^i} \tilde{\mathbf{X}}_j \right) \leq p \exp\{4h(4)Bt^2v^2\}. \quad (\text{A.8})$$

Plugging (A.7) and (A.8) into (A.6) and using the fact that  $\sum_{i_r=0}^1 \tilde{A}_{i_r} = 1 + 2\tilde{L}_1 + \tilde{L}_2$ , we obtain

$$\mathbb{E} \text{Tr} \exp \left( 2t \sum_{i=1}^{2^J} \sum_{k=1}^{m_{q, \tilde{q}}+1} \sum_{j \in Q_{2k-1}^i} \mathbf{X}_j \right) \leq (1 + 2\tilde{L}_1 + \tilde{L}_2)^{2^J} p \exp\{4h(4)Bt^2v^2\}. \quad (\text{A.9})$$

By replacing  $\mathbf{X}$  by  $-\mathbf{X}$ , we obtain

$$\mathbb{E} \text{Tr} \exp \left( -2t \sum_{i=1}^{2^J} \sum_{k=1}^{m_{q, \tilde{q}}+1} \sum_{j \in Q_{2k-1}^i} \mathbf{X}_j \right) \leq (1 + 2\tilde{L}_1 + \tilde{L}_2)^{2^J} p \exp\{4h(4)Bt^2v^2\}. \quad (\text{A.10})$$

Combining (A.5) with (A.9) and (A.10), we get

$$\begin{aligned} \mathbb{E} \operatorname{Tr} \exp \left( t \sum_{j \in K_B} \mathbf{X}_j \right) &\leq \sum_{i_1=0}^{2^0} \cdots \sum_{i_J=0}^{2^{J-1}} \left[ \left( \prod_{k=1}^J A_{k,i_k} \right) (1 + 2\tilde{L}_1 + \tilde{L}_2)^{2^J} p \exp\{4h(4)Bt^2v^2\} \right] \\ &= \left\{ \prod_{k=1}^J (1 + 2L_{k,1} + L_{k,2})^{2^{k-1}} \right\} (1 + 2\tilde{L}_1 + \tilde{L}_2)^{2^J} p \exp\{4h(4)Bt^2v^2\}, \end{aligned} \quad (\text{A.11})$$

where the last equality is followed by  $\sum_{i_k=1}^{2^{k-1}} A_{k,i_k} = (1 + 2L_{k,1} + L_{k,2})^{2^{k-1}}$ .

By using  $\log(1+x) \leq x$  for  $x \geq 0$ , we have

$$\log \mathbb{E} \operatorname{Tr} \exp \left( t \sum_{j \in K_B} \mathbf{X}_j \right) \leq \sum_{k=1}^J 2^{k-1} (2L_{k,1} + L_{k,2}) + 2^J (2\tilde{L}_1 + \tilde{L}_2) + \log[p \exp\{4h(4)Bt^2v^2\}]. \quad (\text{A.12})$$

For simplicity, we denote  $I = \sum_{k=1}^J 2^{k-1} (2L_{k,1} + L_{k,2})$ ,  $II = 2^J (2\tilde{L}_1 + \tilde{L}_2)$  in (A.12).

**Step III** Following calculations similar to [3], we obtain

$$I \leq \frac{32\sqrt{2}}{\log 2} \left[ 1 + \exp \left\{ \frac{1}{\sqrt{2}p} \exp \left( -\frac{\psi_2}{16t} \right) \right\} \right] \frac{t^2}{\psi_2} \exp \left( -\frac{\psi_2}{32t} \right). \quad (\text{A.13})$$

and

$$II \leq 128 \left[ 1 + \exp \left\{ \frac{1}{\sqrt{p}} \exp \left( -\frac{\psi_2}{32t} \right) \right\} \right] \frac{t^2}{\psi_2} \exp \left( -\frac{\psi_2}{64t} \right). \quad (\text{A.14})$$

Hence, by combining (A.4), (A.12), (A.13) and (A.14), we obtain for  $0 < t \leq \min\{1, \frac{\psi_2}{8 \log(\psi_1 B^6 p)}\}$ ,

$$\begin{aligned} &\log \mathbb{E} \operatorname{Tr} \exp \left( t \sum_{j \in K_B} \mathbf{X}_j \right) \\ &\leq \log p + 4h(4)Bt^2v^2 + 151 \left[ 1 + \exp \left\{ \frac{1}{\sqrt{p}} \exp \left( -\frac{\psi_2}{64t} \right) \right\} \right] \frac{t^2}{\psi_2} \exp \left( -\frac{\psi_2}{64t} \right). \end{aligned}$$

This completes the proof of Lemma A.4.  $\square$

## References

1. Andrews, D.W.: Non-strong mixing autoregressive processes. *J. Appl. Probab.* **21**(4), 930–934 (1984)
2. Bai, Z., Yin, Y.: Limit of the smallest eigenvalue of a large dimensional sample covariance matrix. *Ann. Probab.* **21**(3), 1275–1294 (1993)
3. Banna, M., Merlevède, F., Youssef, P.: Bernstein-type inequality for a class of dependent random matrices. *Random Matrices Theory Appl.* **5**(2), 1650006 (2016)
4. Berbee, H.C.: Random Walks with Stationary Increments and Renewal Theory, vol. 112. Mathematisch Centrum, Amsterdam (1979)
5. Blinn, J.: Consider the lowly  $2 \times 2$  matrix. *IEEE Comput. Graph. Appl.* **16**(2), 82–88 (1996)
6. Brand, M.: Fast low-rank modifications of the thin singular value decomposition. *Linear Algebra Appl.* **415**(1), 20–30 (2006)
7. Brillinger, D.R.: Time Series: Data Analysis and Theory. Siam, Philadelphia (2001)
8. Bunea, F., Xiao, L.: On the sample covariance matrix estimator of reduced effective rank population matrices, with applications to fPCA. *Bernoulli* **21**(2), 1200–1230 (2015)
9. Chang, J., Guo, B., Yao, Q.: Principal component analysis for second-order stationary vector time series. *Ann. Stat.* **46**(5), 2094–2124 (2018)
10. Chen, X., Xu, M., Wu, W.B.: Covariance and precision matrix estimation for high-dimensional time series. *Ann. Stat.* **41**(6), 2994–3021 (2013)
11. Davis, C., Kahan, W.M.: The rotation of eigenvectors by a perturbation. iii. *SIAM J. Numer. Anal.* **7**(1), 1–46 (1970)
12. Dedecker, J., Doukhan, P., Lang, G., Leon, J., Louhichi, S., Prieur, C.: Weak Dependence: With Examples and Applications. Springer, New York (2007)
13. Dedecker, J., Prieur, C.: Coupling for  $\tau$ -dependent sequences and applications. *J. Theor. Probab.* **17**(4), 861–885 (2004)
14. Han, F., Liu, H.: ECA: high-dimensional elliptical component analysis in non-gaussian distributions. *J. Am. Stat. Assoc.* **113**(521), 252–268 (2018)
15. Koltchinskii, V., Lounici, K.: Concentration inequalities and moment bounds for sample covariance operators. *Bernoulli* **23**(1), 110–133 (2017a)
16. Koltchinskii, V., Lounici, K.: New asymptotic results in principal component analysis. *Sankhya A* **79**(2), 254–297 (2017b)
17. Koltchinskii, V., Lounici, K.: Normal approximation and concentration of spectral projectors of sample covariance. *Ann. Stat.* **45**(1), 121–157 (2017c)
18. Liu, W., Xiao, H., Wu, W.B.: Probability and moment inequalities under dependence. *Stat. Sin.* **23**(3), 1257–1272 (2013)
19. Lounici, K.: High-dimensional covariance matrix estimation with missing observations. *Bernoulli* **20**(3), 1029–1058 (2014)
20. Mendelson, S.: Empirical processes with a bounded  $\psi_1$  diameter. *Geom. Funct. Anal.* **20**(4), 988–1027 (2010)
21. Mendelson, S., Paouris, G.: On the singular values of random matrices. *J. Eur. Math. Soc.* **16**, 823–834 (2014)
22. Merlevède, F., Peligrad, M., Rio, E.: Bernstein inequality and moderate deviations under strong mixing conditions. *High Dimensional Probability V: The Luminy Volume*, pp. 273–292. Institute of Mathematical Statistics, Beachwood (2009)
23. Merlevède, F., Peligrad, M., Rio, E.: A Bernstein type inequality and moderate deviations for weakly dependent sequences. *Probab. Theory Relat. Fields* **151**(3), 435–474 (2011)
24. Oliveira, R.: Sums of random Hermitian matrices and an inequality by Rudelson. *Electron. Commun. Probab.* **15**, 203–212 (2010)
25. Petz, D.: A survey of certain trace inequalities. *Banach Center Publ.* **30**(1), 287–298 (1994)
26. Rudelson, M.: Random vectors in the isotropic position. *J. Funct. Anal.* **164**(1), 60–72 (1999)
27. Slepian, D.: The one-sided barrier problem for Gaussian noise. *Bell Syst. Techn. J.* **41**(2), 463–501 (1962)
28. Srivastava, N., Vershynin, R.: Covariance estimation for distributions with  $2 + \epsilon$  moments. *Ann. Probab.* **41**(5), 3081–3111 (2013)
29. Talagrand, M.: Upper and Lower Bounds for Stochastic Processes: Modern Methods and Classical Problems. Springer, Berlin (2014)

30. Tikhomirov, K.: Sample covariance matrices of heavy-tailed distributions. *Int. Math. Res. Not.* **2018**(20), 6254–6289 (2017)
31. Tropp, J.A.: An introduction to matrix concentration inequalities. *Found. Trends Mach. Learn.* **8**(1–2), 1–230 (2015)
32. van Handel, R.: Structured random matrices. *Convexity and Concentration*, vol. 161, pp. 107–156. Springer, Berlin (2017)
33. Vershynin, R.: Introduction to the non-asymptotic analysis of random matrices. *Compressed Sensing*, pp. 210–268. Cambridge University Press, Cambridge (2012)
34. Wu, W.B.: Nonlinear system theory: another look at dependence. *Proc. Natl. Acad. Sci. U. S. A.* **102**(40), 14150–14154 (2005)
35. Wu, W.B., Wu, Y.N.: Performance bounds for parameter estimates of high-dimensional linear models with correlated errors. *Electron. J. Stat.* **10**(1), 352–379 (2016)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.