Empirical eigen expansions and uniform bounds

Moritz Jirak

Institut für Mathematik, Unter den Linden 6 D-10099 Berlin,
e-mail: jirak@math.hu-berlin.de

Abstract: Let \( \{X_k\}_{k \in \mathbb{Z}} \in L^2(T) \) be a stationary process with associated covariance operator \( C \). Uniform asymptotic expansions of the corresponding empirical eigenvalues and eigenfunctions are established under optimal dependence assumptions, including both short and long memory processes. In addition, the underlying conditions on the covariance operator (spectral gap) are almost optimal. This allows us to study the relative maximum deviation of the empirical eigenvalues under very general conditions. Among other things, we show convergence to an extreme value distribution, giving rise to the construction of simultaneous confidence sets. Uniform rates of convergence for the relative trace and inverse trace of \( C \) are also obtained.

1. Introduction

Principal component analysis (PCA) has emerged as one of the most important tools in multivariate and highdimensional data analysis. In the latter, Functional principal component analysis (FPCA) is becoming more and more important. A comprehensive overview and some leading examples can be found in [21], [19], [33]. Given a functional time series \( X = \{X_k\}_{k \in \mathbb{Z}} \), it is typically assumed that \( X \) lies in the Hilbert space \( L^2(T) \), where \( T \subset \mathbb{R}^d \) is compact. The fundamental tool in the area of PCA and FPCA - both in theory and practice - is the usage of (functional) principal components (FPC). To fix ideas, let us introduce some notation. If \( \mathbb{E}[\|X_k\|_{L^2}] < \infty \), then the mean \( \mu = \mathbb{E}[X_k] \) and the covariance operator

\[
C(x) = \mathbb{E}[(X_k - \mu, x)(X_k - \mu)], \quad x \in L^2(T)
\]

exist. Here \( \langle \cdot, \cdot \rangle \) denotes the inner product in \( L^2 \), and \( \| \cdot \|_{L^2} \) the corresponding norm. The eigenfunctions of \( C \) are called the functional principal components and denoted by \( e = \{e_j\}_{j \in \mathbb{N}} \), i.e; we have \( C(e_j) = \lambda_j e_j \), where \( \lambda = \{\lambda_j\}_{j \in \mathbb{N}} \) denotes the eigenvalues. The eigenfunctions \( e \) are usually estimated by the empirical eigenfunctions \( \hat{e} = \{\hat{e}_j\}_{j \in \mathbb{N}} \), defined as the eigenfunctions of the empirical covariance operator

\[
\hat{C}_n(x) = \frac{1}{n} \sum_{k=1}^{n} \langle X_k - \bar{X}_n, x \rangle (X_k - \bar{X}_n), \quad x \in L^2(T),
\]
where $\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$. Hence $\hat{C}_n(\hat{e}_j) = \hat{\lambda}_j \hat{e}_j$, where $\hat{\lambda} = \{\hat{\lambda}_j\}_{j \in \mathbb{R}}$ denotes the empirical eigenvalues. Due to the fundamental importance of eigenfunctions and eigenvalues for FPCA and PCA, corresponding results on the asymptotic behaviour of empirical eigenfunctions and values are of high interest. [1] was among the first to give such results, (see also [11]), and established a CLT for $\hat{\lambda}_j$ (resp. $\hat{e}_j$) if $j$ is fixed. Fueled from high-dimensional applications, uniform bounds where $j$ increases with the sample size $n$ have become very important, leading to a significant rise in complexity of the problem. Well-known pathwise bounds are provided in the Lemma given below (cf. [4], [5]).

**Lemma 1.1.** Let $C_n = E[\hat{C}_n]$. If $X \in L^2(\mathcal{T})$ then

$$\left| \hat{\lambda}_j - \lambda_j \right| \leq \left\| \hat{C}_n - C_n \right\|_{L^2}, \quad \left\| \hat{e}_j - e_j \right\|_{L^2} \leq \frac{2\sqrt{2}}{\psi_j} \left\| \hat{C}_n - C_n \right\|_{L^2},$$

where $\psi_j = \min\{\lambda_{j-1} - \lambda_j, \lambda_j - \lambda_{j+1}\}$ (with $\psi_1 = \lambda_1 - \lambda_2$) and $\left\| \cdot \right\|_{L^2}$ denotes the operator norm.

The attractiveness of the above bounds lies in their simplicity, but unfortunately they are far from optimal from a probabilistic perspective. Indeed, the results of [11] tell us that in case of $\hat{\lambda}_j - \lambda_j$ the correct bound should include the additional factor $\lambda_j$, i.e; $\lambda_j \left\| \hat{C}_n - C_n \right\|_{L^2}$. A similar claim can be made for $\left\| \hat{e}_j - e_j \right\|_{L^2}$. In this spirit, based on Lemma 1.1, asymptotic expansions for $\hat{\lambda}_j - \lambda_j$ and $\hat{e}_j - e_j$ which allow for increasing $j$ have been established in [15] [16] [14] (see also [5], [8], [26]). These results have proved to be an indispensable tool in the literature, see for instance [7], [8], [14], [27], [21], [19] to name a few. But the corresponding (asymptotic) analysis is often based on heavy structural assumptions regarding $X$ and the spacings (spectral gap) $\Psi = \{\psi_j\}_{j \in \mathbb{N}}$ of the eigenvalues, severely limiting the range of application. In particular, a common key assumption is that $X$ is an IID sequence, which is rather restrictive, see [18], [19], [29]. In this paper, we dispose of these limitations and derive exact asymptotic expansions of $\hat{\lambda}_j$, $\hat{e}_j$ under optimal dependence assumptions, allowing both for long memory (strong dependence) and short memory (weak dependence). In addition, we only require a 'natural condition' concerning the spectral gap $\Psi$. It turns out that this condition is nearly optimal. One of the key ideas is to circumvent Lemma 1.1, replacing it with a backward induction that makes use of underlying recursions and contraction properties.

As application, we study the relative maximum deviation of the empirical eigenvalues, namely

$$T_{\lambda}^\Lambda = \sqrt{n} \max_{1 \leq j \leq J_n^+} \frac{\left| \hat{\lambda}_j - \lambda_j \right|}{\sigma_{\lambda,j} \hat{\lambda}_j},$$

where $J_n^+ \to \infty$, see Assumption 2.1 for a precise definition of $J_n^+$. Under mild
assumptions, we show that
\[ a_n \left( T_{J_n}^{\lambda_j} - b_n \right) \overset{d}{\to} V, \quad (1.3) \]
where \( V \) is a distribution of Gumbel type. The latter is based on a high dimensional Gaussian approximation, which is of independent interest, see Theorem 5.2. Result (1.3) is particularly important for the construction of simultaneous confidence sets and tests for the relevant number of FPCs to be used for statistical inference or modelling. Two other popular statistical measures in this context are the partial trace and inverse partial trace
\[ \text{tr}_J(C) = \sum_{j=1}^{J} \lambda_j \quad \text{and} \quad \text{tr}_J(C^{-1}) = \sum_{j=1}^{J} \lambda_j^{-1}. \quad (1.4) \]
Quantity \( \text{tr}_J(C)/\mathbb{E}[\|X_0 - \mu\|_2^2] \) is probably the most widely used measure among practitioners for determining the relevant number of FPCs. In a similar context, \( \text{tr}_J(C^{-1}) \) arises in ill-posed problems (cf. [15], [27], [3] and the references therein). In Theorem 2.11 we show that under sharp weak dependence conditions
\[
\max_{1 \leq j \leq J_n} \left| \text{tr}_J(\hat{C}) - \text{tr}_J(C) \right| = O_p(n^{-1/2}) \quad \text{and} \quad \\
\max_{1 \leq j \leq J_n} \left| \text{tr}_J(\hat{C}^{-1}) - \text{tr}_J(C^{-1}) \right| = O_p(n^{-1/2}) \text{tr}_{J_n}(C^{-1}). \quad (1.5)
\]
An outline of the paper can be given as follows. In Section 2 the key expansions of \( \lambda_j \) and \( \hat{\lambda}_j \) are given, alongside some additional results. In particular, we discuss in detail the optimality of the underlying assumptions. Section 2.4 introduces a general notion of weak dependence. We then discuss results (1.3) and (1.5). The proofs of the Eigen expansions are given in Section 3. In Section 5.1, a general high dimensional Gaussian approximation under dependence is established. Based on this result, we prove (1.3) and (1.5) in Section 5.2.

2. Preliminary notation and asymptotic expansions

For \( p \geq 1 \), denote with \( \| \cdot \|_p \) the \( L^p \)-norm \( \mathbb{E}[\cdot|\cdot]^{1/p} \). In the sequel, most of our results will depend on the normalized scores \( \{ \eta_{k,j} \}_{k \in \mathbb{Z}, j \in \mathbb{N}} \) and their empirical covariance \( n^{-1} \eta_k \), which we define as
\[ \eta_{k,j} = \frac{(X_k - \mu, e_j)}{\lambda_j^{1/2}} \quad \text{and} \quad \eta_{k,j} = \sum_{k=1}^{n} (\eta_{k,i} \eta_{k,j} - \mathbb{E}[\eta_{k,i} \eta_{k,j}]). \quad (2.1) \]
In the sequel, we write \( \lesssim, \gtrsim, \sim \) to denote (two-sided) inequalities involving a multiplicative constant. Given a set \( \mathcal{A} \), we denote with \( \mathcal{A}^c \) its complement. We now impose the following conditions on the scores \( \{ \eta_{k,j} \}_{k \in \mathbb{Z}, j \in \mathbb{N}} \) and eigenvalues \( \{ \lambda_j \}_{j \in \mathbb{N}} \).
Assumption 2.1. The sequence \( \{ X_k \}_{k \in \mathbb{N}} \) is stationary, such that for some \( a > 0, m, p \geq 1 \) and \( J_n^+ \in \mathbb{N} \) it holds that

\[
\mathbb{E}[X_k] = 0 \quad \text{and} \quad \max_{i,j \in \mathbb{N}} \| n^{-1/2} \eta_{i,j} \|_q < \infty \quad \text{for} \quad q = p2^{p+4}, \quad p = \lceil m/a \rceil,
\]

(A1) \( \max_{1 \leq i \leq J_n^+} \left\{ n^{-1/2 + a} \sum_{i \neq j} \frac{\lambda_i}{\lambda_j - \lambda_i}, n^{-1/2 + a} \sum_{i \neq j} \frac{\lambda_j}{\lambda_i - \lambda_j} \right\} < \infty \) and \( \lambda_{J_n^+} \geq n^{-m} \).

Let us discuss these assumptions and compare them to the literature. As a general preliminary remark, we note that all of our results have analogues in a general Hilbert space setting \( \mathbb{H} \). Working in \( L^2(T) \) is notationally less burdensome though, and the proofs are simpler. In particular, the Fubini-Tonelli Theorem allows to interchange the order of inner products and expectations.

Dependence assumptions: Assumption (A1) implicitly imposes a dependence assumption on the scores \( \eta_{k,j} \). In contrast to the literature (cf. [10] [15] [16] [26]), we do not require the typical independence assumption. In fact, (A1) is much more general. In Proposition 2.6 we show that (A1) holds under general, sharp weak dependence conditions. This means that if these conditions fail, we no longer have weak dependence. However, much more is valid. Suppose that \( \eta_{k,j} = \sum_{i=0}^{\infty} \alpha_{i,j} c_{k-i,j} \) where \( \{ \epsilon_{i,j} \}_{i,j \in \mathbb{N}} \) is standard Gaussian and IID and \( \alpha_{i,j} \sim i^{-\alpha} \), \( \alpha > 1/2 \). Then we show in Section 2.2 that

\[
\| \hat{C}_n - C_n \|_{L^2} \lesssim n^{-1/2} \quad \text{is equivalent with} \quad \text{'(A1)' holds for any fixed} \quad p \geq 1',
\]

(2.2)

where \( \| \hat{C}_n - C_n \|_{L^2} \) denotes the Hilbert-Schmidt-norm. Hence the rate \( n^{-1/2} \) carries over and (A1) poses no restriction, as long as we consider the CLT-domain (normalization with \( n^{-1/2} \)). In this sense, condition (A1) is optimal (in the CLT-Domain). Interestingly, this also allows for long memory sequences, and we even obtain a CLT for \( \hat{\lambda}_j \) and \( \hat{\epsilon}_j \) under long memory conditions, i.e; \( \sum_{i=1}^{\infty} \alpha_{i,j} = \infty \), see Theorem 2.5. Note that it is shown in [28] that \( \sum_{i=1}^{\infty} | \alpha_i | < \infty \) is necessary for the validity of a CLT for \( \sum_{k=1}^{n} X_{k} \) in an infinite dimensional Hilbert space (a different normalization doesn’t help here, which is different from the univariate case, see [28] for details). Finally, we remark that our method of proof can also be used to derive corresponding results in the non-central domain, i.e; \( \| \hat{C}_n - C_n \|_{L^2} \sim b_n \) with \( \sqrt{n} = o(b_n) \). To keep this exposition at reasonable length, this is not pursued here.

Structural conditions for eigenvalues: (A2) is the key condition regarding the structure of the eigenvalues \( \lambda_j \). The literature (cf. [8] [15] [16] [14] [10]) usually requires polynomial, exponential or convex structures regarding the decay-rate of the eigenvalues and particularly the spacing \( \psi_j \). For instance, a common minimum assumption is that \( \psi_j \gtrsim \lambda_{j-j}^{-1} \), which reflects a polynomial behavior of the eigenvalues \( \lambda_j \). As will be discussed below Theorem 2.3, (A2) turns out to be much weaker, in fact, we shall see that it is nearly optimal. To get a feeling of
the implications of (A2), let us consider the case where \( \lambda_j \) satisfies a convexity condition, i.e;

\[
\text{the function } \lambda(x) : x \mapsto \lambda_x \text{ is convex.}
\]

If (2.3) holds, then one may verify (cf. Lemma 4.1) that

\[
\sum_{i=1 \atop i \neq j}^{\infty} \frac{\lambda_i}{|\lambda_j - \lambda_i|} \lesssim j \log j \quad \text{and} \quad \sum_{i=1 \atop i \neq j}^{\infty} \frac{\lambda_i \lambda_j}{(\lambda_j - \lambda_i)^2} \lesssim j^2,
\]

hence (A2) is valid if \( J_n^+ \lesssim n^{1/2-a}(\log n)^{-1} \). Note that these bounds are not directly influenced by the decay of \( \lambda \) or \( \Psi \). The convexity condition (2.3) itself is mild and includes many cases encountered in the literature (cf. [10]), in particular polynomial or exponential cases

\[
\lambda_j \sim j^{-\rho}, \quad 0 < \rho < 1, \quad |c| < \infty \quad \text{or} \quad \lambda_j \sim j^{-\epsilon}, \quad \epsilon > 1. \quad (\text{EP})
\]

Also note that (A2) implies that the first \( J_n^+ \) eigenvalues are distinct. See [11] for a flavour of results which allow for Eigenspaces with rank greater than one.

Moment assumptions: The existence of all moments (often with additional Gaussian like growth conditions) is usually required in the literature (cf. [10] [15] [16], [26]) in the context of expansions for \( \hat{\lambda}_j, \hat{e}_j \). In contrast, we only require a finite number of moments, which, however, may be large. On the other hand, all of our results will be expressed in terms of the \( \| \cdot \|_p \)-norm, and moving over to the weaker \( O_P(\cdot) \) formulation, the moment assumptions can be lowered.

For stating our results, we introduce the quantity

\[
I_{k,j} = \langle (\hat{C}_n - C_n)(e_j), e_k \rangle, \quad k, j \in \mathbb{N},
\]

which is one of the main contributing parts in the expansions given below. We first give the main results, followed by a discussion and comparison to the literature. For the empirical eigenvalues \( \hat{\lambda}_j \), we have the following.

**Theorem 2.2.** Assume that Assumption 2.1 holds. Then for \( 1 \leq J \leq J_n^+ \)

\[
\left\| \max_{1 \leq j \leq J} \frac{1}{\lambda_j} \left( \hat{\lambda}_j - \lambda_j - I_{j,j} \right) \right\|_p \lesssim \frac{J^{1/p}}{\sqrt{n}} n^{-a}.
\]

The above result provides an exact uniform first-order expansion for \( \hat{\lambda}_j \). For fixed \( j, 1 \leq j \leq J_n^+ \), the factor \( J^{1/p} \) in the bound on the RHS can be dropped.

Next, we state the companion result for the empirical eigenfunctions \( \hat{e}_j \).

**Theorem 2.3.** Assume that Assumption 2.1 holds. Then for \( 1 \leq J \leq J_n^+ \)

\[
\left\| \max_{1 \leq j \leq J} \left\| \frac{1}{\sqrt{\Lambda_j}} \left( e_j - \hat{e}_j + \frac{e_j}{2} \| e_j - \hat{e}_j \|_{L^2}^2 - \sum_{k=1 \atop k \neq j}^{\infty} e_k \frac{I_{k,j}}{\lambda_j - \lambda_k} \right) \right\|_{L^2} \right\|_p \lesssim \frac{J^{1/p}}{\sqrt{n}} n^{-a},
\]
where $\Lambda_j = \sum_{k \neq j}^{\infty} \frac{\lambda_j \lambda_k}{(\lambda_j - \lambda_k)^2}$, and we also have
\[
\| \max_{1 \leq j \leq J} \left( \frac{1}{\Lambda_j} \left( \| \hat{e}_j - e_j \|_{L^2}^2 - \sum_{k \neq j}^{\infty} \frac{I_{k,j}^2}{(\lambda_j - \lambda_k)^2} \right) \right) \|_p \lesssim \frac{J^1/p}{n} n^{-a}.
\]

Theorem 2.3 provides both uniform expansions for $\hat{e}_j$ and the corresponding norm. As before, the factor $J^{1/p}$ in the bound on the RHS can be dropped for fixed $1 \leq j \leq J_n^+$. As an immediate corollary, we obtain a probabilistic version of Lemma 1.1 with correct order.

**Corollary 2.4.** Assume that Assumption 2.1 holds. Then for $1 \leq j \leq J_n^+$
\[
\| \hat{\lambda}_j - \lambda_j \|_p \lesssim \frac{\lambda_j}{\sqrt{n}} \quad \text{and} \quad \| \| \hat{e}_j - e_j \|_{L^2}^2 \|_p \lesssim \frac{\Lambda_j}{n}.
\]

### 2.1. Previous results and comparison

Let us now compare Theorems 2.2 and 2.3 to the literature. It seems that the currently best known expansions in this context can be found in [16]. Among other things, it is required that $\{X_k\}_{k \in \mathbb{Z}}$ is IID, all moments exist, and the error term $ER_{J_n^+}$ in the expansions is of magnitude
\[
ER_{J_n^+} \gtrsim n^{-3/2}(1 - \xi_j)^{-1/2}\psi_j^{-3}\lambda_j^{-1/2}s_j, \quad s_j = \sup_{t \in T} |e_j(t)|, \quad (2.6)
\]
and $\xi_j \in (0, 1)$ is defined as $\xi_j = \inf_{k < j} (1 - \frac{\lambda_k}{\lambda_j})$. We emphasize that this is the overall error term, hence one requires for instance at least $\sqrt{n}ER_{J_n^+} = O(1)$ for the validity of a CLT. If we assume the convexity condition (2.3), we see that (A2) is much weaker. In fact, taking for instance $\lambda_j \sim j^{-\epsilon}$ we find that $ER_{J_n^+} \gtrsim n^{-3/2}(J_n^+)^{3+7/2\epsilon}$. On the other hand, we see from (2.4) that if $J_n^+ \sim n^{1/2-a}$, $a > 0$, we still obtain valid asymptotic expansions, i.e. the expressions containing $I_{k,j}$ are still the principal terms in our expansions, reflecting the exact asymptotic behavior. In stark contrast, $ER_{J_n^+}$ already explodes for a small (resp. large) enough, rendering a vacuous result. Similarly, (A2) is valid if we only require
\[
\max_{1 \leq j \leq J_n^+} n^{-1/2}\psi_j \lesssim n^{-a} \quad \text{for some arbitrary } a > 0, \quad (2.7)
\]
and again obtain valid asymptotic expansions. On the other hand, the actual approximation error $ER_{J_n^+}$ in [16] may even be unbounded, since $1/\lambda_j \to \infty$ as $j$ increases. In this sense, Assumption 2.1 is substantially weaker.

### 2.2. Dependence assumptions: optimality

We first give the following result.
Theorem 2.5. Assume that \( X \) is Gaussian with zero mean and
\[
\sup_{j \in \mathbb{N}} \sum_{k=0}^{\infty} \text{Cov}[\eta_{0,j}, \eta_{k,j}]^2 < \infty. \tag{2.8}
\]
Then (A1) holds. Moreover, if we have in addition (A2) (for \( J_n^+ \) possibly finite), then for any fixed \( 1 \leq j \leq J_n^+ \)
\[
\sqrt{n}(\bar{\lambda}_j - \lambda_j) \xrightarrow{w} \mathcal{N}(0, \lambda_j^2 \sigma_j^2) \quad \text{and} \quad \sqrt{n}(\bar{e}_j - e_j) \xrightarrow{w} \mathcal{N}(0, \Sigma_j),
\]
where \( \xrightarrow{w} \) denotes weak convergence in the corresponding (Hilbert) space, and \( \sigma_j^2 \) (\( \Sigma_j \)) denotes the corresponding variance (operator).

Following the discussion in [2], [6], condition (2.8) is necessary for
\[
n^{-1/2}\|\eta_{i,j}\|_2 < \infty, \quad \text{for any } i, j \in \mathbb{N},
\]
in the sense that if \( \sum_{i,j \leq n} \text{Cov}[\eta_{i,j}, \eta_{k,j}]^2 \gtrsim L(n) \) for some slowly varying function \( L(n) \) and fixed \( j \in \mathbb{N} \), then \( n^{-1/2}\|\eta_{i,j}\|_2 \gtrsim L(n) \). If we restrict ourselves to a more special case we get a strict relation in terms of \( \|\hat{\mathcal{C}}_n - \mathcal{C}_n\|_{L^2} \). Let
\[
\eta_{k,j} = \sum_{i=0}^{\infty} \alpha_{i,j} \epsilon_{k-i,j} \quad \text{with } 0 \leq \alpha_{i,j} \sim i^{-\alpha} \quad \text{and } \epsilon_{k,j} \text{ standard Gaussian IID.}
\]
Then one readily computes that
\[
\|\|\hat{\mathcal{C}}_n - \mathcal{C}_n\|_{L^2} \|_2 \lesssim n^{-1/2} \quad \text{iff } \alpha > 3/4.
\]
However, \( \alpha > 3/4 \) also gives (2.8). Hence by Theorem 2.5 condition (A1) is valid. On the other hand, Lemma 3.4 below yields that (A1) implies \( \|\|\hat{\mathcal{C}}_n - \mathcal{C}_n\|_{L^2} \|_2 \lesssim n^{-1/2} \). Hence we obtain the equivalence in (2.2). Finally, note that the regime \( 1/2 < \alpha \leq 1 \) is generally considered as long memory. Hence by Theorem 2.5 above, we obtain a CLT for \( \bar{\lambda}_j \) and \( \bar{e}_j \) even in the presence of long memory, where \( 3/4 < \alpha \leq 1 \). If \( 1/2 < \alpha < 3/4 \), Non-central limit theorems arise. If \( \alpha \leq 1/2 \), then \( \mathbb{E}[\|X_0\|_{L^2}] = \infty \), which requires a completely different treatment.

2.3. Spectral gap: almost optimality

Next, we discuss the issue of ‘almost optimality’ of condition (A2). To this end, we draw heavily from the noteworthy results of [26]. Suppose that \{\eta_{i,j}\}_{i,j \in \mathbb{N}}\) are IID and satisfy \( \mathbb{E}[|\eta_{i,j}|^{2p}] \leq pC^{p-1} \) for some constant \( C > 0 \). If a structure condition like (EP) holds, then it is shown in [26] that
\[
\mathbb{E}[\|\hat{\epsilon}_j - e_j\|_{L^2}^2] \lesssim \frac{j^2(\log n)^2}{n}. \tag{2.9}
\]
As can be seen from Corollary 2.4, this bound deviates from the optimal one by the additional factor \( (\log n)^2 \). On the other hand, note that in the polynomial case in (EP), this bound is also valid for \( j > J_n^+ \) (we require \( \alpha > 0 \), which is
a slightly larger region. In [26], a lower bound is also provided, which is $\frac{j^2}{n} \wedge 1$. Strictly speaking, it is proven for the projection $\hat{\pi}_j = \hat{e}_j \otimes \hat{e}_j$, where $\otimes$ denotes the one-rank operation

$$u \otimes v(w) = \langle u, w \rangle v, \quad u, v, w \in L^2(T).$$

According to [26], it then holds that (recall that $L$ denotes the operator norm)

$$j^2 n^{-1} \lesssim E[\|\hat{\pi}_j - \pi_j\|_2^2] \lesssim \frac{j^2 (\log n)^2}{n} \wedge 1. \quad (2.10)$$

Heuristically, this may also be inferred from [11]. On the other hand, Corollary 2.4 and elementary computations yield

$$E[\|\hat{\pi}_j - \pi_j\|_2^2] \lesssim \frac{1}{n} \sum_{k=1}^{\infty} \frac{\lambda_j \lambda_k}{(\lambda_j - \lambda_k)^2} \lesssim \frac{j^2}{n}, \quad \text{if } j \leq n^{1/2 - \alpha} (\log n)^{-1}, \quad (2.11)$$

(in the polynomial case) and thus the order of the upper and lower bounds match for $j \leq n^{1/2 - \alpha} (\log n)^{-1}$. If $j \geq n^{1/2}$, Cauchy-Schwarz yields the trivial optimal upper bound. Since $\alpha > 0$ may be chosen arbitrarily small given sufficiently many (all) moments, we find that our conditions on the eigenvalues $\lambda$ are essentially optimal. In other words, we obtain exact expansions and the optimal error bound for almost the complete region of indices $j$ where (2.11) still converges to zero.

### 2.4. Maximum deviation of empirical eigenvalues and trace estimates

As already mentioned, Theorems 2.2 and 2.3 can be used to obtain various fluctuation results for eigenvalues or eigenfunctions. To exemplify this further, we first introduce a popular notion of weak dependence. In the remainder of this section, we assume that for each $j \in \mathbb{N}$, the score sequence $\{\eta_{k,j}\}_{k \in \mathbb{Z}}$ is a causal weak Bernoulli sequence, which can be written as

$$\eta_{k,j} = g_j(\ldots, \epsilon_{k-1,j}, \epsilon_{k,j}) \quad (2.12)$$

for some measurable functions $g_j$ and IID sequences $\{\epsilon_{k,j}\}_{k \in \mathbb{Z}, j \in \mathbb{N}}$. Let $\mathcal{E}_{k,j} = (\epsilon_i, i \leq k)$. To quantify the dependence of $\{\eta_{k,j}\}_{k \in \mathbb{Z}}$, we adopt the coupling idea. Let $\{\epsilon'_{k,j}\}_{k \in \mathbb{Z}, j \in \mathbb{N}}$ be an IID copy of $\{\epsilon_{k,j}\}_{k \in \mathbb{Z}, j \in \mathbb{N}}$ and $\mathcal{E}'_{k,j} = (\mathcal{E}_{-1,j}, \epsilon'_{0,j}, \epsilon_{1,j}, \ldots, \epsilon_{k,j})$ the coupled version of $\mathcal{E}_{k,j}$. Then we define

$$\Omega_p(k) = \max_{j \in \mathbb{N}} \|\eta_{k,j} - \eta'_{k,j}\|_p \quad \text{for } p \geq 1, \quad \text{where } \eta'_{k,j} = g_j(\mathcal{E}'_{k,j}). \quad (2.13)$$

Roughly speaking, $\Omega_p(k)$ measures the overall degree of dependence of $\eta_{k,j} = g_j(\mathcal{E}_{k,j})$ on $\mathcal{E}_{0,j}$ and it is directly related to the data-generating mechanism of the underlying process ( [34] refers to $\Omega_p(k)$ as physical dependence measure).
This dependence concept is well established in the literature, and many popular examples like ARMA, GARCH, many Markov processes etc. fit into this framework (cf. [34], [35]). Consider for example the linear process
\[ \eta_{k,j} = \sum_{l=0}^{\infty} \alpha_l \epsilon_{k-l,j} \]
where \( \{\epsilon_{k,j}\}_{k \in \mathbb{Z}, j \in \mathbb{N}} \) is IID with \( \|\epsilon_{k,j}\|_p < \infty \). Then
\[ \sum_{k=1}^{\infty} \Omega_p(k) < \infty \quad \text{holds iff} \quad \sum_{k=1}^{\infty} |\alpha_k| < \infty. \] (2.14)

In this sense, (2.14) is necessary for a CLT. In fact, if it is violated, one can construct examples such that
\[ \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \eta_{k,j}^{2} = \infty, \quad j \in \mathbb{N}, \]
and a different normalization than \( n^{-1/2} \) is required (cf. [31]). In the sequel, all dependence conditions will be expressed in terms of summability conditions of \( \Omega_p(k) \). We start with the following result.

**Proposition 2.6.** Assume that \( \sum_{k=1}^{\infty} \Omega_{2p}(k) < \infty \) for \( p \geq 2 \). Then
\[ \max_{i,j} \| n^{-1/2} \eta_{i,j} \|_p < \infty. \]

Proposition 2.6 shows that condition \( (A1) \) holds under sharp weak dependence conditions, including a large variety of linear and non-linear processes, see the earlier references above. Related results can be established under different weak dependence conditions, see for instance [30], [32] or [12]. Next, we formally introduce the longrun covariance
\[ \gamma_{\lambda,i,j} = \lim_{n \to \infty} n^{-1} \mathbb{E} \left[ \sum_{k=1}^{n} \sum_{l=1}^{n} (\eta_{k,i}^{2} - 1)(\eta_{l,j}^{2} - 1) \right]. \] (2.15)

In Section 5.1 we show that this is well-defined given Assumption 2.7 below. Moreover, for \( \sigma_{\lambda,h}^{2} = \gamma_{\gamma,h,h} \) we have the usual representation \( \sigma_{\lambda,h}^{2} = \sum_{k \in \mathbb{Z}} \phi_{k,h} \), where \( \phi_{k,h} = \text{Cov}[\eta_{0,h}^{2}, \eta_{k,h}^{2}] \). Denote with
\[ T^{\lambda}_{J} = \sqrt{n} \max_{1 \leq j \leq J} \left| \frac{\hat{\lambda}_{j} - \lambda_{j}}{\sigma_{\lambda,j}^{\lambda}} \right|, \quad T^{Z^{\lambda}}_{J} = \max_{1 \leq h \leq J} \left| Z_{\lambda,h}^{\lambda} \right|, \] (2.16)
where \( \{Z_{\lambda,h}\}_{1 \leq h \leq J} \) is a zero mean sequence of Gaussian random variables with covariance structure \( \Sigma_{J}^{\lambda} = \left( \rho_{\lambda,i,j} \right)_{1 \leq i,j \leq J} \), where \( \rho_{\lambda,i,j} = \gamma_{\lambda,i,j}/\sigma_{\lambda,i}\sigma_{\lambda,j} \). In the sequel, we show that \( T^{\lambda}_{J} \) is close to \( T^{Z^{\lambda}}_{J} \) in probability. To this end, we work under the following assumption.

**Assumption 2.7.** For \( p \geq 1 \) let \( p^{*} = p^{2p+4} \), \( p = \lceil m/a \rceil \), and assume that

\( (B1) \quad \mathbb{E}[X_{k}] = 0 \) and \( (A2) \) holds (with \( p, a, m \) as above) such that
\[ \left( J_{n}^{+} \right)^{1/p} n^{-\alpha} \lesssim n^{-\delta}, \quad \delta > 0, \]
(B2) $\Omega_k(2p^*) \lesssim j^{-b}$, $b > 3/2$.

(B3) $\inf h \sigma_{\lambda,h} > 0$.

Note that these assumptions are mild. In particular, the decay rate $b$ in condition (B2) is completely independent of the underlying dimension $J^+_n$. We then have the following result.

**Theorem 2.8.** Grant Assumption 2.7. Then

$$\sup_{x \in \mathbb{R}} |P(T^\lambda_{J^+_n} \leq x) - P(T^{Z^\lambda}_{J^+_n} \leq x)| \lesssim n^{-C}, \quad C > 0.$$  

The above result provides a Gaussian approximation with an explicit, algebraic rate. Note that no conditions on the underlying covariance structure are required. If we impose a very weak decay assumption on $\gamma_{\lambda,i,j}$, we obtain the limit distribution in the corollary below.

**Corollary 2.9.** Grant Assumption 2.7, and assume in addition $|\gamma_{\lambda,i,j}| \log(|i - j|) = O(1)$ for $|i - j| > 1$.  

Then

$$\lim_{n \to \infty} P(T^\lambda_{J^+_n} \leq u_{J^+_n}(x)) = \exp(-e^{-x}),$$

where $u_m(x) = x/a_m + b_m$ with $a_m = (2 \log m)^{1/2}$ and $b_m = (2 \log m)^{1/2} - (8 \log m)^{-1/2}(\log \log m + 4\pi - 4)$ for $m \in \mathbb{N}$ and $x \in \mathbb{R}$.

**Remark 2.10.** Not that condition (2.17) is the weakest possible currently known, see [22], [23] and [17].

Uniform control measures are an important statistical tool and have many applications. In the present context, Corollary 2.9 allows for the construction of simultaneous confidence bands for $\hat{\lambda}_j$. This in turn is very useful to assess parametric hypothesis and decay rates of the structure of $\lambda$. A particular and important case is the determination of relevant principle components, i.e; where $\lambda_j$ is still large enough subject to some threshold (cf. [21], [19]). Another important measure in this context are the partial trace and inverse partial trace, which were already introduced in (1.4). Concerning these two measures, we have the following result.

**Theorem 2.11.** Assume (A2) and $\sum_{k=1}^{\infty} \Omega_{2p^*}(k) < \infty$ for $p^* = p \times 2^{p+4}$, $p \geq 1$. If $(J^+_n)^{1/p} n^{-1/2} = o(1)$, then

(a) $\max_{1 \leq j \leq J^+_n} \left| \sum_{j=1}^{J^+_n} \hat{\lambda}_j - \sum_{j=1}^{J^+_n} \lambda_j \right| = O_p(n^{-1/2})$,

(b) $\max_{1 \leq j \leq J^+_n} \left| \sum_{j=1}^{J^+_n} \frac{1}{\lambda_j} - \sum_{j=1}^{J^+_n} \frac{1}{\lambda_j} \right| = O_p(n^{-1/2}) \sum_{j=1}^{J^+_n} \frac{1}{\lambda_j}$. 
Note that for any fixed $j \in \mathbb{N}$ we have
\[ \sqrt{n}(\hat{\lambda}_j - \lambda_j) \xrightarrow{w} \mathcal{N}(0, 2\lambda_j^2), \] (2.18)
provided that $X$ is IID and Gaussian (cf. [11]). Hence the above rate in (a) is optimal, which can be inferred from the univariate case where the trace simply is the variance, for instance by the Cramér-Rao lower bound. Whether the additional factor $\sum_{j=1}^{J_n} \frac{1}{\lambda_j}$ in (b) is unavoidable remains open. As demonstrated earlier, condition $\sum_{k=j}^{\infty} \Omega_{2p'}(k) < \infty$ cannot be improved in terms of weak dependence and is sharp. However, if we purely focus on Gaussian processes, we can obtain the same rate under the same conditions as in Theorem 2.5.

3. Proofs of asymptotic expansions

Throughout the proofs, we make the following simplification. We assume that $\mu = 0$, and drop the estimator $\hat{X}_n$ in the definition of the covariance operator $\hat{C}_n$. This makes the proofs notationally less burdened. Note however, that all proofs readily carry over to the more general case $\mu \neq 0$. We now introduce the following additional notation. Given functions $f, g \in L^2(T)$ and a Kernel $B(r,s)$, we write
\[ \int_T fg = \int_T f(r)g(r)dr \quad \text{and} \quad \int_{T^2} B f g = \int_{T^2} B(r,s)f(r)g(s)dr ds. \] (3.1)
If we have $f = g$, then we write $f^2 = f(r)^2$ and otherwise $f g = f(r)g(s)$ in the above notation. We interchangeably use $\langle \cdot, \cdot \rangle$ and $\int_T \cdot$, the latter being more convenient when dealing with Kernels. We also frequently apply Fubini-Tonelli without mentioning it any further. Next, we introduce the covariance Kernel $C$ and its empirical version $\hat{C}_n$ as
\[ C = C(r,s) = \mathbb{E}[X_k(r)X_k(s)], \quad \hat{C}_n = \hat{C}_n(r,s) = \frac{1}{n} \sum_{k=1}^{n} X_k(r)X_k(s). \] (3.2)
Note that $\mathbb{E}[\hat{C}_n] = C$. The proofs of Theorems 2.2 and 2.3 are developed in a series of lemmas. As starting point, we recall the following elementary preliminary result (cf. [5]).

Lemma 3.1. We have the decomposition
\[ \hat{\lambda}_j \int_T e_k(\hat{e}_j - e_j) = \lambda_k \int_T e_k(\hat{e}_j - e_j) + \int_{T^2} (\hat{C}_n - C)e_k e_j + \int_{T^2} (\hat{C}_n - C)e_k(\hat{e}_j - e_j). \] (3.3)
Rearranging terms, we obtain from the above that (provided $\lambda_k \neq \lambda_j$)
\[
\int_T e_k(\hat{e}_j - e_j) = \frac{1}{\lambda_j - \lambda_k} \left( \int_{T^2} (\hat{C}_n - C)e_k e_j \right.
+ \int_{T^2} (\hat{C}_n - C)e_k (\hat{e}_j - e_j) + (\hat{\lambda}_j - \lambda_j) \int_T e_k(\hat{e}_j - e_j) \biggr)
\]
def \[1/\lambda_j - 1/\lambda_k \left( I_{k,j} + II_{k,j} + III_{k,j} \right), \tag{3.4} \]

and
\[
\int_T e_k(\hat{e}_j - e_j) = \frac{\lambda_j - \lambda_k}{\lambda_j - \lambda_k + \lambda_j - \lambda_k} \frac{1}{\lambda_j - \lambda_k} \left( I_{k,j} + II_{k,j} \right). \tag{3.5} \]

Due to the frequent use of relations (3.4), (3.5) it is convenient to use the abbreviation

\[
E_{k,j} = \int_T e_k(\hat{e}_j - e_j) = \langle e_k, \hat{e}_j - e_j \rangle
\]
in the sequel. We also recall the following lemma (cf. [5]).

**Lemma 3.2.** For any \( j \in \mathbb{N} \) we have

\[
\int_T (\hat{e}_j - e_j)(\hat{e}_j - e_j) = \frac{1}{2} \| \hat{e}_j - e_j \|_{L^2}^2 \quad \text{and} \quad \int_T (\hat{e}_j - e_j)(\hat{e}_j - e_j) = -\frac{1}{2} \| \hat{e}_j - e_j \|_{L^2}^2.
\]

We proceed by deriving subsequent bounds for \( I_{k,j}, II_{k,j} \) and \( III_{k,j} \).

**Lemma 3.3.** Assume that Assumption 2.1 holds. Then for \( 1 \leq q \leq p^2 + 4 \) we have

\[
\| I_{k,j} \|_q \lesssim \sqrt{\lambda_j \lambda_k} n^{-1/2} \text{ uniformly for } k,j \in \mathbb{N}.
\]

**Proof of Lemma 3.3.** Using the orthogonality of \( e_j, e_k \) we have

\[
I_{k,j} = \int_{T^2} (\hat{C}_n - C)e_k e_j = n^{-1/2} \sqrt{\lambda_k \lambda_j} n^{-1/2} \eta_{k,j},
\]

hence the claim follows from Assumption 2.1.

**Lemma 3.4.** Assume that Assumption 2.1 holds. Then for \( 1 \leq q \leq p^2 + 3 \) we have

\[
\| \hat{C}_n - C \|_q \lesssim \| \hat{C}_n - C \|_{L^2} \lesssim n^{-1/2}.
\]

**Proof of Lemma 3.4.** Since the Hilbert-Schmidt norm dominates the Operator norm, Parseval’s identity and Lemma 3.3 yield the claim, recalling that \( \mathbb{E}[\| X_0 \|_{L^2}^2] < \infty \) implies \( \sum_{j=1}^{\infty} \lambda_j < \infty \).
Lemma 3.5. Assume that Assumption 2.1 holds. Then for \(1 \leq q \leq p2^{p+4}\) and \(k \in \mathbb{N}\) we have
\[
\max_{1 \leq j \leq J_n^*} \frac{|II_{k,j}|}{\|e_j - e_j\|_q} \lesssim \sqrt{\lambda_k}^{-1/2}.
\]

Proof of Lemma 3.5. It holds that
\[
II_{k,j} = \int_{\mathbb{T}^2} (\hat{C}_n - C)e_k(e_j - e_j) = n^{-1} \sum_{i=1}^{\infty} \sqrt{\lambda_k} \psi_{k,i} E_{i,j}.
\]

Since \(\sum_{i=1}^{\infty} E_{i,j}^2 = \|e_j - e_j\|_2^2\) by Parseval’s identity, the Cauchy-Schwarz inequality gives
\[
\left| \sum_{i=1}^{\infty} \sqrt{\lambda_k} \psi_{k,i} \right| \leq \left( \sum_{i=1}^{\infty} \lambda_k \eta_{k,i}^2 \right)^{1/2} \|e_j - e_j\|_2.
\]

Hence the triangle inequality and Assumption 2.1 yield
\[
\max_{1 \leq j \leq J_n^*} \frac{|II_{k,j}|}{\|e_j - e_j\|_q} \leq n^{-1/2} \sqrt{\lambda_k} \left( \sum_{i=1}^{\infty} \lambda_k \eta_{k,i}^2 \right)^{1/2} \lesssim n^{-1/2} \sqrt{\lambda_k}.
\]

Lemma 3.6. Assume that Assumption 2.1 holds, and let \(A_j = \{|\hat{\lambda}_j - \lambda_j| \leq \psi_j/2\}\). Then
\[
\max_{1 \leq j \leq J_n^*} P(A_j) \lesssim n^{-ap2^{p+4}+4}.
\]

Proof of Lemma 3.6. For the proof, we denote with \(\|M\|\) the maximum (or \(\infty\))-norm for a matrix \(M\). We proceed similarly as in [26], employing an approach of [13]. It suffices to treat the case \(\hat{\lambda}_j \leq \lambda_j - \psi_j/2\), the complementary case can be handled in the same manner. Let \(E^*_j = \text{span}\{e_1, \ldots, e_j\}\). Then using the minimax definition of eigenvalues, one may show that
\[
P(\hat{\lambda}_j \leq \lambda_j - \psi_j/2) \leq P\left( \min_{x \in E^*_j, \|x\|_2 = 1} \int_{\mathbb{T}^2} (\hat{C}_n - C)xx \leq -\psi_j/2 \right).
\]

Denote with \(M_j\) the \(j\)-dimensional matrix defined by \((M_j)_{k,l} = n^{-1/2} \delta_{k,l}, 1 \leq k, l \leq j\), and with \(D_{\lambda}\) the diagonal matrix of size \(j\) with \((D_{\lambda})_{k,k} = \sqrt{\lambda_k - \lambda_{k+1}}\), \(1 \leq k \leq j\). Proceeding as in the proof of [26, Lemma 14], it follows that
\[
P(\hat{\lambda}_j \leq \lambda_j - \psi_j/2) \leq P\left( \|D_{\lambda}^{-1}M_jD_{\lambda}^{-1}\| \geq \psi_j/(2|\lambda_j+1 - \lambda_j|) \right)
\leq P\left( \|D_{\lambda}^{-1}M_jD_{\lambda}^{-1}\| \geq C \right).
\]
for some absolute constant $C > 0$. Let $p^* = p^{2p+4}$. Then by the triangle, Cauchy-Schwarz inequality and Lemma 3.3

$$
\left\| \|D_{\lambda}^{-1} M_j D_{\lambda}^{-1}\| \right\|_{p^*/2} \leq \left\| \frac{1}{n} \sum_{k,l=1}^{j} |\lambda_k - \lambda_{j+1}| |\lambda_l - \lambda_{j+1}| \right\|_{p^*/2} \lesssim \left( \frac{1}{\sqrt{n}} \sum_{k=1}^{j} |\lambda_k - \lambda_{j+1}| \right)^2 \lesssim n^{-2a},
$$

where the last inequality follows from (A2). Hence we conclude via Markov’s inequality that

$$
P(\|D_{\lambda}^{-1} M_j D_{\lambda}^{-1}\| \geq C) \lesssim n^{-pa^*},
$$

and the claim follows.

The next result is our key technical lemma.

**Lemma 3.7.** Assume that Assumption 2.1 holds. Then uniformly for $1 \leq q \leq p^{2p/2+3}$, $k \in \mathbb{N}$ and $1 \leq j \leq J^+$

$$
\left\| II_{k,j} 1(A_j) \right\|_q \lesssim \frac{\sqrt{\lambda_k \lambda_j}}{\sqrt{n}} \left( \left\| \hat{e}_j - e_j \right\|^{2q}_2 + n^{-a} \right).
$$

**Proof of Lemma 3.7.** Note first that by construction of $A_j$, we have that

$$
\left| \frac{\lambda_j - \lambda_l}{\lambda_j - \lambda_j + \lambda_j - \lambda_l} 1(A_j) \right| \leq 2, \quad \text{for } l \neq j. \quad (3.9)
$$

Using the decomposition in (3.5) and bound (3.9), we obtain that

$$
| E_{l,j} 1(A_j) | \leq \frac{2}{|\lambda_j - \lambda_l|} (|I_{l,j}| + |I_{l,j}|) 1(A_j). \quad (3.10)
$$

We now use a backward inductive argument. Let $p_l = p^{2l}$, $\tau \geq 0$, and suppose we have uniformly for $k \in \mathbb{N}$

$$
\left\| II_{k,j} 1(A_j) \right\|_{p_l} \lesssim n^{-1/2} \sqrt{\lambda_k} \left( \sqrt{\lambda_j + n^{-\tau}} \right) \text{ for some } l \leq p + 4. \quad (3.11)
$$

Then we obtain from (3.10), the triangle inequality and Lemma 3.3 that for $l \neq j$

$$
\left\| E_{l,j} 1(A_j) \right\|_{p_l} \lesssim n^{-1/2} \frac{\sqrt{\lambda_j}}{|\lambda_j - \lambda_l|} \left( \sqrt{\lambda_j + n^{-\tau}} \right). \quad (3.12)
$$

Using decomposition (3.6) and the Cauchy-Schwarz inequality, we get

$$
\left\| II_{k,j} 1(A_j) \right\|_{p_{l-1}} \leq \frac{\sqrt{\lambda_k}}{n} \sum_{l=1}^{\infty} \sqrt{\lambda_l} \left\| E_{l,j} 1(A_j) \right\|_{p_l} \| \eta_{k,l} \|_{p_l},
$$
hence we obtain from Lemma 3.2, inequality (3.12) and (A2) that
\[ \|II_{k,j}1(A_j)\|_{p_{l-1}} \lesssim \sqrt{\frac{\lambda_j}{n}} \left( \sqrt{\lambda_j} \|\tilde{e}_j - e_j\|_2^2 \|p_i + \frac{1}{n} \sum_{l=1}^{\infty} \frac{\lambda_l}{|\lambda_l - \lambda_j|} (\sqrt{\lambda_j} + n^{-\tau}) \right) \]
\[ \lesssim \sqrt{\frac{\lambda_j}{n}} \left( \|\tilde{e}_j - e_j\|_2^2 \|p_i + n^{-\sigma} \right) \sqrt{\lambda_j + n^{-\sigma - \tau}}, \quad (3.13) \]
and this bound holds uniformly for \( k \in \mathbb{N} \). Observe that we have now shown the validity of relation (3.11) with the updated value \( \tau = \tau + a \), but with respect to \( p_{l-1} \) instead of \( p_l \). Since \( \lambda_j \gtrsim n^{-m} \) with \( m \geq 1 \), it follows that after at most \( p/2 + 1 = [m/a]/2 + 1 \) iterations we have
\[ \|II_{k,j}1(A_j)\|_{q^{*}} \lesssim \sqrt{\frac{\lambda_k \lambda_j}{n}} \left( \|\tilde{e}_j - e_j\|_2^2 \right)^{2q^{*}} + n^{-\sigma}, \]
where \( q^{*} = p2^{p/2+3} \). By Lemma 3.5, relation (3.11) is true for \( \tau = 0 \) (hence \( n^{\tau} = 1 \)) and \( l = p + 4 \), constituting the basis induction step, hence the proof is complete. Note that we have also shown
\[ \|E_{l,j}1(A_j)\|_{q^{*}} \lesssim n^{-1/2} \sqrt{\frac{\lambda_k \lambda_j}{|\lambda_j - \lambda_l|}}, \quad (3.14) \]
which is of further relevance in the sequel.

\[ \boxdot \]

**Proposition 3.8.** Assume that Assumption 2.1 holds. Then for \( 1 \leq q \leq p2^{p/2+2} \) we have uniformly for \( 1 \leq j \leq J_n^+ \)
\[ \|\tilde{e}_j - e_j\|_2^2 \|q \leq P(A_j^{1/q})^{1/q} + n^{-1} \sum_{\substack{k=1 \atop k \neq j}}^{\infty} \frac{\lambda_k \lambda_j}{(\lambda_k - \lambda_j)^2} \lesssim n^{-2a}. \]

**Proof of Proposition 3.8.** The triangle inequality and Cauchy-Schwarz give
\[ \|\tilde{e}_j - e_j\|_2^2 \|q \leq 2P(A_j^{1/q})^{1/q} \|\tilde{e}_j - e_j\|_2^2 \|1(A_j)\|_q \|. \quad (3.15) \]
We now invoke the ‘traditional’ way of bounding \( \|\tilde{e}_j - e_j\|_2^2 \|1(A_j)\|_q \), (cf. [5], [19]), which uses the inequality
\[ \|\tilde{e}_j - e_j\|_2^2 \| \leq 2 \sum_{\substack{k=1 \atop k \neq j}}^{\infty} E_{k,j}^2. \quad (3.16) \]
Hence using (3.14) and the triangle inequality, we obtain from (A2) that
\[ \|\tilde{e}_j - e_j\|_2^2 \|1(A_j)\|_q \leq 2 \sum_{\substack{k=1 \atop k \neq j}}^{\infty} E_{k,j}^2 \|1(A_j)\|_q \leq \frac{1}{n} \sum_{\substack{k=1 \atop k \neq j}}^{\infty} \frac{\lambda_k \lambda_j}{(\lambda_k - \lambda_j)^2} \lesssim n^{-2a}. \]
Combining this with (3.15) gives the first inequality, Lemma 3.6 and Assumption 2.1 yield the second part.

Note that \( a \leq 1/2 \) and hence \( p/2 \geq m \geq 1 \) and \( 2^{p/2+2} \geq 8 \). Since

\[
\|\|\hat{e}_j - e_j\|_2^2\|_{2q} \leq \sqrt{2}\|\|\hat{e}_j - e_j\|_2\|_{2q}^{1/2}
\]

for \( q \geq 1 \), we obtain the following corollary to Lemma 3.7.

**Corollary 3.9.** Assume that Assumption 2.1 holds. Then for \( 1 \leq q \leq 8p \) we have uniformly for \( k \in \mathbb{N} \) and \( 1 \leq j \leq J_n^+ \)

\[
\|I_{k,j}\|_q \lesssim \frac{\sqrt{\lambda_j \lambda_k}}{\sqrt{n}} n^{-a}.
\]

**Proof of Corollary 3.9.** Lemma 3.5, Lemma 3.6, Lemma 3.7 and Cauchy-Schwarz give

\[
\|I_{k,j}\|_q \leq \|I_{k,j}1(A_j)\|_q + \|I_{k,j}1(A_j^c)\|_q \leq \frac{\sqrt{\lambda_j \lambda_k}}{\sqrt{n}} n^{-a} + \frac{\sqrt{\lambda_k}}{\sqrt{n}} n^{-a} P(A_j^c)^{1/2q}
\]

\[
\lesssim \frac{\sqrt{\lambda_j \lambda_k}}{\sqrt{n}} n^{-a} + \frac{\lambda_k}{\sqrt{n}} n^{-a} n^{-ap2^{p+3}/q}.
\]

Since \( ap2^{p+3}/q \geq a2^p \geq m \), we have \( n^{-ap2^{p+3}/q} \lesssim \lambda_{J_n^+} \) by (A2) and the claim follows.

**Lemma 3.10.** Assume that Assumption 2.1 holds. Then for \( 1 \leq q \leq 4p \)

\[
\|\hat{\lambda}_j - \lambda_j - I_{j,j}\|_q \lesssim \frac{\lambda_j}{\sqrt{n}} n^{-a}, \quad \text{and} \quad \|\hat{\lambda}_j - \lambda_j\|_q \lesssim \frac{\lambda_j}{\sqrt{n}}, \quad \text{uniformly for } 1 \leq j \leq J_n^+.
\]

**Proof of Lemma 3.10.** We have that

\[
\hat{\lambda}_j = \int_{T^2} \hat{C}_n \hat{e}_j \hat{e}_j = \int_{T^2} \hat{C}_n (\hat{e}_j - e_j) \hat{e}_j + \int_{T^2} \hat{C}_n e_j \hat{e}_j
\]

\[
= \hat{\lambda}_j \int_{T^2} (\hat{e}_j - e_j) \hat{e}_j + \int_{T^2} (\hat{C}_n - C) e_j \hat{e}_j + \int_{T^2} C_n e_j \hat{e}_j
\]

\[
= \frac{\hat{\lambda}_j}{2} \|\hat{e}_j - e_j\|_2^2 + \int_{T^2} (\hat{C}_n - C) e_j (\hat{e}_j - e_j) + \int_{T^2} (\hat{C}_n - C) e_j e_j + \int_{T^2} C_n e_j \hat{e}_j.
\]

Since by Lemma 3.2

\[
\int_{T^2} C_n e_j \hat{e}_j = \int_{T^2} C_n e_j (\hat{e}_j - e_j) + \int_{T^2} C_n e_j e_j = -\frac{\lambda_j}{2} \|\hat{e}_j - e_j\|_2^2 + \lambda_j,
\]

\[
\]
we obtain by rearranging terms (if \( \| \bar{e}_j - e_j \|_{L^2}^2 < 2 \))
\[
\hat{\lambda}_j - \lambda_j = \frac{2}{2 - \| \bar{e}_j - e_j \|_{L^2}^2} \left( \int_{T^2} (\mathcal{C}_n - C) e_j e_j + \int_{T^2} (\mathcal{C}_n - C) (\bar{e}_j - e_j) \right) \\
= \frac{2}{2 - \| \bar{e}_j - e_j \|_{L^2}^2} (I_{j,j} + II_{j,j}).
\] (3.17)

Let \( B_j = \{ \| \bar{e}_j - e_j \|_{L^2}^2 \leq 1 \} \). By Lemma 3.3, Proposition 3.8 and the Cauchy-Schwarz inequality we obtain
\[
\left\| I_{j,j} \left( 1 - \frac{2}{2 - \| \bar{e}_j - e_j \|_{L^2}^2} \right) 1(B_j) \right\|_q \lesssim \left\| I_{j,j} \right\|_{2q} \left\| \bar{e}_j - e_j \right\|_{L^2}^2 \lesssim \frac{\lambda_j}{\sqrt{n}} n^{-2a}.
\] (3.18)

Similarly, Corollary 3.9 yields that
\[
\left\| II_{j,j} \left( 1 - \frac{2}{2 - \| \bar{e}_j - e_j \|_{L^2}^2} \right) 1(B_j) \right\|_q \lesssim \frac{\lambda_j}{\sqrt{n}} n^{-a}.
\] (3.19)

Let \( C = \{ \| \mathcal{C}_n - C_n \|_{L^2} \leq 1 \} \). Lemma 3.4 and Markov’s inequality then yield that
\[
P(C^c) \lesssim n^{-2a/2+3}. \] (3.20)

On the other hand, Proposition 3.8 implies that \( P(B_j^c) \lesssim n^{-2a/2+2} \). Since \( m \geq 1, 1/2 > a \) we have \( 2^{p/2} \geq 1/2 + 1/4a + m/2a \) and hence \( n^{-2a/2} \lesssim n^{-1/2-a} \lambda_j \). Combining (3.17), (3.18), (3.19) and (3.20) we obtain from the Cauchy-Schwarz inequality, Lemma 1.1 and Lemma 3.4, that
\[
\left\| \hat{\lambda}_j - \lambda_j - I_{j,j} \right\|_q \lesssim P(B_j^c)^{1/q} + \left\| \mathcal{C}_n - C_n \right\|_{L^q} P(C^c)^{1/2q} + \frac{\lambda_j}{\sqrt{n}} n^{-a} \\
\lesssim \frac{\lambda_j}{\sqrt{n}} n^{-a},
\]
which gives the first claim. The second claim follows from Lemma 3.3.

\[\square\]

**Lemma 3.11.** Assume that Assumption 2.1 holds. Then for \( 1 \leq q \leq 2p \) we have uniformly for \( k \in \mathbb{N} \) and \( 1 \leq j \leq J_n^+ \)
\[
\left\| III_{k,j} 1(A_j) \right\|_q \lesssim \frac{\lambda_j}{n} \frac{\sqrt{\lambda_k \lambda_j}}{\lambda_k - \lambda_j} \lesssim \frac{\sqrt{\lambda_k \lambda_j}}{\sqrt{n}} n^{-a}.
\]

**Proof of Lemma 3.11.** Recall that \( III_{k,j} = (\hat{\lambda}_j - \lambda_j) E_{k,j} \). By the Cauchy-Schwarz inequality and Lemma 3.10, we have that
\[
\left\| III_{k,j} 1(A_j) \right\|_q \lesssim \frac{\lambda_j}{\sqrt{n}} \left\| E_{k,j} 1(A_j) \right\|_{2q}.
\]
Hence the claims follow from inequality (3.14) resp. Assumption (A2).

\[\square\]
For the sake of reference, we state Pisier’s inequality.

**Lemma 3.12.** Let $p \geq 1$ and $Y_j$, $1 \leq j \leq J$ be a sequence of random variables. Then

$$\left\| \max_{1 \leq j \leq J} |Y_j| \right\|_p \leq \left( \sum_{j=1}^{J} \|Y_j\|_p^p \right)^{1/p} \leq J^{1/p} \max_{1 \leq j \leq J} \|Y_j\|_p.$$  

We are now ready to prove Theorems 2.2 and 2.3.

**Proof of Theorem 2.2.** This readily follows from Lemma 3.10 and Lemma 3.12.

**Proof of Theorem 2.3.** We treat the first claim. By Lemma 3.2 we have the decomposition

$$\hat{\epsilon}_j - e_j = -\frac{\epsilon_j}{2} \|\hat{\epsilon}_j - e_j\|_{L^2}^2 + \sum_{k=1}^{\infty} \epsilon_k \frac{I_{k,j} + II_{k,j} + III_{k,j}}{\lambda_j - \lambda_k} \overset{\text{def}}{=} -A_j + B_j. \quad (3.21)$$

Note that by the triangle inequality

$$\|B_j\|_{L^2} \leq \|\hat{\epsilon}_j - e_j\|_{L^2} + \|A_j\|_{L^2} \leq 4.$$

Let $C_j = \sum_{k=1}^{\infty} \epsilon_k \frac{I_{k,j}}{\lambda_j - \lambda_k}$. Then another application of the triangle inequality gives

$$\|\hat{\epsilon}_j - e_j + A_j - C_j\|_{L^2} \leq \|B_j\|_{L^2} + \|C_j\|_{L^2} \leq 4 + \|C_j\|_{L^2}.$$

Hence by the Cauchy-Schwarz inequality and Lemma 3.3

$$\|\|\hat{\epsilon}_j - e_j + A_j - C_j\|_{L^2} \|1(A_j^c)\|_p \lesssim 4P(A_j^c)^{1/p} + P(A_j^c)^{1/2p} \left( \frac{1}{n} \sum_{k=1 \neq j}^{\infty} \frac{\lambda_j \lambda_k}{(\lambda_j - \lambda_k)^2} \right)^{1/2},$$

which by Lemma 3.6 and (A2) (arguing as in the proof of Lemma 3.10) is bounded by

$$\|\|\hat{\epsilon}_j - e_j + A_j - C_j\|_{L^2} \|1(A_j^c)\|_p \lesssim n^{-1/2-a} (\lambda_j + \sqrt{A_j}).$$

Lemma 3.12 and the inequality $\Lambda_j \geq \frac{\lambda_j}{\lambda_{j-1}} \geq \lambda_j \wedge 1$ then show that it suffices to consider event $A_j$. Corollary 3.9 and Lemma 3.11 give

$$\left\| \sum_{k=1 \neq j}^{\infty} \frac{(II_{k,j} + III_{k,j})^2}{(\lambda_j - \lambda_k)^2} 1(A_j) \right\|_p \lesssim n^{-1-a} \sum_{k=1 \neq j}^{\infty} \frac{\lambda_j \lambda_k}{(\lambda_j - \lambda_k)^2}.$$  \quad (3.22)
hence the first claim follows from Lemma 3.12. Next, we treat the second claim. As before Lemma 3.2 yields
\[ \|\tilde{e}_j - e_j\|_{L^2}^2 = \frac{1}{4} \|\tilde{e}_j - e_j\|_{L^2}^4 + \sum_{k=1, k \neq j}^{\infty} \frac{(I_{k,j} + II_{k,j} + III_{k,j})^2}{(\lambda_j - \lambda_k)^2}. \]

Proceeding as in the first claim, one shows that it suffices to consider the event $A_j$. Let $C_j = \{\|\tilde{e}_j - e_j\|_{L^2}^2 \leq n^{-a}\}$. Then proceeding as in Lemma 3.10 we obtain
\[ P(C_j^c) \lesssim n^{-ap^{2/2+2}} \lesssim n^{-p-2ap} \lambda_j^p. \quad (3.23) \]

We thus obtain from Lemma 3.3, Corollary 3.9, Lemma 3.11 and (3.23)
\[ \left\| \left( \left\|\tilde{e}_j - e_j\|_{L^2}^2 - \sum_{k=1, k \neq j}^{\infty} \frac{I_{k,j}^2}{(\lambda_j - \lambda_k)^2} \right) 1(A_j) \right\|_p \lesssim n^{-a} \left\|\tilde{e}_j - e_j\|_{L^2}^2 1(A_j) \right\|_p + \sum_{k=1, k \neq j}^{\infty} \frac{(I_{k,j} + II_{k,j} + III_{k,j})^2 - I_{k,j}^2}{(\lambda_j - \lambda_k)^2} 1(A_j) \right\|_p \]
\[ \lesssim n^{-a} \left\|\tilde{e}_j - e_j\|_{L^2}^2 1(A_j) \right\|_p + n^{-1-2a} \lambda_{j_n} + n^{-1-a} \sum_{k=1, k \neq j}^{\infty} \frac{\lambda_j \lambda_k}{(\lambda_j - \lambda_k)^2}. \quad (3.24) \]

Iterating this inequality once and rearranging terms, Lemma 3.3 yields that
\[ \left\| \left( \left\|\tilde{e}_j - e_j\|_{L^2}^2 - \sum_{k=1, k \neq j}^{\infty} \frac{I_{k,j}^2}{(\lambda_j - \lambda_k)^2} \right) \right\|_p \lesssim n^{-1-2a} \lambda_{j_n} + n^{-1-a} \sum_{k=1, k \neq j}^{\infty} \frac{\lambda_j \lambda_k}{(\lambda_j - \lambda_k)^2}. \quad (3.25) \]

Since $\Lambda_j \geq \frac{\lambda_j}{\lambda_j - 1} \geq \lambda_j \wedge 1$, an application of Lemma 3.12 yields the desired result.

of Corollary 2.4. The claim follows from Lemma 3.3, the triangle inequality and Theorems 2.2 and 2.3.

\[ \square \]

4. Proofs of Lemma 4.1 and Theorem 2.5

We first provide the following result about the convexity relations of $\lambda_x$.

**Lemma 4.1.** If (2.3) holds, then (2.4) is valid.
Proof of Lemma 4.1. For the proof, the following relations are useful, which can be found in [8], [10].

If \( j > k \) and (2.3) holds, then \( k\lambda_k \geq j\lambda_j \) and \( \lambda_k - \lambda_j \gtrsim \left(1 - \frac{k}{j}\right)\lambda_k \).

Moreover, it holds that \( \sum_{k > j} \lambda_k \leq (j + 1)\lambda_j \). (4.1)

Now by (4.1) we have
\[
\sum_{k=1}^{\infty} \frac{\lambda_k}{(\lambda_j - \lambda_k)^2} \leq j^2 \sum_{j > k} \frac{\lambda_j \lambda_k}{(k-j)^2\lambda_k^2} + \sum_{j < k} \frac{k^2\lambda_j \lambda_k}{(k-j)^2\lambda_j^2} + \sum_{j < k} \frac{\lambda_j \lambda_k}{\lambda_j^2} \lesssim j^2.
\]

In the same manner, one shows that
\[
\sum_{k=1}^{\infty} \frac{\lambda_k}{|\lambda_j - \lambda_k|} \lesssim j \log j.
\]

Proof of Theorem 2.5. First note that due to the Gaussianity of \( X \), scores \( \eta_{k,i} \) and \( \eta_{k,j} \) are mutually independent for \( i \neq j \). Given independent standard Gaussian random variables \( X,Y \), the function \( XY - 1 \) is a two-dimensional second degree Hermite polynomial. If \( X = Y \), then \( X^2 - 1 \) is a univariate Hermite polynomial of second degree. We may now invoke Theorem 4 in [2]. The proof is based on the method of moments for partial sums of Hermite polynomials. In particular, using that \( \sup_{j \in \mathbb{N}} \sum_{k=0}^{\infty} \text{Cov}(\eta_{0,j}, \eta_{k,j})^2 < \infty \) it is shown via the Diagram formula that for any fixed \( p \in \mathbb{N} \)
\[
\frac{1}{\sqrt{n}} \max_{1 \leq i,j \leq \infty} \|\eta_{i,j}\|_p < \infty \quad \text{and} \quad \frac{\eta_{i,j}}{\sqrt{n}} \xrightarrow{w} \mathcal{N}(0, \sigma_{i,j}^2).
\]

Hence (A1) and the CLT for \( \hat{\lambda}_j \) follow. In order to prove the CLT for \( \hat{e}_j \) we proceed as follows. Denote with
\[
C_j = \sum_{k=1}^{\infty} e_k \frac{I_{k,j}}{\lambda_j - \lambda_k}, \quad C_{j,d} = \sum_{k=1}^{d} e_k \frac{I_{k,j}}{\lambda_j - \lambda_k} \quad \text{for} \quad d > j.
\]

Due to Theorem 2.3 and Lemma 3.3, we have that
\[
\sqrt{n} \left\| \frac{1}{\sqrt{\Lambda_j}} (\hat{e}_j - e_j - C_j) \right\|_{L^2} = o(1).
\]

It thus suffices to consider \( C_j \). Since \( \sum_{k > d} \lambda_k \to 0 \) as \( d \) increases, Lemma 3.3 implies that for any \( \delta > 0 \) there exists \( d_\delta \in \mathbb{N} \) such that
\[
\sqrt{n} \mathbb{E} [\|C_j - C_{j,d_\delta}\|_{L^2}] \leq \delta.
\]

(4.3)
It now suffices (cf. [24]) to establish that for any fixed $d \in \mathbb{N}$ (which includes the case $d = d_{\delta}$)

$$\sqrt{n}C_{j,d} \overset{w}{\to} \mathcal{N}(0, \Sigma_d), \tag{4.4}$$

where $\Sigma_d \in \mathbb{R}^d \times \mathbb{R}^d$ denotes the corresponding covariance matrix. But, since we have for $i_l \neq j_l$, $l \in \{1, 2\}$ that

$$E[\eta_{i_1,j_1}, \eta_{i_2,j_2}] = 0 \quad \text{if either } i_1 \neq i_2 \text{ or } j_1 \neq j_2,$$

we may apply Theorem 4 in [2] due to sup$_j \in \mathbb{N} \sum_{k=0}^{\infty} \text{Cov}(\eta_{0,j}, \eta_{k,j})^2 < \infty$, which gives (4.4). This completes the proof.

\[\Box\]

5. Proofs of Section 2.4

We need to introduce some further notation. To this end, we slightly reformulate our notion of weak dependence. In the sequel, $\{\epsilon_k\}_{k \in \mathbb{Z}} \in \mathcal{S}$ denotes an IID sequence in some measure space $\mathcal{S}$ and $\mathcal{F}_k = \sigma(\epsilon_j, j \leq k)$ the corresponding filtration. For $d \in \mathbb{N}$, we then consider the variables

$$U_{k,h} = H_h(\mathcal{F}_k), \quad k \in \mathbb{Z}, 1 \leq h \leq d,$$

where $H_h$ are measurable functions. Note that by considering different measure spaces $\mathcal{S}$, we can virtually model any spatial dependence structure we want, with the extreme cases where $U_{k,h} = U_{k,h+1}$ or $U_{k,h}$ and $U_{k,h+1}$ are independent. Compared to Section 2.4, this setup is notationally more convenient, and prevents us from the necessity of considering different sequences $\{\epsilon_{k,h}\}_{k \in \mathbb{Z}}$ for each coordinate $h$. As a measure of dependence, we then consider

$$\theta_{j,p} = \max_{1 \leq h \leq d} \|U_{j,h} - U'_{j,h}\|_p, \quad p \geq 1,$$

where $U_{k,h} = H_h(\mathcal{F}'_k)$, $\mathcal{F}'_k = \sigma(\ldots, \epsilon_{-1}, \epsilon'_0, \epsilon_1, \ldots, \epsilon_k)$, and $\{\epsilon'_k\}_{k \in \mathbb{Z}}$ is an independent copy of $\{\epsilon_k\}_{k \in \mathbb{Z}}$.

5.1. Gaussian approximation for weak dependence

In this section, a high dimensional Gaussian approximation result is established, which is a key ingredient in the proof of Theorem 2.8. This result may be of independent interest. Let $S_{n,h} = \sum_{k=1}^{n} U_{k,h}$, and denote with

$$T_d = \frac{1}{\sqrt{n}} \max_{1 \leq h \leq d} |S_{n,h}|, \quad T'_d = \max_{h \leq d} |Z_h|, \quad \tag{5.1}$$

where $\{Z_h\}_{1 \leq h \leq d}$ is a sequence of zero mean Gaussian random variables. We also formally introduce

$$\gamma_{i,j} = \lim_{n \to \infty} \frac{1}{n} E[S_{n,i}S_{n,j}].$$
existence is shown below in Lemma 5.6. We also put $\sigma^2_h = \gamma_{h,h}$. Throughout this section, we work under the following assumption.

**Assumption 5.1.** The sequence $\{U_{k,h}\}_{k \in \mathbb{Z}}$ is stationary for each $1 \leq h \leq d$, such that for $p > 2$ and $d \lesssim n^\delta$

(C1) $\mathbb{E}[U_{k,h}] = 0$ and $\theta_{j,p} \lesssim j^{-\epsilon}$ with $\epsilon > 3/2$,
(C2) $\delta < p/2 - 1$,
(C3) $\inf_h \sigma_h > 0$.

We then have the following Gaussian approximation result.

**Theorem 5.2.** Grant Assumption 5.1. Then

$$\sup_{x \in \mathbb{R}} \left| P(T_d \leq x) - P(T_d^\beta \leq x) \right| \lesssim n^{-C}, \quad C > 0,$$

where $\{Z_h\}_{1 \leq h \leq d}$ has the same covariance structure as $n^{-1/2}\{S_{n,h}\}_{1 \leq h \leq d}$. Alternatively, we may also choose $(\gamma_{i,j})_{1 \leq i,j \leq d}$ as covariance structure.

The proof of Theorem 5.2 is lengthy, and requires a number of preliminary lemmas. The general idea consists of a coupling argument, exploiting the Bernoulli structure of $U_{k,h}$. Employing a Fuk-Nagaev type inequality (Theorem 2 in [25]) as one of the major tools, we can subsequently reduce the problem to the IID case, where we can resort to the literature.

We first establish some additional notation. Let $K = n^i$, $L = n^l$ such that $n = KL$ and $0 < i, l < 1$. To simplify the discussion, we always assume that $K, L \in \mathbb{N}$. For each $1 \leq l \leq L$, let $\{\epsilon_k\}_{k \in \mathbb{Z}} \in S$ be mutually independent sequences of IID random variables. For $K(l - 1) < k \leq KL$, $1 \leq l \leq L$, denote with

$$(5.2) \quad V_{l,h}(m) = \sum_{k=K(l-1)+m}^{Kl} U_{k,h} + \sum_{k=K(l-1)+1}^{K(l-1)+m} U_{k,h}^{(K,\phi)}$$

and $V_{l,h}^\phi = V_{l,h}^\phi(1)$. The random variables $V_{l,h}^\phi$ play a key role in the proof of Theorem 5.2. Note in particular that $\{V_{l,h}^\phi\}_{1 \leq l \leq L}$ is IID by construction for each $h$. Finally, put $S_{L,h}(V) = \sum_{j=1}^{L} V_{j,h}$ and $S_{L,h}^\phi(V) = \sum_{j=1}^{L} V_{j,h}^\phi$, and note that $S_{n,h} = S_{L,h}(V)$. In the sequel, we make frequent use of the following lemma.

**Lemma 5.3.** Suppose that $\sum_{j=1}^{\infty} \theta_{j,p} < \infty$ for $p \geq 2$. Then

$$\max_{1 \leq h \leq d} \left\| \sum_{k=1}^{n} U_{k,h} \right\|_p \lesssim \sqrt{n}.$$
For the proof and variants of this result, see [35]. The next lemma controls the approximation error between $S_{L,h}(V)$ and $S^*_{L,h}(V)$.

**Lemma 5.4.** Grant Assumption 5.1. For any $K = n^\xi$ with $0 < \xi < 1$ there exists a $\delta > 0$ and a constant $C > 0$ such that

$$P(\left|S_{L,h}(V) - S^*_{L,h}(V)\right| \geq Cn^{1/2-\delta}) \lesssim n^{-\frac{p-2}{2} + p\delta}.$$

**Proof of Lemma 5.4.** Let $x_n = x\sqrt{n}$, $x > 0$. For $1 \leq m < K$ we have that

$$P\left(\left|S_{L,h}(V) - S^*_{L,h}(V)\right| \geq 2x_n\right) \leq P\left(\sum_{l=1}^{L} \sum_{k=K(l-1)+1}^{K(l-1)+m-1} U_{k,h} - U_{k,h}^{(K,\phi)} \geq x_n\right)$$

$$+ P\left(\left|\sum_{l=1}^{L} V_{l,h} - V_{l,h}^\phi(m)\right| \geq x_n\right).$$

Denote with $\alpha_{j,p} = (j^{p/2-1} \theta_{j,p})^{1/(p+1)}$ and $A = \sum_{j=1}^{\infty} \alpha_{j,p}$. Note that by (C1) we have

$$\alpha_{j,p} \lesssim j^{-\mathfrak{g}(p,c)}, \quad \text{where } \mathfrak{g}(p,c) = \frac{p(\epsilon - 1/2) + 1}{p + 1} > 1, \quad (5.3)$$

and thus $A < \infty$. Due to Theorem 2 in [25], there exist constants $C_{p,1}, C_{p,2} > 0$ such that

$$P\left(\left|\sum_{l=1}^{L} \sum_{k=K(l-1)+1}^{K(l-1)+m-1} U_{k,h} - U_{k,h}^{(K,\phi)} \geq x_n\right| \leq \frac{C_{1,p} L m^{\frac{1}{p}}}{x_n} + \sum_{j=1}^{\infty} \exp\left(-\frac{C_{p,2} \alpha_{j,p}^2 x_n}{A^2 L m \theta_{j,2}^2}\right)$$

$$+ \exp\left(-\frac{C_{p,2} x_n^2}{L m \|U_{k,h}\|^2_2}\right).$$

Setting $x = y\sqrt{LmA^{1+1/p}/\sqrt{n}}$, it follows that $\alpha_{j,p}^2 x_n^2/(A^2 L m \theta_{j,2}^2) \geq j^{1-2/p} y^2$ and hence

$$\exp\left(-\frac{C_{p,2} \alpha_{j,p}^2 x_n^2}{A^2 L m \theta_{j,2}^2}\right) \leq \exp\left(-C_{p,2} j^{-1-2/p} y^2\right).$$

Choosing $m$ such that $\sqrt{n}/\sqrt{Lm} = n^{2\delta}$ and $y = n^\delta$, $\delta > 0$, it follows that

$$P\left(\left|\sum_{l=1}^{L} \sum_{k=K(l-1)+1}^{K(l-1)+m-1} U_{k,h} - U_{k,h}^{(K,\phi)} \geq n^{1/2-\delta} A^{1+1/p}\right| \lesssim n^{-\frac{p-2}{2} + p\delta}. \quad (5.4)$$

Next, put $\Delta_{k,h}(U) = U_{k,h} - U_{k,h}^{(K,\phi)}$. By the triangle inequality, we have

$$\left\|\Delta_{k,h}(U) - \Delta_{k,h}(U)\right\|_p \lesssim 2(\theta_{k,p} \wedge \|\Delta_{k,h}(U)\|_p).$$
Let \((k)_K = k \mod K\). Then Theorem 1 in [34] yields that
\[
\max_{1 \leq h \leq d} \|\Delta_{k,h}(U)\|_p^2 = \max_{1 \leq h \leq d} \|U_{k,h} - U_{k,h}^{(K,\nu)}\|_p^2 \lesssim \sum_{j=(k)_K}^\infty \theta_{j,p}^2 \overset{\text{def}}{=} \Theta_{(k)_K,p}.
\]
Since clearly \(\Theta_{(k)_K,p}\) is monotone decreasing, we have \(\Theta_{(k)_K,p} \leq \Theta_{(m)_K,p}\) for \(m \leq k \leq K\). Combining this with the above, it follows that for \(m \leq (k)_K\) (since \(m = (m)_K\))
\[
\max_{1 \leq h \leq d} \|\Delta_{k,h}(U) - \Delta_{k,h}(U)^\nu\|_p \leq 2\left(\theta_{k,p} \wedge \sqrt{\Theta_{m,p}}\right) \overset{\text{def}}{=} \vartheta_{k,p}(m). \quad (5.5)
\]
Put \(\beta_{j,p}(m) = (j^{p/2-1} \vartheta_{j,p}(m))^{1/(p+1)}\) and \(B(m) = \sum_{j=1}^\infty \beta_{j,p}(m)\). Then another application of Theorem 2 in [25] yields that
\[
P\left(\left|\sum_{l=1}^L V_{l,h} - V_{l,h}^\nu(m)\right| \geq x_n\right) \leq C_{1,p} \frac{n}{x_n} + \sum_{j=1}^\infty \exp\left(-\frac{C_{p,2} \beta_{j,p}^2(m) x_n^2}{B^2(m) n \vartheta_{j,p}^2(m)}\right)
+ \exp\left(-\frac{C_{p,2} y_n^2}{\Theta_{m,p}}\right).
\]
Let \(y_n = n^{\delta} \sqrt{Lm}/\sqrt{n} = n^{-\delta}\). Arguing similarly as before, it follows (since \(m = (m)_K\))
\[
P\left(\left|\sum_{l=1}^L V_{l,h} - V_{l,h}^\nu(m)\right| \geq x_n\right) \lesssim \frac{n}{x_n} + \sum_{j=1}^\infty \exp\left(-\frac{C_{p,2} y_n^2}{B(m)^2}\right)
+ \exp\left(-\frac{C_{p,2} y_n^2}{\Theta_{m,p}}\right).
\]
Since \(\Theta_{m,p} \lesssim m^{-2\epsilon+1}\), we conclude
\[
B(m) \lesssim \sum_{j>M} \alpha_{j,p} + \sum_{j=1}^M \left(j^{p/2-1} m^{-p+1/2}\right)^{1/(p+1)} \lesssim M^{-\mathbb{B}(p,c)} + M^{2p-4} m^{-2p+2p \epsilon}.\]
Setting \(m \sim n^\nu, \nu > 0\), balancing the above and choosing \(\delta\) sufficiently small, we obtain
\[
\frac{y_n^2}{B(m)^2} \wedge \frac{y_n^2}{\Theta_{m,p}} \gtrsim n^\delta. \quad (5.6)
\]
This implies that
\[
P\left(\left|\sum_{l=1}^L V_{l,h} - V_{l,h}^\nu(m)\right| \geq n^{1/2-\delta} A^{1+1/p}\right) \lesssim n^{\frac{-p-2}{2} + \delta}.
\]
Note that by the above choice of \(m = n^\nu\) we require that \(L \sim n^{1-4\delta-\nu}\). Choosing \(\nu\) sufficiently close to 1, we can select \(\epsilon \leq 1\) arbitrarily close to 1, which completes the proof. \(\square\)
In the sequel, we also require the following result.

**Lemma 5.5.** Grant Assumption 5.1. Then

\[ P \left( \left| V_{l,h}^o \right| \geq \sqrt{K \log n} \right) \lesssim K^{1-p/2} (\log n)^p. \]

**Proof of Lemma 5.5.** Since \( V_{l,h}^o \overset{d}{=} V_{l,h} \), Theorem 2 in [25] and arguing similarly as in Lemma 5.4 yields

\[ P \left( \left| V_{l,h}^o \right| \geq y \sqrt{L} \right) \lesssim L^{1-p/2} y^p \]

\[ + \exp \left( - \frac{C_p,2 y^2}{\|U_{k,h}\|^2_2} \right). \]

Setting \( y = \log n \), the claim follows.

Next, we establish some useful results concerning the covariances \( \phi_{k,i,j} = E[U_0,i U_k,j] \).

**Lemma 5.6.** Grant Assumption 5.1. Then

(i) \( \sup_{i,j} |\phi_{k,i,j}| \lesssim k^{-c+1/2} \),

(ii) \( \sup_{i,j} \sum_{k=0}^{\infty} |\phi_{k,i,j}| < \infty \),

(iii) \( \gamma_{i,j} = \phi_{0,i,j} + 2 \sum_{k=1}^{\infty} \phi_{k,i,j} < \infty \),

(iv) \( \sum_{k,j=1}^{n} E[U_k,i U_{l,j}] = n \gamma_{i,j} - \sum_{k \in \mathbb{Z}} n \wedge |k| \phi_{k,i,j} \).

**Proof of Lemma 5.6.** Claims (iii) and (iv) are well-known in the literature, and follow from elementary computations from (ii), see for instance [20]. Since (i) implies (ii) due to \( c > 3/2 \), it suffices to establish (i). To this end, let \( U_{k,h}^* = H_k(F_k^*) \), where \( F_k^* = \sigma(\ldots, \xi'_{-1}, \xi'_0, \xi_1, \ldots, \xi_k) \). Since then \( E[U_{k,h}^* | \mathcal{F}_0] = E[U_{k,h}] = 0 \), Cauchy-Schwarz and Jensen inequality yield

\[ \left| E[U_{0,i} U_{k,j}] \right| = \left| E[U_{0,i} E[U_{k,j} | \mathcal{F}_0]] \right| \leq \|U_{0,i}\|_2 \|U_{k,j} - U_{k,j}^*\|_2. \]

Theorem 1 in [34] and (C1) then imply that

\[ \left| E[U_{0,i} U_{k,j}] \right| \lesssim \left( \sum_{l=k}^{\infty} \theta_{l,2}^2 \right)^{1/2} \lesssim k^{-c+1/2}. \]

For \( 1 \leq i, j \leq d \) denote with

\[ \gamma_{i,j}^{(n)} = \frac{1}{n} E[S_{n,i} S_{n,j}], \quad \gamma_{i,j}^{(o,n)} = \frac{1}{n} E[S_{L,i}(V) S_{L,j}(V)]. \]
Remark 5.7. Note that Lemma 5.6 (iv) yields that
\[|\gamma_{i,j} - \gamma_{i,j}^{(n)}| \lesssim \frac{1}{n} \sum_{j=1}^{n} j^{3/2-\varepsilon} + \sum_{j>n}^{\infty} j^{-\varepsilon+1/2} \lesssim n^{3/2-\varepsilon}.\]

Lemma 5.8. Grant Assumption 5.1. Then
\[\max_{1 \leq i,j \leq d} |\gamma_{i,j}^{(n)} - \gamma_{i,j}^{(\diamond,n)}| \lesssim n^{-1/2} L.\]

Remark 5.9. Note that we obtain from Remark 5.7 that
\[|\gamma_{i,j} - \gamma_{i,j}^{(\diamond,n)}| \lesssim n^{-1/2} L + n^{3/2-\varepsilon}.\]

Proof of Lemma 5.8. We have that
\[
\left| \mathbb{E}[S_{L,i}(V)S_{L,j}(V)] - \mathbb{E}[S_{L,i}^o(V)S_{L,j}^o(V)] \right| \leq \sum_{l=1}^{L} \|V_{i,j}^o - V_{i,j}\|_2 \|S_{L,j}^o(V)\|_2
+ \sum_{l=1}^{L} \|V_{i,i}^o - V_{i,i}\|_2 \|S_{L,i}(V)\|_2.
\]

By Lemma 5.3 and (C1) we have
\[\max_{1 \leq h \leq d} \|S_{L,h}^o(V)\|_2 \lesssim \sqrt{n} \quad \text{and} \quad \max_{1 \leq h \leq d} \|S_{L,h}(V)\|_2 \lesssim \sqrt{n}. \quad (5.7)\]

Using the triangle inequality and Theorem 1 in [34], it follows that
\[\max_{1 \leq h \leq d} \sum_{l=1}^{L} \|V_{i,h}^o - V_{i,h}\|_2 \lesssim \max_{1 \leq h \leq d} L \sum_{k=1}^{\infty} \|U_{k,h} - U_{k,h}^o\|_2
\lesssim \sum_{k=1}^{\infty} \sqrt{\sum_{j \geq k} \theta_{j,2}^2} \lesssim L \sum_{k=1}^{\infty} j^{-\varepsilon+1/2} \lesssim L. \quad (5.8)\]

Hence combining (5.7) and (5.8) we obtain
\[\max_{1 \leq i,j \leq d} \left| \gamma_{i,j}^{(n)} - \gamma_{i,j}^{(\diamond,n)} \right| \lesssim n^{-1/2} L. \quad (5.9)\]

Next, we state some Gaussian approximation results. To this end, we require the following condition. For \(\varepsilon, u(\varepsilon) > 0\) we have
\[P\left( \max_{1 \leq h \leq d} \max_{1 \leq i \leq L} |V_{i,h}^o| \geq \sqrt{Ku(\varepsilon)} \right) \leq \varepsilon. \quad (5.10)\]
Denote with
\[ T_{L,d}^0 = \frac{1}{\sqrt{n}} \max_{1 \leq h \leq d} |S_{L,h}^0(V)|, \quad T_d^{Z,0} = \max_{1 \leq h \leq d} |Z_h^0|, \]
where \( \{ Z_h^0 \}_{1 \leq h \leq d} \) is a zero mean Gaussian sequence with covariance structure 
\[ \Sigma_d^{(o,n)} = (\gamma_{i,j}^{X,Y})_{1 \leq i,j \leq d}. \]
We have the following Gaussian approximation result, which is an adaptation of Theorem 2.2 in [9].

**Lemma 5.10.** Assume the validity of (5.10) and that

(i) \( K^{-1/2} \min 1 \leq h \leq d \min_{1 \leq l \leq L} \| V_{i,l}^0 \|_2 > 0, \)
(ii) \( K^{-1/2} \max_{1 \leq h \leq d} \max_{1 \leq l \leq L} \| V_{i,l}^0 \|_4 < \infty. \)

Then it holds that
\[
\sup_{x \in \mathbb{R}} |P(T_{L,d}^0 \leq x) - P(T_d^{Z,0} \leq x)| 
\leq L^{-1/8} \left( \log(dL/\varepsilon) \right)^{7/8} + L^{-1/2} \left( \log(dL/\varepsilon) \right)^{3/2} u(\varepsilon) + \varepsilon.
\]

We also require the following two results, which are Lemmas 2.1 and 3.1 in [9], slightly adapted for our purpose.

**Lemma 5.11.** Let \( \{ X_h \}_{1 \leq h \leq d} \) and \( \{ Y_h \}_{1 \leq h \leq d} \) be zero mean Gaussian sequences, and denote with \( \gamma_{i,j}^{X,Y} \) the corresponding covariances for \( 1 \leq i, j \leq d \).
If \( 0 < \inf_{h} \gamma_{h,h}^{X,Y} \leq \sup_{h} \gamma_{h,h}^{X,Y} < \infty, \) then
\[
\sup_{x \in \mathbb{R}} |P(\max_{1 \leq h \leq d} |X_h| \leq x) - P(\max_{1 \leq h \leq d} |Y_h| \leq x)| 
\lesssim \delta^{1/3} \left( 1 \vee \log(d/\delta) \right)^{2/3},
\]
where \( \delta = \max_{1 \leq i,j \leq d} |\gamma_{i,j}^{X} - \gamma_{i,j}^{Y}|. \)

**Lemma 5.12.** Let \( \{ X_h \}_{1 \leq h \leq d} \) be a zero mean Gaussian sequence, and denote with \( \gamma_{h,h}^{X} \) the corresponding covariances for \( 1 \leq i, j \leq d \).
If \( 0 < \inf_{h} \gamma_{h,h}^{X} \leq \sup_{h} \gamma_{h,h}^{X} < \infty, \) then
\[
\sup_{x \in \mathbb{R}} P(\max_{1 \leq h \leq d} |X_h - \delta| \leq x) 
\lesssim \delta^{1/3} \left( 1 \vee \log(d/\delta) \right)^{2/3},
\]

We are now ready to give the proof of Theorem 5.2.

**Proof of Theorem 5.2.** First note that by Lemma 5.4 and Boole's inequality we have
\[
P(\max_{1 \leq h \leq d} |S_{L,h}(V) - S_{L,h}^0(V)| \geq C_1 n^{1/2 - \delta}) \lesssim d n^{-1/2 + \delta} + d p.
\]
Since \( d \lesssim n^b \) we obtain from (C2) that
\[
P(\max_{1 \leq h \leq d} |S_{L,h}(V) - S_{L,h}^0(V)| \geq C_1 n^{1/2 - \delta}) \lesssim n^{-C_2}, \quad C_2 > 0. \quad (5.11)
\]
Employing this bound, we get that
\[ P(T_d \leq x) \leq P(T^\circ_{L,d} \leq x + C_1 n^{-\delta}) + O(n^{-C_2}). \]

In the same manner one obtains a lower bound, hence
\[
P(T^\circ_{L,d} \leq x - C_1 n^{-\delta}) - O(n^{-C_2}) \leq P(T_d \leq x) \\
\leq P(T^\circ_{L,d} \leq x + C_1 n^{-\delta}) + O(n^{-C_2}). \tag{5.12}
\]

Next, we apply Lemma 5.10 to \( T^\circ_{L,d} \). To this end, we need to verify its conditions.

Note that by the independence of \( V^\circ_{l,h} \), we have that
\[ \gamma^{(\circ,n)}_{h,h} = \frac{1}{LK} \sum_{l=1}^{L} \| V^\circ_{l,h} \|_2^2 = \frac{1}{K} \| V^\circ_{1,h} \|_2^2. \]

Hence we deduce from Lemma 5.6, Lemma 5.8, Remark 5.9 and (C3) that
\[ K^{-1} \| V^\circ_{1,h} \|_2^2 \geq \gamma^{(\circ,n)}_{h,h} - \sigma(1) \geq \sigma^2 - \sigma(1) > 0, \]
uniformly in \( h \), and thus (i) holds. Next we verify (ii). This, however, readily follows from Lemma 3.12 and (C1). Finally, we need to establish (5.10). Set \( u(\varepsilon) = (\log n)^2 \). Using Boole's inequality and Lemma 5.5 gives
\[
P \left( \max_{1 \leq h \leq d} \max_{1 \leq l \leq L} | V^\circ_{l,h} | \geq \sqrt{Ku(\varepsilon)} \right) \\
\leq \sum_{h=1}^{d} \sum_{l=1}^{L} P \left( | V^\circ_{l,h} | \geq \sqrt{Ku(\varepsilon)} \right) \lesssim dLK^{-\frac{3}{2}} (\log n)^p.
\]

By (C2) and choosing \( \varepsilon \) sufficiently close to 1, we get that
\[
P \left( \max_{1 \leq h \leq d} \max_{1 \leq l \leq L} | V^\circ_{l,h} | \geq \sqrt{Ku(\varepsilon)} \right) \lesssim n^{-C_3}, \quad C_3,
\]
and (5.10) holds with \( \varepsilon \sim n^{-C_3} \). Since \( L \sim n^l \) with \( l > 0 \) due to \( \varepsilon < 1 \), Lemma 5.10 yields that
\[
\sup_{x \in \mathbb{R}} | P(T^\circ_{L,d} \leq x) - P(T_d^\circ \leq x) | \lesssim n^{-C_4}, \quad C_4 > 0. \tag{5.13}
\]

Combining this with (5.12), we deduce that
\[
P(Z^\circ_d \leq x - C_1 n^{-\delta}) - O(n^{-C_2}) \leq P(T_d \leq x) \\
\leq P(Z^\circ_d \leq x + C_1 n^{-\delta}) + O(n^{-C_2}). \tag{5.14}
\]

Next, since \( \log d \lesssim \log n \), Lemma 5.12 yields that
\[
\sup_{x \in \mathbb{R}} | P(Z^\circ_d \leq x - C_1 n^{-\delta}) - P(Z^\circ_d \leq x) | \lesssim n^{-\delta} \sqrt{\log n}. \tag{5.15}
\]
In addition, by Remark 5.9
\[
\max_{1 \leq i, j \leq d} |\gamma_{i,j}^{(s,n)} - \gamma_{i,j}| \lesssim n^{-\frac{1}{2}}L + n^{\frac{2}{3} - \epsilon} \lesssim n^{-C_6}, \quad C_6 > 0.
\]
Hence an application of Lemma 5.11 yields
\[
\sup_{x \in \mathbb{R}} |P(Z_d^\circ \leq x) - P(Z_d \leq x)| \lesssim n^{-C_6}, \quad C_6 > 0.
\] (5.16)

5.2. Proofs of Section 2.4

Proof of Theorem 2.8. Denote with
\[
T_{J_n} = \frac{1}{\sqrt{n}} \max_{1 \leq h \leq J_n} \frac{\sum_{k=1}^{n} (\eta^2_{k,h} - 1)}{\sigma_{\lambda,h}}.
\]
We first show that we may apply Theorem 5.2 to \(T_{J_n}^\eta\). To this end, we need to verify Assumption 5.1. Observe that (B2) implies \(\|\eta_{k,h}\|_{p^*} < \infty\) (cf. [34]). Moreover, using \(a^2 - b^2 = (a - b)(a + b)\), it follows from Cauchy-Schwarz
\[
\|\eta_{k,h}^2 - (\eta_{k,h}^2)'\|_{2p^*} \leq 2\|\eta_{k,h} - \eta_{k,h}'\|_{4p^*}\|\eta_{k,h}\|_{4p^*} \lesssim \Omega_k(4p^*) \lesssim k^{-b}.
\]
Since \(b > 3/2\) by (B2), (C1) follows. Next, note that (B1) implies that \(J_n^+ \lesssim n^{p(a-b)}\). Since \(p^*/4 - 1 > p2p^* + 1 > pa\) (recall \(0 < a < 1\)), (C2) holds. Finally, (B3) gives (C3), hence Assumption 5.1 is verified. We proceed with the proof. Note that
\[
I_{j,n} = \frac{\lambda_n}{n} \sum_{k=1}^{n} (\eta_{k,j}^2 - 1), \quad j \in \mathbb{N}.
\]
Introduce the set
\[
\mathcal{M} = \left\{ \max_{1 \leq j \leq J_n^+} \lambda_n^{-1} |\lambda - \lambda_j - I_{j,n}| \geq n^{-1/2 - \delta/2} \right\}.
\]
Then Markov’s inequality together with Theorem 2.2 and (B1) yield
\[
P(\mathcal{M}^c) \lesssim n^{-p_d} \lesssim n^{-C_1}, \quad C_1 > 0.
\] (5.17)
Due to Theorem 5.2 and the above, we have the inequalities
\[
P(T_{J_n}^\lambda \leq x) \leq P(T_{J_n}^\eta \leq x + n^{-\delta/2}) + P(\mathcal{M}^c)
\leq P(T_{J_n}^{Z_n} \leq x + n^{-\delta/2}) + O(n^{-C_2}), \quad C_2 > 0,
\]
where $T^{Z_n}_{J_n^+}$ is as in (2.16). An application of Lemma 5.12 yields that this is further bounded by
\[ P(T^{Z_n}_{J_n^+} \leq x) \leq P(T^{Z_n}_{J_n^+} \leq x) + \mathcal{O}(n^{-C_2} + n^{-\delta/2} \log n). \]
In the same manner, we obtain a lower bound, hence
\[ \sup_{x \in \mathbb{R}} |P(T^{Z_n}_{J_n^+} \leq x) - P(T^{Z_n}_{J_n^+} \leq x)| \lesssim n^{-C_3}, \quad C_3 > 0, \quad (5.18) \]
which completes the proof.

**Proof of Corollary 2.9.** Due to Theorem 2.8, it suffices to show that
\[ P(T^{Z_n}_{J_n^+} \leq u,_{J_n^+}(z)) \rightarrow \exp(-e^{-z}). \]
This, however, follows from Theorem 14 and Theorem 1 in [17].

**Proof of Theorem 2.11.** We only show (b), claim (a) follows in the same manner.
An application of Pisier’s inequality (cf. Lemma 3.12) and Proposition 2.6 give
\[ \left\| \max_{1 \leq j \leq J_n^+} |I_{j,n}^{-1}| \right\|_p \leq n^{-1/2} f^{1/p} \max_{1 \leq j \leq J_n^+} \left\| n^{-1/2} \eta_{j,n} \right\|_p \lesssim n^{-1/2} f^{1/p}. \]
Due to Theorem 2.2 and condition $(J_n^+)^{1/p} n^{-1/2} = \mathcal{O}(1)$ we get
\[ \left\| \max_{1 \leq j \leq J_n^+} \lambda_j^{-1} |\hat{\lambda}_j - \lambda_j| \right\|_p \lesssim \delta_{n,1} \quad (5.19) \]
for some $\delta_n, 1 \rightarrow 0$. Let $\delta_n, 2 \rightarrow 0$ with $\delta_n, 1 = \mathcal{O}(\delta_n, 2)$ and denote with
\[ \mathcal{M}_{\delta_n, 2} = \left\{ \max_{1 \leq j \leq J_n^+} \lambda_j^{-1} |\hat{\lambda}_j - \lambda_j| \leq \delta_n, 2 \right\}. \]
Then Markov’s inequality and (5.19) yield that $P(\mathcal{M}_{\delta_n, 2}^c) \rightarrow 0$. Next, observe that
\[ \left\| \max_{1 \leq j \leq J_n^+} \sum_{j=1}^J \left( \frac{1}{\lambda_j} - \frac{1}{\hat{\lambda}_j} \right) 1_{\mathcal{M}_{\delta_n, 2}} \right\|_1 \leq \sum_{j=1}^J \left( \left\| \frac{\hat{\lambda}_j - \lambda_j}{\lambda_j^2} \right\|_1 + \left\| \frac{(\hat{\lambda}_j - \lambda_j)^2}{\lambda_j^2} \right\|_1 \right) 1_{\mathcal{M}_{\delta_n, 2}}. \]
By construction of $\mathcal{M}_{\delta_n, 2}$, Corollary 2.4 (with Proposition 2.6) yields
\[ \left\| \frac{(\hat{\lambda}_j - \lambda_j)^2}{\lambda_j^2} 1_{\mathcal{M}_{\delta_n, 2}} \right\|_1 \lesssim \frac{1}{n \lambda_j} \quad \text{and} \quad \left\| \frac{\hat{\lambda}_j - \lambda_j}{\lambda_j^2} \right\|_1 \lesssim \frac{1}{\sqrt{n \lambda_j}}, \]
which completes the proof.
References

[1] T.W. Anderson. Asymptotic theory for principal component analysis. *The Annals of Mathematical Statistics*, 34(1):122–148, 03 1963.

[2] M. A. Arcones. Limit theorems for nonlinear functionals of a stationary Gaussian sequence of vectors. *Ann. Probab.*, 22(4):2242–2274, 1994.

[3] Z. Bai and G. H. Golub. Bounds for the trace of the inverse and the determinant of symmetric positive definite matrices, 1996.

[4] R. Bhatia, Ch. Davis, and A. McIntosh. Perturbation of spectral subspaces and solution of linear operator equations. *Linear Algebra Appl.*, 52/53:45–67, 1983.

[5] D. Bosq. *Linear processes in function spaces*, volume 149 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 2000. Theory and applications.

[6] P. Breuer and P. Major. Central limit theorems for non-linear functionals of gaussian fields. *Journal of Multivariate Analysis*, 13(3):425–441, September 1983.

[7] H. Cardot, C. Crambes, A. Kneip, and P. Sarda. Smoothing splines estimators in functional linear regression with errors-in-variables. *Comput. Statist. Data Anal.*, 51(10):4832–4848, 2007.

[8] H. Cardot, A. Mas, and P. Sarda. CLT in functional linear regression models. *Probab. Theory Related Fields*, 138(3-4):325–361, 2007.

[9] V. Chernozhukov, D. Chetverikov, and K. Kato. Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *The Annals of Statistics*, 41(6):2786–2819, 12 2013.

[10] C. Crambes and A. Mas. Asymptotics of prediction in functional linear regression with functional outputs. *Bernoulli*, 19(5B):2627–2651, 11 2013.

[11] J. Dauxois, A. Pousse, and Y. Romain. Asymptotic theory for the principal component analysis of a vector random function: some applications to statistical inference. *J. Multivariate Anal.*, 12(1):136–154, 1982.

[12] J. Dedecker and C. Prieur. New dependence coefficients. Examples and applications to statistics. *Probab. Theory Related Fields*, 132(2):203–236, 2005.

[13] I. Gohberg, S. Goldberg, and M.A. Kaashoek. *Basic classes of linear operators*. Birkhäuser Verlag, Basel, 2003.

[14] P. Hall and J.L. Horowitz. Methodology and convergence rates for functional linear regression. *The Annals of Statistics*, 35(1):70–91, 02 2007.

[15] P. Hall and M. Hosseini-Nasab. On properties of functional principal components analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):109–126, 2006.

[16] P. Hall and M. Hosseini-Nasab. Theory for high-order bounds in functional principal components analysis. *Math. Proc. Cambridge Philos. Soc.*, 146(1):225–256, 2009.

[17] X. Han and W. B. Wu. Portmanteau test and simultaneous inference for serial covariances. *Stat. Sin.*, 24(2):577–599, 2014.

[18] S. Hörmann and P. Kokoszka. Weakly dependent functional data. *Ann. Statist.*, 38(3):1845–1884, 2010.
[19] L. Horváth and P. Kokoszka. Inference for functional data with applications. Springer Series in Statistics. Springer, New York, 2012.
[20] M. Jirak. Change-point analysis in increasing dimension. *J. Multivariate Anal.*, 111:136–159, 2012.
[21] I.T. Jolliffe. *Principal component analysis*. Springer Series in Statistics. Springer-Verlag, New York, second edition, 2002.
[22] M. R. Leadbetter. On extreme values in stationary sequences. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 28:289–303, 1973/74.
[23] M. R. Leadbetter. Extremes and local dependence in stationary sequences. *Probability Theory and Related Fields*, 65:291–306, 1983. 10.1007/BF00532484.
[24] M. Ledoux and M. Talagrand. *Probability in Banach spaces*. Classics in Mathematics. Springer-Verlag, Berlin, 2011. Isoperimetry and processes, Reprint of the 1991 edition.
[25] W. Liu, H. Xiao, and W.B. Wu. Probability and moment inequalities under dependence. *Statist. Sinica*, 23(3):1257–1272, 2013.
[26] A. Mas and F. Ruymgaart. High-dimensional principal projections. *Complex Analysis and Operator Theory*, pages 1–29, 2014.
[27] A. Meister. Asymptotic equivalence of functional linear regression and a white noise inverse problem. *Ann. Statist.*, 39(3):1471–1495, 2011.
[28] F. Merlevède, M. Peligrad, and S. Utev. Sharp conditions for the clt of linear processes in a hilbert space. *Journal of Theoretical Probability*, 10(3):681–693, 1997.
[29] V.M. Panaretos and S. Tavakoli. Fourier analysis of stationary time series in function space. *The Annals of Statistics*, 41(2):568–603, 04 2013.
[30] M. Peligrad and S. Utev. A new maximal inequality and invariance principle for stationary sequences. *Ann. Probab.*, 33(2):798–815, 2005.
[31] M. Peligrad and S. Utev. Central limit theorem for stationary linear processes. *Ann. Probab.*, 34(4):1608–1622, 2006.
[32] M. Peligrad, S. Utev, and W. B. Wu. A maximal $L_p$-inequality for stationary sequences and its applications. *Proc. Amer. Math. Soc.*, 135(2):541–550 (electronic), 2007.
[33] J.O. Ramsay and B.W. Silverman. *Functional data analysis*. Springer Series in Statistics. Springer, New York, second edition, 2005.
[34] W. B. Wu. Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences USA.*, 102:14150–14154, 2005.
[35] W. B. Wu. Strong invariance principles for dependent random variables. *Ann. Probab.*, 35(6):2294–2320, 2007.