A note on quadratic forms of stationary functional time series under mild conditions

Anne van Delft*
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum, Germany

November 25, 2022

Abstract

We study distributional properties of a quadratic form of a stationary functional time series under mild moment conditions. As an important application, we obtain consistency rates of estimators of spectral density operators and prove joint weak convergence to a vector of complex Gaussian random operators. Weak convergence is established based on an approximation of the form via transforms of Hilbert-valued martingale difference sequences. As a side-result, the distributional properties of the long-run covariance operator are established.

keywords: functional data, time series, spectral analysis, martingales

AMS subject classification: Primary: 62HG99, 60G10, 62M15, Secondary 62M10.

1 Introduction

The subject of this paper is quadratic forms of a stationary time series \( \{X_t : t \in \mathbb{Z}\} \) with paths in some function space \( \mathcal{H} \). From a technical perspective, we shall adhere to existing literature and assume \( \mathcal{H} \) is a separable Hilbert space. Each realization is therefore a function. Such Functional time series are of growing interest due to the fact that many processes are almost continuously measured on their domain of definition. While quadratic forms of Euclidean-valued random variables have received considerable attention and have been studied under various dependence conditions [see e.g., 21, 29, 24, 2, 37, 25, and references therein], this is not so much the case for quadratic forms of function-valued random variables. Yet, they do arise naturally in a variety of inference problems. A quadratic form statistic of a functional time series can be given by

\[
\hat{\mathcal{Q}}_T = \sum_{s,t=1}^{T} \Phi_{T,s,t}(X_s \otimes X_t)
\]

where \( \{\Phi_{T,s,t}\}_{s,t=1}^{T} \) is a sequence of bounded linear operators, which will vary depending on the application. Important applications in which statistics of the form (1) arise are those that concern the consistent estimation of the second order characteristics of the process. This is especially relevant for functional data because the smoothness properties of the random functions are

*anne.vandelft@rub.de
encoded in the second order structure and are key in obtaining optimal finite-dimensional representations. For example, if we denote \( I_{H^pH} \) the identity operator on the tensor product space \( H \otimes H \), then the specification \( \Phi_{T,t,s} = \frac{1}{T} \sum_{t=1}^{T} I_{H^pH} \) yields the sample covariance operator. In the case of i.i.d. functional data, this object captures the full second order structure and its eigen-decomposition plays a central role in the reduction to finite dimension of the process’s properties, e.g., via the Karhunen-Loève representation if \( H = \mathbb{L}^2 \). Not surprisingly, the sample covariance operator received considerable attention in the corresponding line of literature [e.g., 14, 9, 35, but also in case of linear processes [see among others 4, 10, 18, and references therein]. However, when there is serial correlation between observations the covariance operator clearly does not capture the full dynamics. For dependent functional data, a more meaningful object is therefore the spectral density operator

\[
\mathcal{F}^{(\lambda)} = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} C_h e^{-i\lambda h}, \quad \lambda \in (-\pi, \pi],
\]

where \( C_h \) is the \( h \)-lag covariance operator of the process \( X := \{X_t : t \in \mathbb{Z} \} \). As an estimator for \( \mathcal{F}^{(\lambda)} \) of a process \( X \) with mean function \( \mu \), one can consider

\[
\hat{\mathcal{F}}^{(\lambda)} = \frac{1}{2\pi T} \sum_{s,t=1}^{T} \left( \phi_{T,t,s} \pi \right) \left( (X_s - \mu) \otimes (X_t - \mu) \right), \quad \lambda \in (-\pi, \pi],
\]

which simply corresponds to the quadratic form in \( \mathcal{F} \) with \( \Phi_{T,t,s}^{(\lambda)} = 2\pi T \phi_{T,t,s}^{(\lambda)} I_{H^pH} \). Here \( \phi(\cdot) \) is an even, bounded function on \( \mathbb{R} \) that is continuous at zero and \( b_T \) is a bandwidth parameter converging to zero at a rate such that \( b_T T \to \infty \) as the sample size \( T \) tends to infinity. The properties of this estimator and its relation to the smoothed periodogram operator are discussed in detail in Section 4. For \( \lambda = 0 \), \( 2\pi \hat{\mathcal{F}}^{(0)} \) is an estimator of the long-run covariance operator. Because it arises as the limiting covariance operator of the sample mean function, properties of the long-run covariance operator have been studied in several contexts within the framework of \( L^m_p \)-approximability [see e.g., 13, 16, 1].

Frequency domain analysis of functional time series, i.e., the case \( \lambda \neq 0 \), has received considerably less attention than time domain analysis. Yet, not only does frequency domain analysis (and hence the spectral density operator) arise naturally in various applications, it allows in particular to capture the full second order dynamics of dependent functional data. It can therefore be seen to take on a similar role for dependent functional data as the covariance operator takes on in the case of i.i.d. functional data. In fact, it allows to reduce the uncountably infinite variation to a countably infinite space in an optimal manner via a dynamic Karhunen-Loève representation provided the function space is sufficiently smooth. Moreover, frequency domain based inference methods enable powerful nonparametric tools for hypothesis testing. Because of its relevance for dependent functional data, estimators of \( \mathcal{F}^{(\lambda)} \) in the context of \( L^m_p \)-dependence as well as under functional cumulant-mixing conditions were introduced earlier this decade. Under \( L^m_p \) approximability, [13] considered dynamic principal components for stationary functional time series and obtained a consistency result for a lag window estimator. Under cumulant-mixing conditions, [30] derived consistency and asymptotic normality of a smoothed periodogram operator estimator. Estimation and distributional properties of an estimator for a time-varying spectral density operator were derived in [11], who introduced a framework for locally stationary functional time series. Note that all of the aforementioned estimators can be written in the form \( \mathcal{F} \). It is worth mentioning that these works have paved the way for frequency domain-based inference of functional time series, leading to an upsurge in the available literature in the past few years [see e.g., 41, 17, 26, 33, 12, 23, and references therein].
Cumulant tensors and spectral cumulant tensors can be shown to form Fourier pairs, provided appropriate summability conditions are satisfied. The consideration of functional cumulant mixing conditions as in [30] can therefore to some extent be seen to provide a natural framework for the derivation of sampling properties. Yet, the central limit theorem and consistency result as derived in [30] rely on existence of all moments and summability conditions of the cumulant tensors. In certain applications such required summability conditions might be too strong and worthwhile to be relaxed. To the best of the author’s knowledge, there is currently no CLT available under $L^p_m$-dependence and the consistency rate available in this setting [13] is sub-optimal compared to the one derived under cumulant mixing conditions in [30].

Broadly speaking, the goal of this paper is therefore twofold. We aim to derive a general central limit theorem for quadratic forms of stationary functional time series under sharp moment conditions. At the same time, we aim to obtain the best possible convergence and consistency rates for the aforementioned applications. It is worth mentioning that our conditions on the dependence structure are also weaker than those considered within the $L^p_m$-dependence framework. Underlying our approach is an approximation of the quadratic form with a Hilbertian-valued martingale process. To construct this process, we shall use a martingale approximation of the quadratic form. The idea to approximate a normalized partial sum process via a related martingale process was first put forward by [20]. [38] introduced this approach to derive distributional theorems for a set of quadratic forms as in (1). In Section 4, we focus on the estimation of the spectral density operator and long-run covariance as particular applications. More specifically, the central limit theorem for quadratic forms of stationary functional time series under sharp moment conditions. At the same time, we aim to obtain the best possible convergence and consistency rates for the aforementioned applications. It is worth mentioning that our conditions on the dependence structure are also weaker than those considered with within the $L^p_m$-dependence framework. Underlying our approach is an approximation of the quadratic form with a Hilbertian-valued martingale process. To construct this process, we shall use a martingale approximation of the quadratic form. The idea to approximate a normalized partial sum process via a related martingale process was first put forward by [20]. [38] introduced this approach to derive distributional

The structure of this note is as follows. In Section 2, we introduce necessary notation and conditions. In Section 3, we explain the approach in more detail and provide a joint central limit theorem for a set of quadratic forms as in (1). In Section 4, we focus on the estimation of the spectral density operator and long-run covariance as particular applications. More specifically, consistency rates and distributional properties are established. In Section 5, we relate the mild assumptions made in this paper to functional cumulant mixing conditions. Various technical results and proofs are relegated to the Appendix.

2 Framework

Throughout this paper, we will focus on random variables taking values in some separable Hilbert space, say $H$. For elements of $H$, we shall denote the inner product by $\langle \cdot, \cdot \rangle$ and the induced norm by $\| \cdot \|_H$. We let $H_1 \otimes H_2$ denote the Hilbert tensor product of the Hilbert spaces $(H_j, \langle \cdot, \cdot \rangle_{H_j})_{j=1,2}$. This Hilbert space can be constructed from the algebraic tensor product $H_1 \otimes_{alg} H_2$ together with a bilinear map $\psi : H_1 \times H_2 \rightarrow H_1 \otimes_{alg} H_2$ that satisfies $\langle \psi(x_1, x_2), \psi(y_1, y_2) \rangle = \langle x_1, y_1 \rangle_{H_1} \langle x_2, y_2 \rangle_{H_2}$ for $x_1, y_1 \in H_1$ and $x_2, y_2 \in H_2$ and then taking the completion with respect to the induced norm [see e.g., [22] for details]. By the associative law, this can be extended to construct a Hilbert space, $\bigotimes_{i=1}^n H_i$, from the algebraic $n$-fold tensor product. Next, we denote by $S_{\infty}(H_1, H_2)$ the Banach space of bounded linear operators $A : H_1 \rightarrow H_2$ equipped with the operator norm $\|A\|_{\infty} = \sup_{\|g\|_{H_1} \leq 1} \|Ag\|_{H_2}, g \in H_1$. An operator $A \in S_{\infty}(H) := S_{\infty}(H, H)$ is called non-negative definite if $\langle Ag, g \rangle \geq 0$ for all $g \in H$. It is called self-adjoint if $\langle Af, g \rangle = \langle f, A^* g \rangle = \langle f, Ag \rangle$ for all $f, g \in H$, where $A^*$ denotes the adjoint of $A$. The conjugate operator of $A$ can be defined as $\overline{Ag} = (\overline{A} \overline{g})$, where $\overline{g}$ denotes the complex conjugate of $g \in H$. For $A, B, C \in S_{\infty}(H)$ we define the Kronecker tensor product $(A \otimes B) C = ACB^\dagger$, while the transpose Kronecker tensor product is given by $(A \otimes B) C = (A \otimes B)^\dagger C$. A bounded linear operator $A : H_1 \rightarrow H_2$ belongs to the class of...
Hilbert-Schmidt operators, denoted by \( S_2(H_1, H_2) \), if it has finite Hilbert-Schmidt norm \( \|A\|_2 := (\sum_{i=1}^{\infty} \| A(\chi_i) \|_{H_2}^2)^{1/2} \), where \( \{\chi_i\}_{i=1}^{\infty} \) is an arbitrary orthonormal basis of \( H_1 \). We say \( A \) is trace class and denote \( A \in S_1(H) \) if \( A \) has finite trace class norm \( \|A\|_1 = \sum_{i=1}^{\infty} \langle A(\chi_i), \chi_i \rangle \). The space \( (S_2(H_1, H_2), \|\cdot\|_2) \) is a Hilbert space with inner product \( \langle A, B \rangle_S = \text{Tr}(AB^*) = \sum_{i=1}^{\infty} \langle A(\chi_i), B(\chi_i) \rangle_{H_2}, A, B \in S_2(H_1, H_2) \). For \( f, g, v \in H \), define the tensor product \( f \otimes g : H \otimes H \to H \) as the bounded linear operator \( (f \otimes g)v = \langle v, g \rangle f \). The mapping \( \mathcal{T} : H \otimes H \to S_2(H) \) defined by the linear extension of \( \mathcal{T}(f \otimes g) = f \otimes g \) is an isometric isomorphism.

A \( H \)-valued random element \( X \) over a probability space \( (\Omega, \mathcal{F}, \mathbb{P}) \) is a strongly measurable function \( X : (\Omega, \mathcal{F}, \mathbb{P}) \to (H, \mathcal{B}) \), where \( \mathcal{B} \) denotes the \( \sigma \)-algebra on \( H \). We denote \( X \in \mathcal{L}_H^p \) if \( \| X \|_{\mathcal{L}_H^p} := (\mathbb{E}\| X \|^p_H)^{1/p} < \infty \). Observe that \( \mathcal{L}_H^p \) is a Banach space w.r.t. the norm \( \| \cdot \|_H \) and for \( p = 2 \) it is a Hilbert space with equipped with the inner product \( \langle \cdot, \cdot \rangle \). For \( X \in \mathcal{L}_H^1(\Omega, \mathcal{A}, \mathbb{P}) \) and \( \mathcal{A}_0 \) a sub-algebra of \( \mathcal{A} \), we define the conditional expectation \( \mathbb{E}[\cdot | \mathcal{A}_0] : \mathcal{L}_H^1(\Omega, \mathcal{A}, \mathbb{P}) \to \mathcal{L}_H^1(\Omega, \mathcal{A}_0, \mathbb{P}) \) to be the mapping such that

\[
\int_A \mathbb{E}[X|\mathcal{A}_0]d\mathbb{P} = \int_A Xd\mathbb{P} \quad A \in \mathcal{A}_0,
\]

where the expectations should all be understood in the sense of a Bochner integral. Note that classical properties of conditional expectations remain valid in the context of separable Hilbert spaces. The cross-covariance operator between two zero-mean elements \( X, Y \in \mathcal{L}_H^2 \) is given by \( \text{Cov}(X, Y) = \mathbb{E}(X \otimes Y) \) and belongs to \( S_1(H) \). We note in particular that \( \| X \|_{L_2}^2 = \text{Tr}(\text{Var}(X \otimes X)) \), where \( \text{Var}(X) = \text{Cov}(X, X) \). For a filtration \( \{\mathcal{G}_t\} \) of \( \sigma \)-algebras of \( \mathcal{A} \), we shall make extensive use of projection operators defined by

\[
P_k = \mathbb{E}[\cdot | \mathcal{G}_k] - \mathbb{E}[\cdot | \mathcal{G}_{k-1}], \quad k \in \mathbb{Z}
\]

which are linear operators on \( \mathcal{L}_H^1 \) and are strongly orthogonal elements in \( \mathcal{L}_H^2 \), i.e.,

\[
\text{Cov}(P_k(X_1), P_j(X_2)) = 0_H \quad \forall X_1, X_2 \in \mathcal{L}_H^2 \text{ and } k \neq j \in \mathbb{Z}.
\]

Finally, we let \( \Rightarrow_N \) indicate convergence in distribution as \( N \to \infty \), where \( N \in \mathbb{N} \).

## 3 Main result

Throughout this article, we are interested in weakly stationary functional time series \( \{X_t : t \in \mathbb{Z}\} \) taking values in \( \mathcal{L}_H^2 \). In particular this means that the mean \( \mathbb{E}X_t = \mu \) and the \( h \)-lag covariance operator \( C_h \) are invariant under translations in time, i.e., \( C_h = \mathbb{E}(X_{t+h} - \mu) \otimes (X_0 - \mu) \). Without loss of generality, we shall assume that the data are centered. When the mean is unknown one can consider centering the data by subtracting the sample mean function (see Remark 4.1). Furthermore, we assume the process admits a representation of the form

\[
X_t = g(e_t, e_{t-1}, \ldots),
\]

where \( \{e_t : t \in \mathbb{Z}\} \) is an i.i.d. sequence of elements in some measurable space \( S \) and where \( g : S^\infty \to H \) is a measurable function. Functional processes with such representation are widely applicable and allow for example for nonlinear dynamics [e.g., 16]. It is clear from this representation that \( X \) is stationary and ergodic and we can consider the filtration \( \mathcal{G}_t = \sigma(e_t, e_{t-1}, \ldots) \). Moreover, it is straightforward to show that a stationary ergodic process can be written as

\[
X_h = \sum_{j=-\infty}^{h} P_j(X_h) \tag{3}
\]
To prove (i), it suffices to observe that

\[ \nu_{\mathbb{H},p}(X_t) = \|X_t - \mathbb{E}[X_t|\mathcal{G}_t,\{0\}]\|_{\mathbb{H},p}. \] (4)

Additionally, define the \( m \)-dependent version

\[ X_t^{(m)} = P_{t,t-m}X_t = \mathbb{E}[X_t|\sigma(\epsilon_t, \epsilon_{t-1}, \ldots, \epsilon_{t-m})]. \]

The following summarizes the assumption on the dependence structure made throughout this paper.

**Assumption 3.1.** \( \{X_t : t \in \mathbb{Z}\} \) is a centered stationary functional time series in \( L^p_{\mathbb{H}} \) such that

\[ \sum_{j=0}^{\infty} \nu_{\mathbb{H},p}(X_j) < \infty \] (5)

with \( p = 4 \).

Observe that Assumption 3.1 is weaker than \( L^p_{\mathbb{H}} \)-dependence as introduced in [16] and than physical dependence as given in [39].

**Lemma 3.0** (Assumption 5 versus \( L^p_{\mathbb{H}} \) and physical dependence). Define the random functions \( \tilde{X}_t^{(m)} = g(\epsilon_t, \ldots, \epsilon_{t-m+1}, \epsilon_{t-m+1}', \ldots) \) and \( \tilde{X}_t' = g(\epsilon_t, \ldots, \epsilon_1, \epsilon'_0, \epsilon_{-1}, \ldots) \), respectively. Then

1. \( \sum_{t=0}^{\infty} \|X_t - \tilde{X}_t^{(m)}\|_{\mathbb{H},p} < \infty \) implies \( \sum_{t=0}^{\infty} \nu_{\mathbb{H},p}(X_t) < \infty \).
2. \( \sum_{t=0}^{\infty} \|X_t - \tilde{X}_t'\|_{\mathbb{H},p} < \infty \) implies \( \sum_{t=0}^{\infty} \nu_{\mathbb{H},p}(X_t) < \infty \).

**Proof.** To prove (i), it suffices to observe that

\[ \nu_{\mathbb{H},p}(X_t) \leq \|X_t - P_{t,1}(\mathbb{E}[X_t|\mathcal{G}_t,\{0\}])\|_{\mathbb{H},p} + \|P_{t,1}(\mathbb{E}[X_t|\mathcal{G}_t,\{0\}]) - \mathbb{E}[X_t|\mathcal{G}_t,\{0\}]\|_{\mathbb{H},p} \]

\[ \leq 2\|X_t - P_{t,1}(X_t)\|_{\mathbb{H},p} = 2\|\mathbb{E}[X_t|\mathcal{G}_t] - \mathbb{E}[\tilde{X}_t^{(m)}|\mathcal{G}_t]\|_{\mathbb{H},p} \]

\[ \leq 2\|X_t - \tilde{X}_t^{(m)}\|_{\mathbb{H},p}. \]

Similarly, \( \nu_{\mathbb{H},p}(X_t) \leq \|X_t - X_t'\|_{\mathbb{H},p} + \|\mathbb{E}[X_t'|\mathcal{G}_t,\{0\}] - \mathbb{E}[X_t|\mathcal{G}_t,\{0\}]\|_{\mathbb{H},p} \leq 2\|X_t - X_t'\|_{\mathbb{H},p} \), from which (ii) follows.

Additionally, note that by Jensen’s inequality and the contraction property of the conditional expectation

\[ \|P_{0}(X_t)\|_{\mathbb{H},p} = \|\mathbb{E}[X_t - \mathbb{E}[X_t|\mathcal{G}_t,\{0\}]]\|_{\mathbb{H},p} \leq \nu_{\mathbb{H},p}(X_t). \]

Hence, under condition (5) we have \( \sum_{j=0}^{\infty} \|P_{0}(X_j)\|_{\mathbb{H},p} < \infty \). It is moreover interesting to relate Assumption 3.1 to summability of functional cumulant mixing conditions. In particular, the assumption of at least summability of the fourth order cumulant operator in either \( \|\cdot\|_2 \) or \( \|\cdot\|_1 \) is often made in available literature [see e.g. [30], [41], [33]]. These type of conditions are generally stronger than those considered in this paper and generalizations of the coefficients in [4] that can...
control for the interdependencies are required for such mixing conditions to hold. We postpone the discussion of these conditions to Section 5.

The assumption \( \{X_t : t \in \mathbb{Z}\} \in \mathcal{L}_H^2 \) ensures a finite second order structure of a random element of the form \( X_t \otimes X_s \). Note that the latter can be viewed as a random element of \( S_2(H) \), i.e., it is a measurable mapping from \( (\Omega, \mathcal{A}) \) into \( (S_2(H), \mathcal{B}) \) and thus \( X_t \otimes X_s \in \mathcal{L}_H^2 \). Existence of a limit of the quadratic form in (1) requires conditions on both the weight sequence as well as on the dependence structure. To elaborate on the latter, condition (5) has two implications (Proposition 3.1 and Proposition 3.2, resp.), which we shall make use of in order to derive distributional properties of the quadratic form. Denote the functional Discrete Fourier Transform (DFT) of the stationary process \( X \) by

\[
\mathcal{F}^{(\lambda)}_T = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} X_t e^{-i\lambda t}.
\]

Provided the dependence structure of the process decays fast enough, the limiting variance of (6) is given by the spectral density operator in (2). As the next statement shows, this is the case for processes that satisfy Assumption 3.1 with \( p = 2 \).

**Proposition 3.1.** Suppose Assumption 3.7 with \( p = 2 \) is satisfied. Then \( \sum_{h \in \mathbb{Z}} \|C_h\|_2 < \infty \) and \( \mathcal{F}^{(\lambda)} \) exists as a non-negative definite Hermitian element of \( S_2(H) \) for all \( \lambda \in (-\pi, \pi) \). Furthermore,

\[
\lim_{T \to \infty} \text{Var}(\mathcal{F}^{(\lambda)}_T) = \mathcal{F}^{(\lambda)}.
\]

**Proof of Proposition 3.1.** We obtain by orthogonality of the projections, stationarity and Jensen’s inequality

\[
\sum_{h=1}^{\infty} \|C_h\|_2 \leq \sum_{h=0}^{\infty} \sum_{j=-\infty}^{0} \mathbb{E}\|P_0(X_{h-j}) \otimes P_0(X_{j})\|_2 \leq \sum_{h=0}^{\infty} \sum_{j=-\infty}^{0} \sqrt{\mathbb{E}\|P_0(X_{h-j})\|_H^2 \mathbb{E}\|P_0(X_{j})\|_H^2} \leq \left( \sum_{j=0}^{\infty} \|P_0(X_j)\|_{H,2} \right)^2 < \left( \sum_{j=0}^{\infty} \|X_j\|_{H,2} \right)^2 \leq \infty
\]

and similarly for \( h < 0 \). Hence, \( \mathcal{F}_\lambda = \frac{1}{\sqrt{2\pi}} \sum_{h \in \mathbb{Z}} C_h e^{-i\lambda h} \) converges in norm \( \|\cdot\|_2 \) for all \( \lambda \in (-\pi, \pi) \). It follows that \( \mathcal{F}^{(\lambda)} \) is a non-negative definite, Hermitian \( S_2(H) \)-valued density function over frequencies that satisfies \( C_h = \int_{-\pi}^{\pi} \mathcal{F}^{(\lambda)} e^{i\lambda} d\lambda \) [e.g., 12, Thm 3.7]. Moreover, from the dominated convergence theorem one obtains

\[
\lim_{T \to \infty} \text{Var}(\mathcal{F}^{(\lambda)}_T) = \lim_{T \to \infty} \sum_{h \in \mathbb{T}} (1 - \frac{|h|}{T}) \mathbb{E}(X_h \otimes X_0) e^{-i\lambda h} = \mathcal{F}^{(\lambda)} \quad \lambda \in (-\pi, \pi).
\]

\( \blacksquare \)

Hence, \( \mathcal{F}^{(\lambda)} \) in (2) exists as a limit of Césaro averages of \( \{C_h e^{-i\lambda h} : h \in \mathbb{Z}\} \) in \( S_2(H) \). Without stronger assumptions, such as summability conditions, derivations of several distributional properties of the quadratic form in (1) do not appear obvious. Yet, Assumption 3.1 allows to proceed via an approximating \( S_2(H) \)-valued random process. Underlying this approximation is the following process

\[
D_{m,k,T}^{(\lambda)} := \frac{1}{\sqrt{2\pi}} \sum_{t=0}^{T} P_k(X_{t+k}) e^{-i\lambda t}.
\]

The second order structure of (8) is closely related to that of \( \mathcal{F}^{(\lambda)}_T \), but moreover has several useful properties that we shall make extensive use of.
Proposition 3.2. Under the conditions of Assumption 3.1 with \( p = 4 \), we have for all \( \lambda \in (-\pi, \pi] \),

(i) The process \( D_{m,k}^{(\lambda)} := D_{m,k,\infty}^{(\lambda)} \) exists and forms an \( m \)-dependent stationary martingale difference sequence with respect to the filtration \( \mathcal{F}_k \) in \( \mathcal{L}^4_H \).

(ii) The process \( \{D_{\infty,0,y}^{(\lambda)}\}_{T \geq 1} \) is Cauchy in \( \mathcal{L}^4_H \) with limit

\[
D_0^{(\lambda)} := \sum_{t=0}^{\infty} P_0(X_t) e^{-it\lambda}
\]

and the process \( \{D_{\infty,0,y}^{(\lambda)} \otimes D_{\infty,0,y}^{(\lambda)}\}_{T \geq 1} \) is Cauchy in \( \mathcal{L}^2_{S_2(H)} \) with limit \( D_0^{(\lambda)} \otimes D_0^{(\lambda)} \).

(iii) \( \lim_{m \to \infty} \lim_{T \to \infty} \text{Tr} \left[ \Pi_{ijkl} (\text{Var}(D_{m,0,y}^{(\lambda)}) \otimes \text{Var}(D_{m,0,y}^{(\lambda)}) \right] = \lim_{m \to \infty} \text{Tr} \left[ \Pi_{ijkl} (\mathcal{F}_{m}^{(\lambda)} \otimes \mathcal{F}_{m}^{(\lambda)}) \right] < \infty. \]

Here, \( \Pi_{ijkl} \) denote the permutation operator on \( \bigotimes_{i=1}^{4} H \) that permutes the components of a tensor product of simple tensors according to the permutation \( (1,2,3,4) \mapsto (i,j,k,l) \), that is, \( \Pi_{ijkl} (x_1 \otimes \cdots \otimes x_4) = (x_i \otimes \cdots \otimes x_l) \). The details of the proof can be found in Appendix A. For fixed \( m \) and \( p = 2 \), the first statement is almost immediate from the properties of the projection operators which form martingale difference sequences with respect to \( \mathcal{F}_k \) and the fact that the process \( \{X_t^{(m)}\} \) is \( m \)-dependent. For \( p = 4 \), the proof of the above statements requires extensions of inequalities such as Burkholder’s inequality for linear transforms of Hilbert-valued martingales; see Appendix A. The Cauchy property will be necessary to verify several aspects of the distributional properties, including verification of tightness on the function space of the quadratic form. Proposition 3.2(iii) shows in particular that the iterated limit in \( T \) and \( m \), respectively, of a certain functional of the variance operators of the family of martingale processes \( \{D_{m,0,y}^{(\lambda)}\}_{m,0,y} \), \( m \geq 1 \) converges to that of the corresponding functional of \( \mathcal{F}_{m}^{(\lambda)} \), i.e., of the limiting variance operators of the fDFT, and that this functional is finite.

Next, we require the following conditions on the sequence of weight operators. We assume that we have a representation \( \Phi_{T,s,t} = (\phi_{T,s,t} \otimes I_H) \), where \( \phi_{T,s,t} \in S_{\infty}(H) \) such that \( \Phi_{T,s,t} = \Phi_{T,s,t}^\dagger \). Observe that this is an operator in \( S_{\infty}(S_2(H)) \) with the property

\[
(\phi_{T,s,t} \otimes I_H)(X_s \otimes X_t) = \phi_{T,s,t}(X_s) \otimes I_H(X_t) = (\phi_{T,s,t} \otimes I_H)^\dagger (X_s \otimes X_t) = I_H(X_s) \otimes \phi_{T,s,t}^\dagger (X_t). \tag{9}
\]

Note that the identity operator can be replaced with any arbitrary bounded linear operator \( B_T \in S_{\infty}(H) \). Additionally, we require a few technical conditions to ensure that the weights are "well-behaved", i.e., the quadratic form exists as a well-defined random element of \( S_2(H) \) for which no degenerate (non-Gaussian) limiting distributions can arise.

Assumption 3.2 (Conditions on \( \phi : Z \times Z \rightarrow S_{\infty}(H) \)). Let \( T \in \mathbb{N} \) and \( \lambda \in (-\pi, \pi] \). Let \( A_{T,t} : Z \rightarrow S_{\infty}(H) \) be a continuous mapping such that \( A_{T,t} = A_{T,-t} \), \( \forall t \in Z \) and set \( \phi_{T,t}^{(A)} = A_{T,t} e^{i\lambda t} \). Denote

\[
\|\Phi_T\|_F^2 := \sum_{t=1}^{T} \sum_{s=1}^{T} \|\phi_{T,t-s}\|_\infty^2 \quad \text{and} \quad \phi_T^2 := \sum_{t=1}^{T} \|\phi_{T,t}\|_\infty^2.
\]

We assume,

(i) \( T \phi_T^2 = O(\|\Phi_T\|_F^2) \);

(ii) \( \max_{1 \leq t \leq T} \|\phi_{T,t}\|_\infty^2 = \max_{1 \leq t \leq T} \|A_{T,t}\|_\infty^2 = o(\phi_T^2) \);

(iii) \( \sum_{t=1}^{T} \|A_{T,t} - A_{T,t-1}\|_\infty^2 = o(\phi_T^2) \);}
(iv) \( \sum_{j=1}^{T-1} \sum_{s=1}^{j-1} \| \sum_{t=j+1}^{T} \phi_{T,s-t} \otimes \phi_{T,j-t} \|_{\infty}^{2} = o(\| \Phi_{T} \|_{p}^{4}). \)

Note that the first condition simply ensures a balance in order, i.e., the left-hand side is of the same order as the total sum of weights operator when the latter is viewed as a function-valued operator on \( Z \times Z \). Together with the second, this means the norm of none of the individual weight contributions dominates the order of the variance. The third condition ensures a “smooth” contribution of each component \( \Phi_{T,s,t}(X_{s} \otimes X_{t}) \) to the total mass of the quadratic form. The fourth condition is required to ensure that, as the overlap of the two bivariate operator-valued functions over \( Z \times Z \) gets smaller, the contribution to the total mass must become negligible. Observe that for the examples mentioned in the introduction where \( \phi_{T,s}^{(j)} \) are scalar-valued, the norms \( \| \cdot \|_{\infty} \) can be replaced by \( | \cdot | \). Condition (iv) on the kernel then simply means a bandwidth parameter \( b_{T} \ll 1 \) must ensure a local smoothing occurs. As will become clear in the next section, it predictably excludes that the periodogram operator without smoothing can provide an asymptotically Gaussian consistent estimator of the spectral density operator. Many different weight functions used for the consistent estimation of \( \mathcal{F}_{\lambda}^{4} \) will satisfy the above conditions, including the common choice of a bounded piecewise continuous lag window function with compact support, provided the bandwidth parameter ensures condition (iv) holds true (see Section 4).

In order to derive the properties of the quadratic form, a natural and common approach is to decompose \( \hat{\mathcal{Q}}_{T} \) into off-diagonal elements and diagonal elements as follows

\[
\hat{\mathcal{Q}}_{T} = \sum_{t=2}^{T-1} \sum_{s=1}^{t-1} \phi_{T,s}^{(j)}(X_{s} \otimes X_{t}) + \left( \sum_{t=2}^{T-1} \sum_{s=1}^{t-1} \phi_{T,s}^{(j)}(X_{s} \otimes X_{t}) \right)^{\top} + \sum_{1 \leq t \leq T} \Phi_{T,t,t}(X_{t} \otimes X_{t}). \tag{10}
\]

The main ingredient to the proof is to use that the off-diagonal elements, after centering around their mean, can be approximated by the process

\[
\mathcal{M}_{T,m}^{(\lambda)} = \sum_{t=2}^{T-1} \sum_{s=1}^{t-1} \phi_{T,s,t}^{(j)} \left( D_{m,t}^{(j)} \otimes D_{m,s}^{(j)} \right), \tag{11}
\]

where the functionals \( D_{m,t}^{(j)} \) are defined via (8) in Proposition 3.2(i). The intuition is therefore similar in spirit to the strategy applied in the Euclidean setting [see e.g., [27, 37]]. We emphasize that the aim of this paper is not the same nor can the weak convergence result in our paper be seen as a trivial extension of these works. We aim to derive consistency rates and joint distributional convergence of a set of operators where the quadratic form is very general, consisting of operator-valued weight operators of a Hilbertian-valued stochastic process. The derivation of the operator approximations and of the distributional properties, including the verification of tightness on the function space, are therefore far more involved. The convenient properties of \( \mathcal{L}^{2}_{\mathcal{F}(\mathcal{M})} \) are given in the next statement.

**Proposition 3.3.** Let \( \mathcal{M}_{T,m}^{(\lambda)} \) as defined in (11). Under Assumption 3.1 with \( p = 4 \) and fixed \( m \), the process

\[
\left\{ \| \Phi_{T} \|_{F}^{-1} \cdot \mathcal{M}_{T,m}^{(\lambda)} \right\}_{T \geq 1}
\]

is a martingale process in \( \mathcal{L}^{2}_{\mathcal{F}(\mathcal{M})} \) with respect to the filtration \( \{ \mathcal{F}_{T} \} \) for all fixed \( \lambda \in (-\pi, \pi] \).

**Proof of Proposition 3.3.** It is immediate that \( \mathcal{M}_{T,m}^{(\lambda)} \) is adapted to the filtration \( \mathcal{F}_{T} \). Secondly, from the properties of the operators \( \{ \phi_{T,s,t} \} \) we can write

\[
\mathcal{M}_{T,m}^{(\lambda)} = \sum_{t=2}^{T-1} \sum_{s=1}^{t-1} \phi_{T,s,t}^{(j)} \left( D_{m,t}^{(j)} \otimes D_{m,s}^{(j)} \right) = \sum_{t=2}^{T} D_{m,t}^{(j)} \otimes \left( \sum_{s=1}^{t-1} \phi_{T,s,t}^{(j)} D_{m,s}^{(j)} \right).
\]
From Proposition 3.2(i), $D^{(\lambda)}_{m,t}$ forms a stationary martingale difference sequence in $\mathcal{L}^2_H$ with respect to $\mathcal{G}_t$. Hence, using orthogonality of the increments and Lemma A.1 yields

$$E\|\mathcal{M}^{(\lambda)}_{T,m}\|_2^2 \leq \sum_{t=2}^T E\left\|D^{(\lambda)}_{m,t} \otimes \left(\sum_{s=1}^{t-1} \phi^{(\lambda)}_{T,s,t} D^{(\lambda)}_{m,s}\right)\right\|_2^2 \leq \|D^{(\lambda)}_{m,0}\|_{H,4}^2 \|D^{(\lambda)}_{m,0}\|_{H,4}^2 \sum_{t=2}^T \sum_{s=1}^{t-1} \|\phi^{(\lambda)}_{T,s,t}\|_2^2.$$ 

Noting that $\sum_{t=2}^T \sum_{s=1}^{t-1} \|\phi^{(\lambda)}_{T,s,t}\|_2^2 \approx 1/2 \|\Phi_T\|_{F}^{-1} = 1/2 \|\Phi_T\|_{F}^{-1}$, we obtain $\|\Phi_T\|_{F}^{-1} E\|\mathcal{M}^{(\lambda)}_{T,m}\|_2^2 < \infty$. Finally, observe that

$$E\left[D^{(\lambda)}_{m,t} \otimes \left(\sum_{s=1}^{t-1} \phi^{(\lambda)}_{T,s,t} D^{(\lambda)}_{m,s}\right)\right] = E\left[D^{(\lambda)}_{m,t} | \mathcal{G}_{t-1}\right] \otimes \left(\sum_{s=1}^{t-1} \phi^{(\lambda)}_{T,s,t} D^{(\lambda)}_{m,s}\right) = O_H$$

where we used that $\sum_{s=1}^{t-1} \phi^{(\lambda)}_{T,s,t} D^{(\lambda)}_{m,s}$ is $\mathcal{G}_{t-1}$-measurable and that $D^{(\lambda)}_{m,t}$ is a $H$-valued martingale with respect to $\mathcal{G}_t$. The result now follows. \hfill \Box

The following theorem states the distributional properties of the quadratic form.

**Theorem 3.1** (asymptotic normality of $\hat{\mathcal{M}}^{(\lambda)}_T$). Let $\{X_t\}$ be a random sequence with paths in a separable Hilbert space $H$ for which assumption (2.1) holds with $p = 4$ and suppose that the sequence $\{\Phi^{(\lambda)}_T\}$ satisfies Assumption 3.2. Then the quadratic form in (1) satisfies

$$(\|\Phi_T\|_{F}^{-1/2}(\hat{\mathcal{M}}^{(\lambda)}_T - E(\hat{\mathcal{M}}^{(\lambda)}_T)))_{j=1, \ldots, d} \Rightarrow (\mathcal{M}^{(\lambda)}_j)_{j=1, \ldots, d}$$

where, $\hat{\mathcal{M}}^{(\lambda)}_j$, $j = 1, \ldots, d$ are jointly complex Gaussian elements of $S_2(H)$

$$\begin{pmatrix} \Re(\hat{\mathcal{M}}^{(\lambda)}_j) \\ \Im(\hat{\mathcal{M}}^{(\lambda)}_j) \end{pmatrix} \sim \mathcal{N}(S_2(H)^{d \times (S_2(H)^d)}(O_H \Phi) \begin{pmatrix} \Re(\Gamma + \Sigma) \\ \Im(\Gamma + \Sigma) \end{pmatrix}) + \begin{pmatrix} \Re(\Gamma - \Sigma) \\ \Im(\Gamma - \Sigma) \end{pmatrix}.$$ 

The $(i,j)$-th element of the covariance operator is given by

$$\Gamma_{i,j} = \eta(\lambda_i + \lambda_j)4\pi^2 \left(\mathcal{F}^{(\lambda)}_i \otimes \mathcal{F}^{(\lambda)}_j + 1_{[0,\pi]} \mathcal{F}^{(\lambda)}_i \otimes \mathcal{F}^{(\lambda)}_j \right)$$

and of the pseudocovariance operator by

$$\Sigma_{i,j} = \eta(\lambda_i + \lambda_j)4\pi^2 \left(1_{[0,\pi]} \mathcal{F}^{(\lambda)}_i \otimes \mathcal{F}^{(\lambda)}_j + \mathcal{F}^{(\lambda)}_i \otimes \mathcal{F}^{(\lambda)}_j \right),$$

and where $\eta(x) = 1$ for $x = 2\pi z$, $z \in \mathbb{Z}$ and zero otherwise.

In particular, for distinct frequencies $\lambda_1, \ldots, \lambda_d \in [0,\pi]$, $\Gamma$ and $\Sigma$ are $d \times d$ diagonal matrices with $S_1(H \otimes H)$-valued components and hence for such choice of frequencies, the components of $\hat{\mathcal{M}}^{(\lambda)}_j$ are asymptotically independent.

**Proof of Theorem 3.1** We consider the sequence of processes $\{\xi^{(\lambda)}_T : T \in \mathbb{N}\}$ where

$$\xi^{(\lambda)}_T := (\|\Phi^{(\lambda)}_T\|_{F}^{-1/2}(\hat{\mathcal{M}}^{(\lambda)}_T - E(\hat{\mathcal{M}}^{(\lambda)}_T))).$$

Observe that $\{\xi^{(\lambda)}_T : T \in \mathbb{N}\}$ is a measurable stochastic processes with sample paths in the Hilbert space $S_2(H)$. We shall verify the two conditions in Lemma 3.1 to show weak convergence in $S_2(H)$ [see e.g., 3].

**Lemma 3.1** (weak convergence). Let $\{\xi_T : T \in \mathbb{N}\}$ be a stochastic process with sample paths in a separable Hilbert space. If the following two conditions are satisfied
The finite dimensional distributions of $\xi_T$ converge to those of $\xi$ a.e.;

ii) The family of laws $\mathcal{P} := (\mathbb{P}_T)_{T \in \mathbb{N}}$ of $\{\xi_T : T \in \mathbb{N}\}$ is tight.

Then, $\xi_T \Rightarrow T \xi$.

First we derive that, for all $m \geq 1$, $\xi^A_{T,m} \Rightarrow T \xi^A_m$, where $\xi^A_m$ defines a zero-mean Gaussian element of $S_2(H)$ and where the double indexed process is given by

$$\xi^A_{T,m} := (\|\Phi_T^2\|_F)^{-1/2}(\mathcal{M}^{(A)}_{T,m} + \mathcal{M}^{(A)}_{T,m})$$

with $\mathcal{M}^{(A)}_{T,m}$ as in (11). By Proposition 3.3, for every fixed $m$, $\mathcal{M}^{(A)}_{T,m}$ is a martingale process in $\mathcal{L}^2(S_2(H)(\Omega, \mathcal{A}, \mathbb{P})$ with respect to the filtration $\{\mathcal{F}_T\}$. Note the same holds for $\mathcal{M}^{(B)}_{T,m}$. Let $(\chi_i)_{i \geq 1}$ be an orthonormal basis of $H$. Then $(\chi_{i,l',l}) := (\chi_i \otimes \chi_{l'})_{i,l'}$ defines an orthonormal basis of $H \otimes H$ and we shall denote

$$\xi^A_{T,m}(\chi) = \langle \xi^A_{T,m}, \chi \rangle.$$

for any $\chi \in H \otimes H$. The result below shows that, for a finite set of frequencies, the finite-dimensional distributions of $\xi^A_{T,m}$ for fixed $m$ converge jointly to those of $\xi_m$ as $T \to \infty$, where these are asymptotically independent at distinct frequencies.

**Theorem 3.2.** Suppose the conditions of Theorem 3.1 hold. Then, for a finite set of distinct frequencies $\lambda_1, \ldots, \lambda_d \in [0, \pi)$, for all $m \geq 1$ and any $\chi_{i,l',l} \in H \otimes H$, we have

$$\xi^A_{T,m}(\chi_{i,l',l}) = \langle \xi^A_{T,m}, \chi \rangle.$$

where

$$\Gamma^A_m(\chi_{i,l',l}) = 4\pi^2 \left( \mathcal{F}_m^{(A)}(\chi_{i,l',l}) \mathcal{F}_m^{(A)}(\chi_{i,l',l}) + 1_{[0,\pi]}(\mathcal{F}_m^{(A)}(\chi_{i,l',l}) \mathcal{F}_m^{(A)}(\chi_{i,l',l})) \right)$$

and

$$\Sigma^A_m(\chi_{i,l',l}) = 4\pi^2 \left( 1_{[0,\pi]}(\mathcal{F}_m^{(A)}(\chi_{i,l',l}) \mathcal{F}_m^{(A)}(\chi_{i,l',l})) + \mathcal{F}_m^{(A)}(\chi_{i,l',l}) \mathcal{F}_m^{(A)}(\chi_{i,l',l}) \mathcal{F}_m^{(A)}(\chi_{i,l',l}) \right),$$

where $\mathcal{F}_m^{(A)}(\chi_{i,l',l}) = (E(D_{m,0}^{(A)}(\chi_{i,l',l})))$. The proof is tedious and relegated to Appendix B. Next, we show that $\forall m \geq 1, \{\xi^A_{T,m}, T \geq 1\}$ is tight. In order to verify tightness we shall use the following result, which is a particular case of [36, Theorem 3] who considers tightness criteria for more general Schauder decomposable Banach spaces.

**Lemma 3.2** (tightness on a separable Hilbert space). Let $(\chi_{i,l'})$ be an orthonormal basis of $H \otimes H$. A family of probability measures $\mathcal{P} := (\mathbb{P}_T)_{T \in \mathbb{N}}$ on $S_2(H)$ is tight if and only if

i) $\forall k \geq 1: \lim_{h \to \infty} \sup_T \mathbb{P}_T \left( \{ x \in S_2(H) : \sum_{i,l',k} |(x, \chi_{i,l'})|^2 > h \} \right) = 0$;

ii) $\forall \varepsilon > 0: \lim_{k \to \infty} \sup_T \mathbb{P}_T \left( \{ x \in S_2(H) : \sum_{i,l',i+l',k} |(x, \chi_{i,l'})|^2 > \varepsilon \} \right) = 0$.

In order to verify the first condition, note that, since $k$ is fixed

$$\lim_{h \to \infty} \sup_T \left( \sum_{i,l',k} |(\xi^A_{T,m}, \chi_{i,l'})|^2 > h \right) \leq \sum_{i,l',k} \lim_{h \to \infty} \sup_T \left( |(\xi^A_{T,m}, \chi_{i,l'})|^2 > h \right),$$

and hence the first condition is implied by

$$\forall l, l' \geq 1: \lim_{h \to \infty} \sup_T \left( |(\xi^A_{T,m}, \chi_{i,l'})|^2 > h \right) = 0.$$  

(13)
for which we moreover have
\[ \mathbb{P}\left( |\langle \xi_{T,m'}^A, \chi_{l}^r \rangle|^2 > h \right) \leq \mathbb{P}\left( \Re(\langle \xi_{T,m'}^A, \chi_{l}^r \rangle)^2 > h/2 \right) + \mathbb{P}\left( \Im(\langle \xi_{T,m'}^A, \chi_{l}^r \rangle)^2 > h/2 \right). \]

Since the real and imaginary part of the random variables \( \langle \xi_{T,m'}^A, \chi_{l}^r \rangle \) converge to real-valued random variables by Theorem 3.2, the corresponding sequence of probability measures is tight on \((\mathbb{R}, \mathcal{B})\). (13) therefore follows from the continuous mapping theorem. In order to verify the second condition of Lemma 3.2, note that by Markov's inequality it suffices to prove that
\[ \limsup_{k \to \infty} \sum_{T} \mathbb{E}|\xi_{m,T}^A(\psi_{l}^r)|^2 = 0. \]

Firstly, observe that \( \mathbb{E}|\xi_{m,T}^A(\psi_{l}^r)|^2 \geq 0 \) and \( \langle \text{Var}(\xi_{m,T}^A(\psi_{l}^r)(\psi_{l}^r)) \rangle \geq 0 \). Note then from (12) that for any \( \chi_{l}^r \in H \otimes H \)
\[ \lim_{T \to \infty} \mathbb{E}|\xi_{m,T}^A(\chi_{l}^r)|^2 = \Gamma_m^A(\chi_{l}^r) = \text{Var}(\xi_{m,T}^A(\chi_{l}^r)) < \infty. \]
Together with Parseval's identity the monotone convergence and by definition of the (transpose) Kronecker tensor product, Theorem 3.2 implies
\[ \limsup_{T \to \infty} \mathbb{E}\|\xi_{T,m}^A\|^2 \leq \sum_{l,l'} \mathbb{E}\|\xi_{T,m}^A(\psi_{l}^r)^2 \]
\[ = \sum_{l,l'} 4\pi^2 \left( \mathcal{F}_m^{(A)}(\psi_{l}^r)(\mathcal{F}_m^{(A)}(\psi_{l}^r) + 1_{[0,\pi]}(\mathcal{F}_m^{(A)}(\psi_{l}^r)(\mathcal{F}_m^{(A)}(\psi_{l}^r))) \right) \]
\[ = 4\pi^2 \text{Tr}\left( \mathcal{F}_m^{(A)} \otimes \mathcal{F}_m^{(A)} \right) + \text{Tr}\left(1_{[0,\pi]}(\mathcal{F}_m^{(A)} \otimes \mathcal{F}_m^{(A)}) \right) = \mathbb{E}\|\xi_{m}^A\|^2. \]

From Proposition 3.2 iii.) we find immediately that
\[ \mathbb{E}\|\xi_{m}^A\|^2 = \text{Tr}(\text{Var}(\xi_{m}^A)) < \infty. \]
Consequently, we can choose an \( \epsilon > 0 \) such that for all \( k \geq k_0 \)
\[ |\text{Tr}(\text{Var}(\xi_{m}^A)) - \sum_{l,l' \leq k_0} \langle \text{Var}(\xi_{m}^A(\psi_{l}^r)), (\psi_{l}^r) \rangle | < \epsilon. \]

From the pointwise convergence (15) and from the sequence convergence in (16), we obtain
\[ \lim_{T \to \infty} \sum_{l,l' \geq k_0} \mathbb{E}|\xi_{m,T}^A(\psi_{l}^r)|^2 = \lim_{T \to \infty} \left( \sum_{l,l'=1} \mathbb{E}|\xi_{m,T}^A(\psi_{l}^r)|^2 - \sum_{l,l' < k_0} \mathbb{E}|\xi_{m,T}^A(\psi_{l}^r)|^2 \right) \]
\[ \leq \text{Tr}(\text{Var}(\xi_{m}^A)) - \sum_{l,l' \leq k_0} \langle \text{Var}(\xi_{m}^A(\psi_{l}^r)), (\psi_{l}^r) \rangle. \]
In other words, there must exist a \( T_0 \) such that for all \( T \geq T_0 \) and \( k \geq k_0 \)
\[ |\mathbb{E}\|\xi_{m,T}^A\|^2 - \sum_{l,l' < k} \mathbb{E}|\xi_{m,T}^A(\psi_{l}^r)|^2 | < \epsilon. \]
Moreover, we can choose a \( \tilde{k} \geq k_0 \) such that for all \( 1 \leq T < T_0 \),
\[ \mathbb{E}\|\xi_{m,T}^A\|^2 - \sum_{l,l' < \tilde{k}} \mathbb{E}|\xi_{m,T}^A(\psi_{l}^r)|^2 < \epsilon. \]
(14) now follows by taking max ($k, \tilde{k}$). Therefore, we have established both conditions of Lemma 3.1 and thus $\xi^{1}_{T,m} \Rightarrow_{T} \xi^{1}_{m}$. Next, we will show that $\xi^{1}_{T,m} \Rightarrow_{m} \hat{\xi}^{1}$ where $\hat{\xi}^{1}$ denotes the limiting process given in Theorem 3.1. We again verify the conditions of Lemma 3.1. From Theorem 3.2 and Proposition 3.2(iii.), we find

$$
\lim_{m \to \infty} E\|\xi^{A}_{m}\|_{2}^{2} = \lim_{m \to \infty} 4\pi^{2} \text{Tr} \left( \mathcal{F}^{(1)} \otimes \mathcal{F}^{(1)} + 1_{[0,\pi]} (\mathcal{F}^{(1)} \otimes \mathcal{F}^{(1)}) \right) = 4\pi^{2} \text{Tr} \left( \mathcal{F}^{(1)} \otimes \mathcal{F}^{(1)} + 1_{[0,\pi]} (\mathcal{F}^{(1)} \otimes \mathcal{F}^{(1)}) \right) = E\|\hat{\xi}^{1}\|_{2}^{2} < \infty.
$$

(17) Recall then that Theorem 3.2 shows that for fixed $m$ and any $\chi \in H \otimes H$, $\xi^{1}_{m}(\chi)$ is a zero mean complex-valued Gaussian random variable. Hence $\xi^{1}_{m}(\chi) \Rightarrow_{m} \hat{\xi}^{1}(\chi)$ if we can show that the covariance structure satisfies

$$
\lim_{m \to \infty} \Gamma_{m}(\chi_{l'}) = \Gamma(\chi_{l'}) \quad \lim_{m \to \infty} \Sigma_{m}(\chi_{l'}) = \Sigma(\chi_{l'}),
$$

(18) where $\Gamma$ and $\Sigma$ are the covariance and pseudocovariance operator given in Theorem 3.1. This however follows immediately from (17). Hence, $\xi^{1}_{m}(\chi_{l'}) \Rightarrow_{T} \hat{\xi}^{1}(\chi_{l'})$ showing the finite-dimensional distributions converge. Similar to (13) this implies that

$$
\forall l, l' \geq 1: \lim_{h \to \infty} \sup_{m} \mathbb{P} \left( |(\xi^{1}_{m},\chi_{l'})|^{2} > h \right) = 0.
$$

(19) Hence, condition Lemma 3.2(i) is satisfied. The tightness condition Lemma 3.2(ii) is satisfied if

$$
\lim_{k \to \infty} \sup_{m} \sum_{l+l' > k} E|\xi^{1}_{m}(\psi_{l'})|^{2} = 0.
$$

(20) From the pointwise convergence (18) and the convergence of (17) as $m \to \infty$, this now however follows similarly to the proof of (20). Altogether, this establishes $\xi^{1}_{T,m} \Rightarrow_{T} \xi^{1}_{m} \Rightarrow_{m} \hat{\xi}^{1}$. Finally, it remains to show $\xi^{1}_{T,m} \Rightarrow_{T} \xi^{1}_{m} \Rightarrow_{T} \hat{\xi}^{1}$, for which we make use of the next lemma.

**Lemma 3.3.** Under the conditions of Theorem 3.1

$$
\lim_{m \to \infty} \limsup_{T \to \infty} \frac{1}{\|\Phi_{T}\|_{F}} \left\| \hat{\xi}^{1} - E\hat{\xi}^{1} - M_{T,m}^{(1)} - M_{T,m}^{(1)} \right\|_{S_{2,2}} = 0.
$$

(21) The proof can be found in Appendix C. Since $S_{2}(H)$ is a complete metric space, let $F$ be a closed set of $S_{2}(H)$ and fix $\epsilon > 0$. Then

$$
\mathbb{P}(\xi^{1}_{T} \in F) \leq \mathbb{P}(\|\xi^{1}_{T,m} - \xi^{1}_{m}\|_{2} \geq \epsilon) + \mathbb{P}(\xi^{1}_{T,m} \in \{x: \|x - y\|_{2} \leq \epsilon, y \in F\})
$$

and since by the weak convergence of $\xi^{1}_{T,m} \Rightarrow_{T} \xi^{1}_{m} \Rightarrow_{m} \hat{\xi}^{1}$, we have

$$
\lim_{m \to \infty} \limsup_{T \to \infty} \mathbb{P}(\xi^{1}_{T,m} \in \{x: \|x - y\|_{2} \leq \epsilon, y \in F\}) \leq \mathbb{P}(\hat{\xi}^{1} \in \{x: \|x - y\|_{2} \leq \epsilon, y \in F\}).
$$

Using then Lemma 3.3 Markov’s inequality yields

$$
\limsup_{T \to \infty} \mathbb{P}(\xi^{1}_{T} \in F) \leq \mathbb{P}(\hat{\xi}^{1} \in \{x: \|x - y\|_{2} \leq \epsilon, y \in F\}),
$$

so that taking $\epsilon \to 0$, completes the proof. \(\Box\)
4 Estimation of the spectral density operator

In this section, we focus on the application of the above theorem to estimate the spectral density operator

$$\mathcal{F}(\lambda) = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} C_h e^{-ih\lambda}. $$

Proofs of the statements in this section are postponed to Appendix D. It is well-known that under various conditions [see e.g., 15, 30] an asymptotically unbiased estimator is given by the periodogram operator

$$\mathcal{F}^A_T := \mathcal{D}^A_T \otimes D^A_T$$

where $D^A_T$ are the fDFT of $X$ given in Section 2. Note that by construction, this operator is hermitian, non-negative definite and $\lambda \mapsto \mathcal{F}^A_T$ is $2\pi$-periodic. From (7), we can immediately conclude that, under the stated conditions, the periodogram operator is indeed an asymptotically unbiased estimator of $\mathcal{F}(\lambda)$. It can however never be consistent because it is based upon one frequency observation. A consistent estimator of the spectral density operator can be obtained via smoothing the operator-valued function $\lambda \mapsto \mathcal{F}^A_T$ over neighboring frequency ordinates, i.e., via convolving the periodogram operator with a window function $K$. For example, it is very common to consider an estimator of the form

$$\hat{\mathcal{F}}^\omega = \frac{1}{b} \int_{-\infty}^{\infty} \mathcal{F}(\omega - \lambda) \mathcal{D}^A_T \otimes D^A_T d\lambda,$$

where $K : \mathbb{R} \rightarrow \mathbb{R}_+$ is assumed to be an even, non-negative weight function that is integrable. Under Assumption 3.1 with $p = 4$, it is immediate from an application of the Cauchy-Schwarz inequality and Lemma A.2(i) that $\sup_{\lambda} \|\mathcal{F}^A_T\|_{S^2_\mathbb{R}} = O(1)$ uniformly in $T$. By Holder’s inequality, (22) therefore exists as an element of $\|\|_{S^2_\mathbb{R}}$. In order to exploit the results from the previous section, we however require the estimator can be formulated in terms of a quadratic form. As remarked in the introduction, we consider

$$\hat{\mathcal{F}}^\omega = \frac{1}{2\pi T} \sum_{s,t=1}^{T} \left( (X_s - \mu) \otimes (X_t - \mu) \right) w(b_T(t-s)) e^{i\omega(t-s)}. $$

Note that $\hat{\mathcal{F}}^\omega = \frac{1}{2\pi T} \hat{\Phi}_{T,\omega}$ with $\Phi_{T,\omega} = \phi_{T,\omega}(s-t)I_{H \otimes H} = w(b_T(t-s)) e^{i\omega(t-s)} I_{H \otimes H}$ thus yields the representation in terms of the quadratic form introduced in the previous section. Provided $w(\cdot)$ and $K(\cdot)$ form Fourier pairs, there is a clear connection between (22) and (23). Namely, a change of variables gives

$$\hat{\mathcal{F}}^\omega = \frac{1}{b} \int_{-\infty}^{\infty} \mathcal{F}(\omega - \lambda) \mathcal{D}^A_T \otimes D^A_T d\lambda = \int_{-\infty}^{\infty} K(x) \mathcal{F}^0 + x b_T d\lambda$$

$$= \int_{-\infty}^{\infty} K(x) \frac{1}{2\pi T} \sum_{s,t=1}^{T} e^{-i(\omega + x b_T)(s-t)} (X_s \otimes X_t) dx$$

$$= \int_{-\infty}^{\infty} K(x) \frac{1}{2\pi T} \sum_{s,t=1}^{T} (X_s \otimes X_t) e^{-i\omega(s-t)} \int_{-\infty}^{\infty} K(x) e^{ix b_T(t-s)} dx$$

$$= \frac{1}{2\pi T} \sum_{s,t=1}^{T} (X_s \otimes X_t) w(b_T(t-s)) e^{i\omega(t-s)},$$

where the equality is with respect to $\|\|_{S^2_\mathbb{R}}$. In order to verify consistency and asymptotic normality, we require the following assumptions on the weight function $w(\cdot)$ in (23).
Assumption 4.1. Let \( w \) be an even, bounded function on \( \mathbb{R} \) with \( \lim_{x \to 0} w(x) = 1 \) that is continuous except at a finite number of points. Furthermore, suppose that \( \lim_{b \to 0} b \sum_{h \in \mathbb{Z}} w^2(bh) = \kappa \) where 
\[
\kappa := \int_{\mathbb{R}} w^2(x) \, dx < \infty \quad \text{such that} \quad \sup_{0 \leq b \leq 1} b \sum_{|h| \geq M/b} w^2(bh) \to 0 \quad \text{as} \quad M \to \infty.
\]

Observe that these are rather mild conditions for window functions and includes a wide range of common choices [see e.g. [4]]. Under these conditions we can obtain consistency in mean square of the spectral density operator.

Theorem 4.1. Suppose Assumption 3.1 with \( p \geq 2 \) and Assumption 4.1 are satisfied. Then,

(i) \( \sup_{\lambda \in [0, \pi]} \| \hat{F}^\lambda_T - F^\lambda_T \|_{S_2, p}^2 \to 0 \) if \( b_T \to 0 \) as \( T \to \infty \) such that \( b_T T \to \infty \).

(ii) If, in addition, \( \sum_{h \in \mathbb{Z}} h \| P_0(X_h) \|_{H, 2} < \infty \) and \( w(x) - 1 = O(x) \) as \( x \to 0 \), then

\[
\| \hat{F}^\lambda_T - F^\lambda_T \|_{S_2, p}^2 = O \left( \frac{1}{b_T T} + b_T^2 \right)
\]

uniformly in \( \lambda \in [0, \pi] \).

Note that Theorem 4.1 does not rely on a martingale approximation to exist but relies on the ergodicity properties of the underlying process. Without loss of generality, we can restrict to the interval \([0, \pi]\) since the mappings \( \lambda \mapsto \hat{F}^\lambda_T \) and \( \lambda \mapsto F^\lambda_T \) are even and \( 2\pi \)-periodic. Under Assumption 3.1 with \( p = 4 \) and Assumption 4.1 we in fact obtain that \( \sup_{\lambda \in [0, \pi]} \mathbb{E} \| \hat{F}^\lambda_T - F^\lambda_T \|_2^2 = O \left( \frac{1}{b_T^2} \right) \). It is however often of importance to specify the rate of consistency and hence to control the order of the bias in norm. As given in the second part of the statement, this requires mild additional conditions on the smoothness of the process as well as a smoothness condition of the weight function around 0.

Remark 4.1 (If the function \( \mu \) is unknown). In case the mean function \( \mu \) is unknown, we can instead consider the estimator

\[
\hat{F}^{(\lambda)} = \frac{1}{2\pi T} \sum_{s, t = 1}^T \left( (X_s - \hat{\mu}) \otimes (X_t - \hat{\mu}) \right) w(b_T(t - s))e^{i\lambda(t - s)},
\]

where \( \hat{\mu} = \frac{1}{T} \sum_{j=1}^T X_T \) denotes the sample mean function and which defines a random element of \( H \). We obtain the following error bound with the estimator in (23), which shows the results in this section are not affected by centering the data using the sample mean.

Lemma 4.1. Suppose Assumption 3.1 with \( p = 4 \) is satisfied and Assumption 4.1 holds. Then

\[
\sup_{\lambda \in [0, \pi]} \mathbb{E} \| \hat{F}^{(\lambda)} - \hat{F}^{(\lambda)} \|_2^2 = O((b_T T)^{-2}).
\]

More generally, if \( X \in \mathcal{L}^{2p}_H \), then \( \| \hat{F}^{(\lambda)} - \hat{F}^{(\lambda)} \|_{S_2, p} = O((b_T T)^{-1}), \ p \geq 1. \)

A consequence of Theorem 4.1 is that dynamic functional principal component analysis and hence an optimal dimension reduction of functional time series can be derived under weaker conditions than currently available [31, 15]. A central role for such techniques is played by the empirical eigenvalues of an estimator \( \hat{F}^\lambda_T \) of \( F^\lambda_T \). More specifically, since \( F^\lambda_T \) and \( \hat{F}^\lambda_T \) are compact self-adjoint non-negative definite operators, they admit representations of the form

\[
F^\lambda_T = \sum_{j=1}^{\infty} \beta^\lambda_j \Pi^\lambda_j \quad \text{and} \quad \hat{F}^\lambda_T = \sum_{j=1}^{\infty} \hat{\beta}^\lambda_j \hat{\Pi}^\lambda_j,
\]
where the sequence of eigenvalues \( \{ \beta_j^1 \}_{j=1} \) is a non-increasing sequence of positive real numbers tending to zero and where \( \Pi_j^1 = \phi_j^1 \otimes \phi_j^1 \) and \( \hat{\Pi}_j^1 = \hat{\phi}_j^1 \otimes \hat{\phi}_j^1 \) are the \( j \)-th eigenprojectors of \( \mathcal{F}_j^1 \) and \( \hat{\mathcal{F}}_j^1 \), respectively. Using the result of \( \text{Theorem 4.1}(i) \), we obtain consistency of the empirical eigenprojectors and eigenvalues for their population counterparts.

**Proposition 4.1.** Suppose \( \text{Assumption 3.1} \) with \( p = 4 \) and \( \text{Assumption 4.1} \) are satisfied. Then for each \( j \geq 1 \) and each \( \lambda \in [0, \pi] \), and for bandwidths satisfying \( b_T \to 0 \) as \( T \to \infty \) such that \( b_T \to \infty \),

(i) \( |\beta_j^1 - \hat{\beta}_j^1| \xrightarrow{P} 0 \);

(ii) If in addition \( \inf_{l\neq j} |\beta_l^1 - \beta_j^1| > 0 \), we have \( \|\hat{\Pi}_j^1 - \Pi_j^1\|_2 \xrightarrow{P} 0 \).

The proof of (i) and (ii) follow (almost) immediately from results of \([13]\) and \([28]\), respectively. Details are given in the Appendix. Note that the consistency is given for the projectors since the empirical eigenfunctions can only be identified up to rotation on the unit circle, that is, we obtain a version \( c(\lambda)\hat{\phi}_j^1 \), for some \( c(\lambda) \in \mathbb{C} \) with \( |c(\lambda)|^2 = 1 \). In contrast, the projectors are invariant to the rotation.

The next result is the joint distributional convergence of a set of estimators at distinct frequencies to uncorrelated Gaussian elements of \( S_2(H) \).

**Theorem 4.2.** Suppose \( \text{Assumption 3.1} \) with \( p = 4 \) and \( \text{Assumption 4.1} \) are satisfied. Let \( \lambda_1, \ldots, \lambda_d \in [0, \pi] \) be distinct. Then, for bandwidths satisfying \( b_T \to 0 \) such that \( b_T T \to \infty \) as \( T \to \infty \)

\[
\sqrt{b_T T} \left( \mathcal{F}_j^1 - \mathbb{E} \mathcal{F}_j^1 \right)_{j=1,\ldots,d} \Rightarrow T \left( \tilde{\mathcal{F}}_j^1 \right)_{j=1,\ldots,d}
\]

where \( \mathcal{F}^1_j, j = 1, \ldots, d \) are zero-mean jointly independent complex Gaussian elements of \( S_2(H) \), with covariance operator

\[
\text{Cov}(\mathcal{F}^1_j, \mathcal{F}^1_{j'}) = \kappa^2 \left( \mathcal{F}(\lambda_j) \otimes \mathcal{F}(\lambda_{j'}) + 1_{[0,\pi]} \mathcal{F}(\lambda_j) \otimes \mathcal{T} \mathcal{F}(\lambda_{j'}) \right)
\]

and with pseudocovariance operator

\[
\text{Cov}(\tilde{\mathcal{F}}^1_j, \tilde{\mathcal{F}}^1_{j'}) = \kappa^2 \left( 1_{[0,\pi]} \mathcal{F}(\lambda_j) \otimes \mathcal{F}(\lambda_{j'}) + \mathcal{F}(\lambda_j) \otimes \mathcal{T} \mathcal{F}(\lambda_{j'}) \right).
\]

If the conditions of \( \text{Theorem 4.1}(ii) \) are also satisfied, then

\[
\sqrt{b_T T} \left( \mathcal{F}^1_j - \mathbb{E} \mathcal{F}^1_j \right)_{j=1,\ldots,d} \Rightarrow T \left( \tilde{\mathcal{F}}^1_j \right)_{j=1,\ldots,d}
\]

for bandwidths satisfying \( b_T = o(T^{-1/3}) \).

Observe that if \( \lambda_j \in [0, \pi] \), then \( \mathcal{F}^1_j \) is real Gaussian. Finally, we obtain the following corollary on the distributional properties of the estimator of the long run covariance operator, which can be seen to improve upon the results in \([11]\) and \([30]\).

**Corollary 4.1.** Under the conditions of \( \text{Theorem 4.2} \)

\[
\sqrt{b_T T} 2\pi \left( \mathcal{F}^{(0)} - \mathcal{F}^{(0)} \right) \Rightarrow T \mathcal{N}_{S_2(H)}(0, 4\pi^2 \Gamma^{(0)})
\]

where \( \Gamma^{(0)} = \kappa^2 \left( \mathcal{F}^{(0)} \otimes \mathcal{F}^{(0)} + \mathcal{F}^{(0)} \otimes \mathcal{T} \mathcal{F}^{(0)} \right) \).
5 Relation to functional cumulant mixing conditions

Unlike the majority of existing literature [see e.g.,\([30, 41, 33, 11]\)], the results of this paper do not require higher order functional cumulant mixing conditions of the form

\[
\sum_{t_1, \ldots, t_{n-1} \in \mathbb{Z}} \| \text{cum}(X_{t_1}, \ldots, X_{t_{n-1}}, X_0) \|_2 < \infty. \tag{25}
\]

Here, the \(n\)-th order joint cumulant tensor of \(X_{t_1}, \ldots, X_{t_n} \in \mathcal{L}_H^n\), which can be viewed as an element of \(S_2(\otimes^{\lceil \frac{n+1}{2} \rceil} H, \otimes^{\lfloor \frac{n}{2} \rfloor} H)\), is defined by

\[
\text{cum}(X_{t_1}, \ldots, X_{t_n}) = \sum_{\nu = (v_1, \ldots, v_p)} (|\rho| - 1)!(-1)^{|\rho|-1} \Pi_\nu \left( \otimes_{t=1}^\rho E \left[ \otimes_{i \in v_t} X_{t_i} \right] \right), \tag{26}
\]

where the summation extends over all unordered partitions \(\nu\) of \([1, \ldots, n]\) and \(\Pi_\nu\) denotes the permutation operator that maps the components of the tensor back into the original order, that is, \(\Pi_\nu(\otimes_{t=1}^\rho X_{t_i}) = X_{t_{\nu_1}} \otimes \cdots \otimes X_{t_{\nu_\rho}}\) [see e.g.,\([11]\)]. The cumulant tensor in (26) measures the joint statistical dependence of order \(n\) and condition (25) ensures the span of dependence is small enough for the existence – in a Hilbert-Schmidt sense – of \(n\)-th order spectral cumulant tensors (functional polyspectra), which can be seen to capture nonlinear dynamics of the process for \(n \geq 3\). When not assumed explicitly, it can in general be difficult to verify if conditions of the form (25) are satisfied. Because of their natural relation to higher order functional polyspectra and their frequent usage as an underlying assumption in existing literature on statistics that are of the form (1), it is however of interest to understand how these relate to the cumulative measure in Assumption 3.1, which is more easily tractable. In order to do so, we make a few simple observations. Firstly, cumulant operators can be viewed as generalizations of covariance operators, which is easily seen from rewriting (26) as

\[
\text{cum}(X_{t_1}, \ldots, X_{t_n}) = E[X_{t_1} \otimes \cdots \otimes X_{t_n}] - \sum_{\nu_1 |\rho| \neq 1} \Pi_\nu \left( \otimes_{t=1}^\rho \text{cum}(X_{t_i}; i \in v_t) \right). \tag{27}
\]

Hence, the \(n\)-th order cumulant tensor can be interpreted as a measure of the \(n\)-th order joint dependence corrected for by all lower order joint dependencies, i.e., it captures the interaction between the \(n\) variables that is not captured by any subset of the \(n\) variables. Secondly, the cumulant of order \(n\) of a stationary process is translation invariant and is a function of \(n-1\) time differences w.r.t. a chosen base time, e.g., \(\text{cum}(X_{t_1}, \ldots, X_{t_n}) = \text{cum}(X_{t_1-t_n}, \ldots, X_{t_{n-1}-t_n}, X_0)\).

For \(p = n = 2\), a direct relation is then in fact given in Proposition 3.1 which shows that \(\sum_{t=0}^\infty \|X_0 - E[X_0|\mathcal{G}_{0,-t}]\|_{\mathcal{H}_2} < \infty\) implies \(\sum_{t \in \mathbb{Z}} \|C_t\|_2 = \sum_{t \in \mathbb{Z}} \|\text{cum}(X_t, X_0)\|_2 < \infty\). Note that this is intuitive since Assumption 3.1 provides a cumulative measure of the dependence of \(X_0\) on \(\epsilon_{-t}\), i.e., the dependence on the element of the data-generating mechanism \(t\) lags apart. This captures all second order dynamics; there is only one direction from the base time in which the dependence span needs to be controlled for and Assumption 3.1 with \(p = n = 2\) suffices.

This measure appears however inadequate to capture the higher order dynamics that come into play for \(n \geq 3\). More specifically, the coefficients of dependence (1) cannot directly control for the interactions in the various directions since the lags between the \(n-1\) indices cannot be exploited. Tediuous calculations in the appendix indicate that in order to capture the magnitude of these dynamics, we require a generalization of (4) to

\[
v_{X, \mathcal{H}, p}(j_1, \ldots, j_{n-1}) := \left\| X_0 + \sum_{i=1}^{n-1} (-1)^i \sum_{1 \leq l_i < \ldots < l_j \leq n-1} E[X_0|\mathcal{G}_{0,-j_1-\ldots-j_{n-1}}] \right\|_{\mathcal{H}, p} \tag{28}
\]
for \( j_1, \ldots, j_{n-1} \geq 0 \). Observe that (28) accounts for the dependence in all \( n - 1 \) directions from the base time simultaneously, corrected for by the "over-counted" interactions (in spirit of the exclusion-inclusion principle). It therefore fully encapsulates the dynamics of a \( n \)-th order joint cumulant tensor. We remark that a slightly weaker (yet less intuitive) condition was derived in the calculations in terms of composite projection operators (see (57)). Using (28), we obtain the following sufficient condition to ensure summability of cumulant tensors.

**Theorem 5.1.** Suppose \( \{X_t : t \in \mathbb{Z}\} \) is a centered stationary functional time series in \( \mathcal{L}^p_H \) with \( p = n \geq 2 \). Suppose moreover that

\[
\sum_{j_1, \ldots, j_k = 0}^{\infty} \nu_{X,H,p}(j_1, \ldots, j_k) < \infty \quad \text{for } k = n - 1. \tag{29}
\]

Then, for all \( i \leq p \),

\[
\sum_{t_1, \ldots, t_{i-1} \in \mathbb{Z}} \| \text{cum}(X_{t_1}, \ldots, X_{t_{i-1}}, X_0) \|^2 < \infty.
\]

For \( k \geq 2 \), (29) can be seen to capture the nonlinear dynamics in the process, while it coincides with the standard measure in (5) for \( k = 1 \). Condition (29) is weak in the sense that for \( k \geq 2 \) it is in fact identically zero for linear processes, in which case (29) does not impose additional constraints. Neither summability of cumulants nor Assumption 3.1 are however required for linear processes [see e.g., 30, 11]. The following provides an upper bound on (29) in terms of a weighted version of the standard coefficients of dependence.

**Proposition 5.1.** For \( p \geq 2 \),

\[
\sum_{j_1, \ldots, j_k = 0}^{\infty} \nu_{X,H,p}(j_1, \ldots, j_k) \leq 2^{-k+1} \sum_{j=0}^{\infty} j^{k-1} \nu_H(X_j) \quad \text{for all } 1 \leq k \leq p - 1. \tag{30}
\]

Observe that for \( k = 1 \) this is an equality. Additionally, note that the exponential factor in the upper bound on the right hand side of (30) increases in \( k \). As this provides an upperbound for the \( k + 1 \)-th order cumulant tensor, this can be seen to control the interaction in the various directions (and hence the dependence span) in a very rough way. It is worth mentioning that our findings corroborate with those in [40] who obtain a similar bound to ensure summability of the third order cumulant of a univariate time series.

**Acknowledgements** This work has been supported by the Collaborative Research Center “Statistical modeling of nonlinear dynamic processes” (SFB 823, Teilprojekt A1, C1) of the German Research Foundation (DFG). The author kindly thanks Holger Dette and Davide Giraudo for some useful discussions. Furthermore, the author sincerely thanks the Editor and two anonymous referees for their constructive comments that helped to produce an improved version of the original paper.

**References**

[1] Berkes, I., Horváth L. & Rice, G. On the asymptotic normality of kernel estimators of the long run covariance of functional time series. *Journal of Multivariate Analysis*, 144:150–175 (2017).

[2] Bhansali, R., Giraitis, L. & Kokoszka, P. Approximations and limit theory for quadratic forms of linear variables stochastic processes and their applications. *Stochastic processes and their applications*, 117:71–95 (2007).
[3] Billingsley, P. *Convergence of Probability Measures.* Wiley, New York (1968).

[4] Bosq, D. Estimation of mean and covariance operator of autoregressive processes in Banach spaces *Statistical Inference for Stochastic Processes,* 5:287–306 (2002).

[5] Brillinger, D. (1981). *Time Series: Data Analysis and Theory.* McGraw Hill, New York.

[6] Brockwell, P. & Davis, R. *Time Series: Theory and Methods.* Springer, New York (1991).

[7] Burkholder, D. L. Sharp inequalities for martingales and stochastic integrals. *Astérisque,* tome 157-158, p. 75–94 (1988).

[8] Cerovecki, C. & Hörmann, S. On the CLT for discrete Fourier transforms of functional time series. *Journal of multivariate analysis,* 154:282–295 (2017).

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

[10] Dehling, H. & Sharipov, O.S. Estimation of mean and covariance operator for Banach space valued autoregressive processes with dependent innovations *Statistical Inference for Stochastic Processes,* 2:137–149 (2005).

[11] van Delft, A. & Eichler, M. Locally stationary functional time series. *Electronic Journal of Statistics,* 12(1), 107–170 (2018).

[12] van Delft, A. & Eichler, M. A note on Herglotz’s Theorem for time series on the function space. *Stochastic processes and their applications* (2019), https://doi.org/10.1016/j.spa.2019.10.006

[13] Gohberg, I., Goldberg, S. & Kaashoek, M.A. *Classes of Linear Operators Vol. 1.* Birkhäuser, Verlag (1990).

[14] Grenander, U. *Abstract inference.* Wiley, New York (1981).

[15] Hörmann, S., Kidziński, L. & Hallin, M. Dynamic functional principal components. *The Journal of the Royal Statistical Society: Series B,* 77:319–348 (2015).

[16] Hörmann, S. & Kokoszka, P. Weakly dependent functional data. *The Annals of Statistics,* 38:1845–1884 (2010).

[17] Hörmann, S., Kokoszka, P & Nisol, G. Detection of periodicity in functional time series. *The Annals of Statistics,* forthcoming. (2017).

[18] Horváth, L. & Kokoszka, P. *Inference for Functional Data with Applications.* Springer, New York (2012).

[19] Horváth, L., Kokoszka, P & Reeder, R. Estimation of the mean of functional time series and a two-sample problem. *The Journal of the Royal Statistical Society: Series B,* 75:103–122 (2013).

[20] Gordin M. I. The central limit theorem for stationary processes. *Soviet. Math. Dokl.,* 10:1174–1176 (1969).

[21] de Jong, P. A central limit theorem for generalized quadratic forms. *Probability Theory and Related Fields,* 75:261–277 (1987).
[22] Kadison, R.V. & Ringrose, J. R.  *Fundamentals of the Theory of Operator Algebras. Graduate Studies in Mathematics*. Amer. Math. Soc., Providence, RI (1997).

[23] Kokoszka, P. & Mohammadi Jouzdani, N. Frequency domain theory for functional time series: variance decomposition and an invariance principle. *Researchgate Preprint* (2019).

[24] Kokoszka, P. & Taqqu, M. S. The asymptotic behaviour of quadratic forms in heavy-tailed strongly dependent random variables. *Stochastic Processes and their Applications*, 66:21–40 (1997).

[25] Lee, J. & Subba Rao, S. A note one general quadratic forms of nonstationary stochastic processes. *American Journal of Theoretical and Applied Statistics*, 51:949–968 (2017).

[26] Leucht, A., Paparoditis, E. & Sapatinas, T. Testing equality of spectral density operators for functional linear processes. *arXiv:1804.03366* (2018).

[27] Liu, W. & Wu, W. B. Asymptotics of spectral density estimates. *Econometric Theory*, 26:1218–1245 (2010).

[28] Mas, A. & Menneteau, L. Perturbation Approach Applied to the Asymptotic Study of Random Operators. *in: High Dimensional Probability III, vol. 55, Birkhäuser*, pp. 127–133 (2003).

[29] Mikosch, T. Functional limit theorems for random quadratic forms. *Stochastic processes and their applications*, 37:81–98 (1991).

[30] Panaretos, V. & S. Tavakoli. Fourier analysis of stationary time series in function space. *The Annals of Statistics* 41: 568–603 (2013a).

[31] Panaretos, V. & S. Tavakoli. Cramér–Karhunen–Loève representation and harmonic principal component analysis of functional time series. *Stochastic Processes and their Applications* 123: 2779–2807 (2013b).

[32] Peligrad, M. & Wu, W. B. Central limit theorem for Fourier transforms of stationary processes. *The Annals of Probability*, 38(5):2009–2022 (2010).

[33] Pham, T. and Panaretos, V. Methodology and convergence rates for functional time series regression. *Statistica Sinica*, 28:2521–2539 (2018).

[34] Pourahmadi, M. *Foundations of Time Series Analysis and Prediction Theory*. Wiley, New York (2001).

[35] Rice, J. & Silverman, B. Estimating the mean and covariance structure nonparametrically when the data are curves. *The Journal of the Royal Statistical Society: Series.*, 53:233–243 (1991).

[36] Suquet, C. Tightness in Schauder decomposable Banach spaces. *Translations–American Mathematical Society*, 193:201–224 (1996).

[37] Wu, W. B. & Shao, X. A Limit Theorem for Quadratic Forms and Its Applications. *Econometric Theory*, 23(5):930–951 (2007).

[38] Wu, W. B. Fourier transforms of stationary processes. *Proc. Amer. Math. Soc.*, 133:285–293 (2005).
Lemma A.1. \[ \text{By Burkholder’s inequality (31)} \]

A Inequalities for $H$-valued martingales and linear transforms

Let $H$ be a Hilbert space. For a probability space $(\Omega, \mathcal{F}, \mathcal{G}_\infty, P)$ and $\mathcal{G} = \{\mathcal{G}_t\}_{t \geq 0}$ a non-decreasing sequence of sub-$\sigma$-fields of $\mathcal{G}_\infty$, let $\{M_t\} \in \mathcal{L}_H^p$ be a martingale with respect to $\mathcal{G}$ and note that we can write $M_n = \sum_{k=0}^n D_k$, where $\{D_k\}$ denotes its difference sequence. Additionally denote the variable

\[ V_n(M) = \left( \sum_k \|D_k\|_{H}^2 \right)^{1/2} \]

which we call the square function of $M$. It was shown [7, theorem 3.1] that for $H$-valued martingales, we have for $1 < p < \infty$

\[ (p^* - 1)^{-1} (\mathbb{E}|V(M)|^p)^{1/p} \leq (\mathbb{E} \|M\|_{H}^p)^{1/p} \leq (p^* - 1)(\mathbb{E}|V(M)|^p)^{1/p} \tag{31} \]

where $p^* = \max(p, \frac{p}{p-1})$. As a consequence we have the following lemma, which extends lemma 1 of [37].

**Lemma A.1.** Let $\{M_k\}_{k=1,...,n} \in \mathcal{L}_H^p$, $p > 1$, be a martingale with respect to $\mathcal{G}$ with $\{D_k\}$ denoting its difference sequence and let $\{A_k\}_{k=1,...,n} \in S_\infty(H)$. Then, for $q = \min(2, p)$,

\[ \left\| \sum_{k=1}^n A_k(D_k) \right\|_{H, p}^q \leq K^q_{\text{p}} \sum_{k=1}^n \|A_k\|_{\infty}^q \|D_k\|_{H, p}^q \]

where $K^q_{\text{p}} = (p^* - 1)^q$ with $p^* = \max(p, \frac{p}{p-1})$.

**Proof of Lemma A.1.** By Burkholder’s inequality (31)

\[ \left\| \sum_{k=1}^n A_k(D_k) \right\|_{H, p}^q = \left( \mathbb{E} \left( \sum_{k=1}^n A_k(D_k) \right)^p \right)^{q/p} \leq (p^* - 1)^q \left( \mathbb{E} \left( \sum_{k=1}^n \|A_k(D_k)\|_{H}^2 \right)^{1/2} \right)^{q/p}. \]

Let $p < 2$. Then, applying the inequality $|\sum_{k} x_k|^r \leq \sum_{k} |x_k|^r$ for $r < 1$ to $x_k = \|A_k(D_k)\|_{H}^2$, we obtain

\[ (p^* - 1)^q \left( \mathbb{E} \left( \sum_{k=1}^n \|A_k(D_k)\|_{H}^2 \right)^{1/2} \right)^{q/p} \leq (p^* - 1)^q \left( \mathbb{E} \left( \sum_{k=1}^n \|A_k(D_k)\|_{H}^p \right)^{q/p} \right) \]

\[ \leq \left( \sum_{k=1}^n \|A_k\|_{\infty}^p \mathbb{E} \|D_k\|_{H}^p \right)^{q/p} \]

\[ \leq (p^* - 1)^q \sum_{k=1}^n \|A_k\|_{\infty}^q \|D_k\|_{H, p}^q, \]

[39] Wu, W. B. Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences.*, 102:14150–14154 (2005).

[40] Zhang, D. & Wu, W.B. Asymptotic Theory for Estimators of High-Order Statistics of Stationary Processes. *IEEE Transactions on Information Theory*, 64:4097–4922 (2018).

[41] Zhang, X. White noise testing and model diagnostic checking for functional time series. *Journal of Econometrics.*, 194:76–95 (2016).
where the one before last inequality follows from Holder’s inequality for operators and from the fact that \( p = q \). For \( p \geq 2, q = 2 \). Therefore, an application of Minkowski’s inequality to \( \| \cdot \|_{c,p/q} \) and Holder’s inequality yield in this case

\[
(p^* - 1)^q \left( \mathbb{E} \left( \sum_{k=1}^{n} A_k(D_k) \right)^{q/p} \right)^{1/q} \leq (p^* - 1)^q \left( \sum_{k=1}^{n} A_k(D_k) \right)^{q/p} \leq (p^* - 1)^q \left( \sum_{k=1}^{n} A_k \mathbb{E} \| D_k \|_{H}^{p} \right)^{q/p}.
\]

**Lemma A.2.** For \( t = 1, \ldots, n \), let \( \{X_t\} \) be a zero-mean stationary ergodic process in \( \mathcal{L}_H^p \) and \( \{A_t\} \in S_{\infty}(H) \). Then,

\[
\begin{align*}
(i) \quad & \left\| \sum_{t=1}^{n} A_t(X_t) \right\|_{H^p} \leq K_p \| A_n \|_{\ell_q} D_{p,1,0}, \\
(ii) \quad & \left\| \sum_{t=1}^{n} A_t X_t^{(m)} \right\|_{H^p} \leq K_p \| A_n \|_{\ell_q} \Delta_{p,1,0}, \\
(iii) \quad & \left\| \sum_{t=1}^{n} A_t (X_t - X_t^{(m)}) \right\|_{H^p} \leq K_p \| A_n \|_{\ell_q} \Delta_{p,1,m+1}.
\end{align*}
\]

where \( \Delta_{p,q,m} = \sum_{j=m}^{\infty} \nu_{H^p,j,m} \) and \( \| A_n \|_{\ell_q} = \sum_{t=1}^{n} \| A_t \|_{\ell_q} \).

**Proof.** Using \[3\] and Lemma A.1 (i) directly follows. For (ii), by stationarity

\[
\| P_j(X_t^{(m)}) \|_{H^p} = \mathbb{E} \| X_{t-j} - X_{t-j,0} \|_{H^p} \leq \nu_{H^p,j}(X_{t-j})
\]

where we abbreviated \( X_{t-j,0} = \mathbb{E}[X_{t-j} | \mathcal{G}_{t-j,0}] \) and therefore (ii) follows from (i). Finally, if we write \( X_t - X_t^{(m)} = \sum_{j=1}^{\infty} \mathbb{E}[X_t | \mathcal{G}_{t-j}] - \mathbb{E}[X_t | \mathcal{G}_{t-j+1}] \) then \( D_{t,j} := \mathbb{E}[X_t | \mathcal{G}_{t-j}] - \mathbb{E}[X_t | \mathcal{G}_{t-j+1}] \) for \( t = n, \ldots, 1 \) defines a martingale difference with respect to the backward filtration \( \mathcal{G}(\epsilon_t, \ldots, \epsilon_i), i = 0, -1, \ldots \) (iii) now follows from noting by the contraction property and stationarity

\[
\| D_{t,j} \|_{H^p} = \mathbb{E} \| (X_t - X_{t,t-j}) | \mathcal{G}_{t,j} \|_{\ell_1} \leq \mathbb{E} \| (X_t - X_{t,t-j}) \|_{H^p} = \nu_{H^p,j}(X_j).
\]

**Proof of Proposition 3.2.** (i) Since the process \( \{X_t^{(m)}\} \) is \( m \)-dependent it is immediate that the \( D_{m,k} \) are also \( m \)-dependent. Hence, we may write \( D_{m,k}^{(t)} = \sum_{t=0}^{m} P_0(X_t^{(m)}) e^{-it\lambda} \). By orthogonality, \( \| D_{m,k} \|_{H}^{2} \leq \sum_{t=0}^{\infty} \| \mathbb{E}[P_0(X_t)] \|_{H}^{2} < \infty \). Next, observe that

\[
\mathbb{E}[D_{m,k}^{(t)} | \mathcal{G}_{k-1}] = \frac{1}{2\pi} \sum_{t=0}^{\infty} \mathbb{E}[\mathbb{E}[X_{t+k}^{(m)} | \mathcal{G}_{k}] - \mathbb{E}[X_{t+k}^{(m)} | \mathcal{G}_{k-1}] | \mathcal{G}_{k-1}] e^{-it\lambda} = 0
\]

by the properties of the conditional expectation.

(ii) Under Assumption 3.1 with \( p = 4 \), we obtain from Lemma A.1

\[
\mathbb{E}[D_{m,k}^{(t)} \otimes D_{m,k}^{(t)}]^{2} = \mathbb{E}[D_{m,k}^{(t)}]^{4} \leq \left( \sum_{t=0}^{\infty} \| P_0(X_t^{(m)}) \|_{H}^{2} \right)^{2}.
\]
Recall that \( T_r \) yields where we used that \( \{ Y \} \) have \( E_n \subseteq 2^{D} \) the limit satisfies \( \int \). By stationarity and by the fact that the integral of the complex exponential yields the constraint \( -D \) \( G \) \( Y \) \( h \) \( T \) \( n \) \( 2 \) \( \| Y \|_4 \). Now observe that for all \( n_1, n_2 \in \mathbb{N} \)

\[
\mathbb{E} \left\| Y_{n_2} \otimes Y_{n_2} - Y_{n_1} \otimes Y_{n_1} \right\|_4^2 \leq 2 \mathbb{E} \left\| (Y_{n_2} - Y_{n_1}) \otimes Y_{n_2} \right\|_4^2 + 2 \mathbb{E} \left\| (Y_{n_2} - Y_{n_1}) \otimes (Y_{n_2} - Y_{n_1}) \right\|_4^2 \\
\leq 2 \mathbb{E} \left\| (Y_{n_2} - Y_{n_1}) \right\|_4^2 \mathbb{E} \| Y_{n_2} \|_4^2 + 2 \mathbb{E} \left\| (Y_{n_2} - Y_{n_1}) \right\|_4^2 \mathbb{E} \left\| (Y_{n_2} - Y_{n_1}) \right\|_4^2 \\
\leq 2 \mathbb{E} \left\| (Y_{n_2} - Y_{n_1}) \right\|_4 \mathbb{E} \left\| Y_{n_2} \right\|_4 \mathbb{E} \left\| (Y_{n_2} - Y_{n_1}) \right\|_4 \\
\leq 4(\epsilon N)^{1/2},
\]

where we used that \( \{ Y \} \) is Cauchy in \( \mathcal{L}_4 \) from which it follows that for all \( \epsilon > 0 \) there exists an \( N \) such that for \( n_1, n_2 \geq N \), \( \mathbb{E} \left\| (Y_{n_2} - Y_{n_1}) \right\|_4 < \epsilon \) and \( \mathbb{E} \left\| Y_n \right\|_4 < N \). Next we prove (iii). First we need to prove that

\[
\lim_{m \to \infty} \lim_{T \to \infty} \text{Tr} \left( \text{Var}(D^{(A)}_{m,0,T}) \right) = \text{Tr}(\mathcal{F}(A)) < \infty. \tag{34}
\]

Recall that \( \text{Tr}(\text{Var}(D^{(A)}_{m,0,T})) = \mathbb{E} \left\| D^{(A)}_{m,0,T} \right\|_H^2 \) where the latter is finite uniformly in \( m \) and \( T \) because the limit satisfies \( \mathbb{E} \left\| D^{(A)}_{0,T} \right\|_H^2 < \infty \) by property (ii). We shall therefore proceed similar to \[32, 8\].

By stationarity and by the fact that the integral of the complex exponential yields the constraint \( t - s = h \)

\[
\int_{-\pi}^{\pi} \mathbb{E} \left\| D^{(A)}_{m,0,T} \right\|_H^2 e^{i t \omega} d\omega = \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathbb{E} \left( \sum_{t=0}^{T} P_0(X^{(m)}_t), \sum_{s=0}^{T} P_0(X^{(m)}_s) \right) e^{i(t-s)\omega} d\omega \\
= \mathbb{E} \sum_{t=h}^{T} \langle P_0(X^{(m)}_t), P_0(X^{(m)}_{t-h}) \rangle.
\]

Since \( \mathcal{G}_t \subseteq \mathcal{G}_h \forall t \geq h \), the properties of the conditional expectation show that, for any \( m \geq 1 \), we have \( \mathbb{E} \left[ X^{(m)}_t | \mathcal{G}_h \right] \equiv \mathbb{E} \left[ X^{(m)}_0 | \mathcal{G}_h \right], \forall t \geq h \). Moreover, \( X^{(m)}_h \) is \( \mathcal{G}_h \)-measurable. Therefore, we obtain by orthogonality of the projection operators and stationarity that

\[
\mathbb{E} \sum_{t=h}^{T} \langle P_0(X^{(m)}_t), P_0(X^{(m)}_{t-h}) \rangle = \mathbb{E} \sum_{t=h}^{T} P_0(X^{(m)}_t), \sum_{s=h}^{T} P_0(X^{(m)}_s) \\
= \mathbb{E} \sum_{t=h}^{T} P_{-s}(\mathbb{E}[X^{(m)}_0 | \mathcal{G}_h]), \sum_{s=h}^{T} P_{-s}(\mathbb{E}[X^{(m)}_h | \mathcal{G}_h])).
\]

By ergodicity and from (ii) \( \left\{ D^{(A)}_{m,0,T} \right\}_{T \geq 1} \) is Cauchy in \( \mathcal{L}_2 \). Thus, \( \lim_{T \to \infty} \sum_{t=h}^{T} P_{-s}(\mathbb{E}[X^{(m)}_0 | \mathcal{G}_h]) \equiv \mathbb{E}[X^{(m)}_0 | \mathcal{G}_h] \) and \( \lim_{T \to \infty} \sum_{s=h}^{T} P_{-s}(\mathbb{E}[X^{(m)}_h | \mathcal{G}_h]) \equiv X^{(m)}_h \). Therefore, continuity of the inner product yields

\[
\lim_{T \to \infty} \mathbb{E} \sum_{t=h}^{T} P_{-s}(\mathbb{E}[X^{(m)}_0 | \mathcal{G}_h]), \sum_{s=h}^{T} P_{-s}(\mathbb{E}[X^{(m)}_h | \mathcal{G}_h]))
\]
\[
\mathbb{E}(\mathbb{E}[X_0^{(m)}|\mathcal{G}_h], X_0^{(m)}) = \mathbb{E}(X_0^{(m)}, X_0^{(m)}) = \text{Tr}(C_h^{(m)}),
\]
where we used the tower property. Hence, \( \lim_{T \to \infty} \frac{1}{T} \mathbb{E}[D(h)^{(\lambda)}_{m,0,T}]^2 e^{ihx} d\lambda = \text{Tr}(C_h^{(m)}) \). But this holds in particular for \( m = \infty \), i.e., for the process \( \lim_{m \to \infty} X^m_t = X_t \). Now observe that the conditions of the classical Féjer-Lebesgue theorem are satisfied and therefore

\[
\lim_{T \to \infty} \text{Tr}(\mathcal{F}^{(\lambda)}_T) = \lim_{T \to \infty} \sum_{\lambda \leq T} (1 - \frac{h}{T})\mathbb{E}(X_h, X_0) e^{-ihx} = \mathbb{E}[D(h)^{(\lambda)}_0]^2 = \text{Tr}(\mathcal{F}^{(\lambda)}) < \infty, \tag{35}
\]
where we used again property (ii) in order to obtain the finite trace. Let \( \mathcal{D}_{m,T}^{\omega} \) denotes the functional DFT of \( X_t^{(m)} \). Clearly, we have immediately from the above as well that

\[
\lim_{m \to \infty} \lim_{T \to \infty} \text{Tr}(\mathcal{D}_{m,T}^{\omega}) = \lim_{m \to \infty} \mathbb{E}[D(h)^{(\lambda)}_m]^2 = \lim_{m \to \infty} \text{Tr}(\mathcal{F}^{(\lambda)}_m) = \text{Tr}(\mathcal{F}^{(\lambda)}) \tag{36}
\]
where \( 2\pi \mathcal{F}^{(\lambda)}_m = \sum_{\lambda \leq m} \mathbb{E}(X_0^{(m)} \otimes X_0^{(m)}) e^{-ihx} \) and where we applied the dominated convergence theorem which is justified by (33). This proves (34). Consequently, non-negative definiteness allows us to conclude that \( \mathcal{F}^{(\lambda)}_m \in S_1(H) \) for all \( m \geq 1 \) and any \( \lambda \in (-\pi, \pi] \). Then, using the permutation operator is a unitary operator, Holders’ inequality for operators yields

\[
\left\| \Pi_{ijkl}(\mathcal{F}^{(\lambda)} \otimes \mathcal{F}^{(\lambda)}) \right\|_1 \leq \|\Pi_{ijkl}\|_\infty \|\mathcal{F}^{(\lambda)} \otimes \mathcal{F}^{(\lambda)}\|_1 \leq \|\mathcal{F}^{(\lambda)}\|_2^2 = (\mathbb{E}[D(h)^{(\lambda)}_0]^2)^2 \leq \mathbb{E}[D(h)^{(\lambda)}_0]^4 < \infty, \tag{37}
\]
where we applied (35) in the equality and Jensen’s inequality together with property (ii) in the last inequality. From continuity of \( \mathcal{F}^{(\lambda)} \) and the dominated convergence theorem together with (36), we obtain

\[
\lim_{m \to \infty} \lim_{T \to \infty} \text{Tr}(\Pi_{ijkl} \mathcal{F}(\mathcal{D}_{m,0,T}^{(\lambda)} \otimes \mathcal{F}(\mathcal{D}_{m,0,T}^{(\lambda)})) = \text{Tr}(\Pi_{ijkl} \mathcal{F}^{(\lambda)} \otimes \mathcal{F}^{(\lambda)}) < \infty.
\]

\[
\square
\]

**B Joint convergence of finite-dimensional distributions of \( \xi_{T,m}^{(\lambda)} \)**

**Proof of Theorem 3.2** We recall that

\[
\xi_{T,m}^{(\lambda)} := (\mathbb{F}_T \|^{(2)}_F)^{-1/2} (\mathcal{M}_{T,m}^{(\lambda)} + \mathcal{M}_{T,m}^{(\lambda)}).
\]

We want to show that \( \{\xi_{T,m}^{(\lambda)}\}_{T,m} \) are converging jointly to complex Gaussian elements of \( S_2(H) \). From Proposition 3.3, we know that \( \xi_{T,m}^{(\lambda)} \) define martingales in \( \mathcal{L}^2_{S_2(H)}(\Omega, \mathcal{F}, \mathbb{P}) \) with respect to the filtration \( \{\mathcal{G}_T\} \). Below we shall prove convergence of the finite-dimensional distributions via a martingale central limit theorem on the linear combinations. To make this precise, let \( U = \{u_1, \ldots, u_d, v_1, \ldots, v_d \in H\} \). For any \( u, v \in H \) note that we can define the natural filtration of the process \( \{\xi_t(u_1, v_1)\}_{t} \) over \( (\Omega, \mathcal{F}, \mathbb{P}) \) by \( \{\mathcal{G}_t(u, v)\} \). In the following, we let \( \mathcal{G}_t(u_1, v_1) = \sigma(\{\xi_t(u_1, v_1)\}_{t \in \mathbb{Z}}) \) to be the natural filtration over \( (\Omega, \mathcal{F}, \mathbb{P}) \) of the projected process \( \{\xi_t \otimes \xi_s, u_j \otimes v_j \}_{t,s \in \mathbb{Z}} \). Correspondingly, we denote the projection operator \( P_0(u_j, v_j) := \mathbb{E}[\mathcal{G}_0(u_j, v_j)] - \mathbb{E}[\mathcal{G}_0(u_1, v_1)] \). More generally, let \( \mathcal{G}_t(U) = \sigma(\{\xi_t(u_1, v_1), \ldots, \xi_{t-d}(u_1, v_1), \ldots, \xi_{t-d}(u_1, v_1), (X_{t_2}, u)\}) \) for all \( t = t_1 \geq \ldots \geq t_2d \) and \( P_0(U) \) the corresponding projection operator. Observe then that

\[
\frac{1}{\mathbb{F}_T \|^{(2)}_F} \left( \mathcal{M}_{T,m}^{(\lambda)}(u_1, v_1) + \mathcal{M}_{T,m}^{(\lambda)}(u_1, v_1) \right)
\]
defines a well-defined martingale process in \( L^2_{c}(\Omega, \mathcal{A}, \mathbb{P}) \) with respect to the filtration \( \mathcal{G}_T(u, v) \).

In order to derive joint convergence of the finite-dimensional distributions it suffices to show that, for any \( a_1, \ldots, a_d \in \mathbb{R} \) and \( \lambda_i \pm \lambda_j \neq 0 \mod 2\pi \), the process

\[
\frac{1}{\|\Phi_T\|_F} \sum_{j=1}^{d} a_j \left( \langle \mathcal{M}^{(\lambda_j)}_{T,m}(u_j), v_j \rangle + \langle \mathcal{M}^{(\lambda_j)}_{T,m}(u_j), v_j \rangle \right)
\]

converges to a zero-mean complex normal random variable with covariance

\[
\sum_{j=1}^{d} a_j \langle G_m(u_j), v_j \rangle = 4\pi^2 \sum_{j=1}^{d} a_j \left( \langle \mathcal{F}^{(\lambda_j)}_m(v_j), v_j \rangle \langle u_j, \mathcal{F}^{(\lambda_j)}_m(u_j) \rangle + 1_{[0,\pi]}(\langle \mathcal{F}^{(\lambda_j)}_m(u_j), v_j \rangle \langle \mathcal{F}^{(\lambda_j)}_m(u_j), v_j \rangle) \right)
\]

and pseudocovariance

\[
\sum_{j=1}^{d} a_j \langle \Sigma_m(u_j), v_j \rangle = 4\pi^2 \sum_{j=1}^{d} a_j \left( 1_{[0,\pi]}(\langle \mathcal{F}^{(\lambda_j)}_m(u_j), v_j \rangle \langle \mathcal{F}^{(\lambda_j)}_m(u_j), v_j \rangle) + \langle \mathcal{F}^{(\lambda_j)}_m(u_j), v_j \rangle \langle \mathcal{F}^{(\lambda_j)}_m(u_j), v_j \rangle \right).
\]

Note that this process is adapted to the filtration \( \mathcal{G}_T(U) \). We shall do this by means of the Cramér-Wold device. We first decompose the functional processes \( \mathcal{M}^{(\lambda_j)}_{T,m} \) as

\[
\mathcal{M}^{(\lambda_j)}_{T,m} = \sum_{t=1}^{T} D^{(\lambda_j)}_{m,t} \otimes \sum_{s=1}^{t-1} \phi^{(\lambda_j)}_{T,s-t} D^{(\lambda_j)}_{m,t} + \sum_{s=1}^{t-1} \phi^{(\lambda_j)}_{T,s-t} D^{(\lambda_j)}_{m,t}.\]

The following lemma shows the second sum is of lower order in norm.

**Lemma B.1.** Under the conditions of Theorem 3.1

\[
\left\| \sum_{t=1}^{T} D^{(\lambda_j)}_{m,t} \otimes \left( \sum_{s=1}^{t-1} \phi^{(\lambda_j)}_{T,s-t} D^{(\lambda_j)}_{m,s} \right) \right\|_{S_{2,2}} = o(\|\Phi_T\|_F).
\]

This implies in turn that we can focus on the distributional properties of the projections of the operators

\[
\sum_{t=4m+1}^{T} D^{(\lambda_j)}_{m,t} \otimes N^{(\lambda_j)}_{m,t} \quad \text{and} \quad \left( \sum_{t=4m+1}^{T} D^{(\lambda_j)}_{m,t} \otimes N^{(\lambda_j)}_{m,t} \right)^{\dagger},
\]

where

\[
N^{(\lambda_j)}_{m,t} := \sum_{s=1}^{t-4m} \phi^{(\lambda_j)}_{T,s-t} D^{(\lambda_j)}_{m,s}.\]

From Proposition 3.3 it is immediate that both terms in (38) constitute well-defined martingales in \( L^2_{S_1[H]}(\Omega, \mathcal{A}, \mathcal{G}_T, \mathbb{P}) \). Consequently, projecting these on fixed \( u, v \in H \), we obtain the following two martingale processes with paths in \( \mathbb{C} \)

\[
\left\langle \sum_{t=4m+1}^{T} D^{(\lambda_j)}_{m,t} \otimes N^{(\lambda_j)}_{m,t}, v \otimes u \right\rangle_s = \sum_{t=4m+1}^{T} \langle D^{(\lambda_j)}_{m,t}, v \rangle \langle N^{(\lambda_j)}_{m,t}, u \rangle, \]

\[
\left\langle \left( \sum_{t=4m+1}^{T} D^{(\lambda_j)}_{m,t} \otimes N^{(\lambda_j)}_{m,t} \right)^{\dagger}, v \otimes u \right\rangle_s = \sum_{t=4m+1}^{T} \langle D^{(\lambda_j)}_{m,t}, u \rangle \langle N^{(\lambda_j)}_{m,t}, v \rangle.
\]

In order to apply a martingale central limit theorem on the sum of (40) and (41) and over \( j = 1, \ldots, d \), we must verify Lindeberg’s condition is satisfied. Without loss of generality we do this
for (40) and for fixed $u, v$ as the result is immediate to carry over to a finite sum over $j$. To ease notation in the following, we set $(D_{m,t}^{(\lambda)}(x)) := D_{m,t}^{(\lambda)}(x)$ and $(N_{t}^{(\lambda)}(x)) := N_{t}^{(\lambda)}(x)$ for any $x \in H$. Recall the inequality $\mathbb{E}[\|Y\|^2_{L^1}1_{\|Y\| > \epsilon}] \leq \frac{1}{\epsilon^2}\mathbb{E}[\|Y\|^4_{L^1}]$ which holds for any $Y \in L^4$. Hence applying this for $H = \mathbb{C}$, we have

$$\sum_{t=1+4m}^{T} \mathbb{E}\left\{ |D_{m,t}^{(\lambda)}(v)N_{t}^{(\lambda)}(u)|^2 \right\} \leq \frac{1}{\epsilon^2} \sum_{t=1+4m}^{T} \mathbb{E}[|D_{m,t}^{(\lambda)}(v)N_{t}^{(\lambda)}(u)|^4].$$

Lindeberg’s condition is therefore satisfied if we can show that the term on the right hand side is of order $o(\|\Phi_T\|^4_{F})$. Since the $|D_{m,t}^{(\lambda)}|$ are $m$-dependent and by definition of (40) $|\Phi_T| \leq 4m$, $D_{m,t}^{(\lambda)}$ and $N_t$ are independent. Therefore, using Lemma A.1 with $H = \mathbb{C}$ yields

$$\sum_{t=1+4m}^{T} \mathbb{E}[|D_{m,t}^{(\lambda)}(v)N_{t}^{(\lambda)}(u)|^4] \leq \|D_0\|^4_{L^4} \|\Phi_T\|^4_{F} \sum_{t=1+4m}^{T} \mathbb{E}[N_{t}^{(\lambda)}(u)N_{t}^{(\lambda)}(u)]^2 \leq \|D_0\|^4_{L^4} \|\Phi_T\|^4_{F} \sum_{t=1+4m}^{T} (\sum_{l=1+4m}^{T} \|N_{t}^{(\lambda)}(u)\|^2_{L^2})^2 \leq \|D_0\|^4_{L^4} \|\Phi_T\|^4_{F} \sum_{t=1+4m}^{T} \mathbb{E}[|\Phi_T|] \sum_{j=1}^{d} a_j \Gamma_j = O(T\|\Phi_T\|^4_{F})$$

and similarly for (41), showing that Lindeberg's condition is satisfied. It therefore remains to verify that the conditional variance satisfies

$$\frac{1}{\|\Phi_T\|^2_{F}} \sum_{t=1+4m}^{T} \mathbb{E}\left[ \left( \sum_{j=1}^{d} a_j D_{m,t}^{(\lambda)}(v) N_{t}^{(\lambda)}(u) + D_{m,t}^{(\lambda)}(u) N_{t}^{(\lambda)}(v) \right)^2 \right] \leq \sum_{j=1}^{d} a_j \Gamma_j$$

and that the conditional pseudocovariance satisfies

$$\frac{1}{\|\Phi_T\|^2_{F}} \sum_{t=1+4m}^{T} \mathbb{E}\left[ \left( \sum_{j=1}^{d} a_j D_{m,t}^{(\lambda)}(v) N_{t}^{(\lambda)}(u) + D_{m,t}^{(\lambda)}(u) N_{t}^{(\lambda)}(v) \right)^2 \right] \leq \sum_{j=1}^{d} a_j \Sigma^{\lambda}_{m}.$$  

Moreover, observe that we can write $\mathbb{E}(\cdot|\Phi_{t-1}^{(U)}) = \sum_{k=1}^{m} P_{t-k}^{(U)}(\cdot) + \mathbb{E}(\cdot|\Phi_{t-4m}^{(U)})$. We will show that the sum of projections are of lower order. For (44), orthogonality of the $P_{t-k}^{(U)}(\cdot)$ and the contraction property of the expectation give

$$\sum_{t=1+4m}^{T} \max_{j} \|D_{m,t}^{(\lambda)}(v)\|^4_{L^4} \|N_{t}^{(\lambda)}(u)\|^4_{L^4}$$

where we used again that $D_{m,t}^{(\lambda)}$ and $N_{t}^{(\lambda)}$ are independent for any $\lambda$ and where the order follows in a similar manner to (42). Furthermore, observe that, for any $x \in U$ and any $\lambda$, $N_{t}^{(\lambda)}(x)$ is $\Phi_{t-4m}$ measurable. The left-hand side of (43) therefore equals

$$\sum_{t=1+4m}^{T} \mathbb{E}\left[ \left( \sum_{j=1}^{d} a_j D_{m,t}^{(\lambda)}(v) N_{t}^{(\lambda)}(u) + D_{m,t}^{(\lambda)}(u) N_{t}^{(\lambda)}(v) \right)^2 \right] \Phi_{t-m-1}^{(U)} + O(1)$$
We require the following lemma.

**Lemma B.2.** Let \( \{D^{(i)}_{m,t}\} \in \mathcal{L}_H^4 \) be a \( H \)-valued martingale difference process. Then, provided that conditions (i) and (iv) of [Assumption 3.2](#) are satisfied, we have

\[
\left\| \frac{1}{\|T\|_F^2} \sum_{t=1}^{T} M^{(i)}_t \otimes M^{(j)}_t - \mathbb{E} M^{(i)}_0 \otimes M^{(j)}_0 \right\|_{S_{2,2}} = o(1),
\]

where \( M^{(i)}_t := \sum_{s=1}^{t-1} \phi^{(i)}_{t,s} D^{(i)}_{m,s} \).

Since norm convergence implies convergence in the weak operator topology, we obtain for any \( u, v \in H \)

\[
\left\| \frac{1}{\|T\|_F^2} \sum_{t=1}^{T} N^{(i)}_t(u)N^{(j)}_t(v) - \mathbb{E} N^{(i)}_t(u)N^{(j)}_t(v) \right\|_{c,2} = o(1).
\]

Therefore, we may replace them with their expectation in (43) and (44) in order to obtain, respectively, for (43)

\[
\frac{1}{\|T\|_F^2} \sum_{t=1}^{T} \sum_{j=1}^{d} a_j^2 (|\lambda_j^{(i)}(u)|^2 + |\lambda_j^{(i)}(v)|^2) + |\lambda_j^{(i)}(v)|^2 |\lambda_j^{(i)}(u)|^2 + N^{(i)}_t(u)N^{(i)}_t(v)D_{m,t}^{(i)}(u)D_{m,t}^{(i)}(v)
\]

while (44) becomes

\[
\frac{1}{\|T\|_F^2} \sum_{t=1}^{T} \sum_{j=1}^{d} a_j^2 (|\lambda_j^{(i)}(u)|^2 + |\lambda_j^{(i)}(v)|^2) + |\lambda_j^{(i)}(v)|^2 |\lambda_j^{(i)}(u)|^2 + N^{(i)}_t(u)N^{(i)}_t(v)D_{m,t}^{(i)}(u)D_{m,t}^{(i)}(v)
\]

\[
+ \sum_{i \neq j} a_i a_j (N^{(i)}_t(u)N^{(i)}_t(v)D_{m,t}^{(i)}(u)D_{m,t}^{(i)}(v)) + N^{(i)}_t(v)N^{(i)}_t(v)D_{m,t}^{(i)}(u)D_{m,t}^{(i)}(v)) + o_p(1),
\]
and for (44)

\[
\frac{1}{\|\Phi_T\|^2_F} \sum_{t=1+4m}^{T} \sum_{i,j=1}^{d} a_i^2 \left[ \mathbb{E}[N_t^{(\lambda_i)}(u_i)N_t^{(\lambda_j)}(v_j)]\mathbb{E}[D_{m,t}^{(\lambda_i)}(v_i)D_{m,t}^{(\lambda_j)}(u_j)] + \mathbb{E}[N_t^{(\lambda_i)}(v_i)N_t^{(\lambda_j)}(v_j)]\mathbb{E}[D_{m,t}^{(\lambda_i)}(u_i)D_{m,t}^{(\lambda_j)}(u_j)] \right] + 2\mathbb{E}[N_t^{(\lambda_i)}(u_i)N_t^{(\lambda_j)}(v_j)\mathbb{E}[D_{m,t}^{(\lambda_i)}(v_j)D_{m,t}^{(\lambda_j)}(u_j)]
\]

(45)

Next, we make use of the following auxiliary result.

**Lemma B.3.** Let \( \{D_{m,t}^{(\lambda)}\} \in \mathcal{L}_H^2 \) be a \( H \)-valued martingale difference process and let conditions (ii) and (iii) in Assumption 3.2 be satisfied. Furthermore, assume \( \lambda_1 \pm \lambda_2 \neq 0 \mod 2\pi \). Then for any \( u, v \in H \),

\[
\sum_{t=1+4m}^{T} \|\mathbb{E}[N_t^{(\lambda_1)}(u)N_t^{(\lambda_2)}(v)] = o(\|\Phi_T\|^2_F),
\]

where \( N_t^{(\lambda)} \) is as defined in (39).

Suppose first that \( d = 1 \). It follows from this Lemma B.3 that the third and fourth term of (45) and the first two terms of (46) will be of lower order if \( \lambda \neq 0, \pi \). Hence, from Proposition 3.2

\[
\mathbb{E}(D_{m,t}^{(\lambda)}(u)D_{m,t}^{(\lambda)}(v)) = \sum_{|k| \leq m} \langle c_k^{(m)}(v), u \rangle e^{-ikk} = 2\pi \langle \mathcal{F}_m^{(\lambda)}(v), u \rangle.
\]

If \( \lambda = 0 \mod \pi \), we also have \( \mathbb{E}(D_{m,t}^{(\lambda)}(u)D_{m,t}^{(\lambda)}(v) = 2\pi \langle \mathcal{F}_m^{(\lambda)}(v), (u) \rangle \). Note that the latter is real for \( v = u \). Hence, we obtain for (45) and (46)

\[
\frac{1}{\|\Phi_T\|^2_F} \sum_{t=1+4m}^{T} \mathbb{E}[N_t^{(\lambda_i)}(u)_i^{(\lambda_i)}(v)_j^{(\lambda_i)}(v)]\mathbb{E}[D_{m,t}^{(\lambda_i)}(v_i)D_{m,t}^{(\lambda_i)}(u_i)] + \mathbb{E}[N_t^{(\lambda_i)}(v_i)N_t^{(\lambda_i)}(v_j)]\mathbb{E}[D_{m,t}^{(\lambda_i)}(u_i)D_{m,t}^{(\lambda_i)}(u_j)] + 2\mathbb{E}[N_t^{(\lambda_i)}(u_i)N_t^{(\lambda_i)}(v_j)\mathbb{E}[D_{m,t}^{(\lambda_i)}(v_j)D_{m,t}^{(\lambda_i)}(u_i)]
\]

and

\[
\frac{1}{\|\Phi_T\|^2_F} \sum_{t=1+4m}^{T} \mathbb{E}(D_{m,t}^{(\lambda)}(u)D_{m,t}^{(\lambda)}(v))^2 + \mathbb{E}(N_t^{(\lambda_i)}(v_i)D_{m,t}^{(\lambda_i)}(u_i)) + \mathbb{E}(N_t^{(\lambda_i)}(v_i)N_t^{(\lambda_i)}(v_j))\mathbb{E}[D_{m,t}^{(\lambda_i)}(u_i)D_{m,t}^{(\lambda_i)}(u_j)]
\]

\[
= \frac{8\pi^2 \|\Phi_T\|^2_F}{\sum_{t=1+4m}^{T} \sum_{s=1}^{T-4m} w_{t-s}^2} \left( \langle \mathcal{F}_m^{(\lambda)}(v), u \rangle \langle \mathcal{F}_m^{(\lambda)}(u), u \rangle + 1_{[0,\pi]}(\langle \mathcal{F}_m^{(\lambda)}(v), u \rangle \langle \mathcal{F}_m^{(\lambda)}(v), u \rangle) \right)
\]

\[
- 4\pi^2 \left( \langle \mathcal{F}_m^{(\lambda)}(v), u \rangle \langle \mathcal{F}_m^{(\lambda)}(u), u \rangle + 1_{[0,\pi]}(\langle \mathcal{F}_m^{(\lambda)}(v), u \rangle \langle \mathcal{F}_m^{(\lambda)}(v), u \rangle) \right)
\]

(46)

respectively, as \( T \to \infty \). This completes the proof for \( d = 1 \). Next, suppose that \( d > 1 \). If \( \lambda_i \pm \lambda_j \neq 0 \mod 2\pi \), then by Lemma B.3 the cross terms are of lower order. Together this establishes the stated convergence in (43) and (44), respectively.
B.1 Proofs auxiliary statements

**Proof of Lemma B.2.** Let $Y_{T-1} = \sum_{t=2}^{T} M_{t-1}^{(A)} \otimes M_{t-1}^{(A)} - \|\Phi_T\|_2^2 E M_0^{(A)} \otimes M_0^{(A)}$. Observe that from the properties of $\{M_t^{(A)}\}$, the process $Y_{T-1}$ is $\mathcal{G}_{T-1}$ measurable, stationary and ergodic. Ergodicity of the underlying process and a telescoping argument allows us therefore to write

$$
\mathbb{E}\|Y_{T-1}\|_2^2 = \mathbb{E}\left\| \sum_{j=-\infty}^{0} P_j \left( \sum_{t=2}^{T} M_{t-1}^{(A)} \otimes M_{t-1}^{(A)} \right) \right\|_2^2 
= \sum_{j=-\infty}^{0} \mathbb{E}\|P_j \left( \sum_{t=2}^{T} M_{t-1}^{(A)} \otimes M_{t-1}^{(A)} \right)\|_2^2 + \mathbb{E}\sum_{j=1}^{T-1} \mathbb{E}\|P_j \left( \sum_{t=2}^{T} M_{t-1}^{(A)} \otimes M_{t-1}^{(A)} \right)\|_2^2,
$$

where we used orthogonality of the projection operators $\{P_j\}$. We first consider the first term on the right hand side for which we have $j < 1$. Since $\{D_{m,s}\}$ has uncorrelated increments

$$
P_j(M_{t-1}^{(A)} \otimes M_{t-1}^{(A)}) = \sum_{s,s' = 1}^{t-1} \sum_{j} \phi_{T,s,t}^{(A)} D_{m,s} \otimes \phi_{T,s',t}^{(A)} D_{m,s'} = \sum_{s = 1}^{t-1} \sum_{j} \phi_{T,s,t}^{(A)} D_{m,s} \otimes \phi_{T,s,t}^{(A)} D_{m,s},
$$

where we used that linear operators and expectation operators commute, i.e., $\mathbb{E}(AX \otimes AX) = \mathbb{E}((A \otimes A)(X \otimes X))$ for $A \in S_{\infty,H}$, $X \in \mathcal{L}_{H}$. Consequently, linearity, orthogonality of the projections, Minkowsk’s inequality and stationarity of $\{D_{m,s}\}$ yield

$$
\sum_{j=-\infty}^{0} \mathbb{E}\|P_j \left( \sum_{t=2}^{T} M_{t-1}^{(A)} \otimes M_{t-1}^{(A)} \right)\|_2^2 
\leq \sum_{j=-\infty}^{0} \left( \sum_{t=2}^{T} \mathbb{E}\|P_j \left( M_{t-1}^{(A)} \otimes M_{t-1}^{(A)} \right)\|_{S_{2,2}} \right)^2
= \sum_{j=-\infty}^{0} \left( \sum_{t=2}^{T} \left( \phi_{T,s,t}^{(A)} \otimes P_0(D_{m,s-j}^{(A)} \otimes D_{m,s-j}^{(A)}) \right) \right) \|_{S_{2,2}}^2
\leq \sum_{j=-\infty}^{0} \left( \sum_{t=2}^{T} \left( \phi_{T,s,t}^{(A)} \otimes P_0(D_{m,s-j}^{(A)} \otimes D_{m,s-j}^{(A)}) \right) \right) \|_{S_{2,2}}^2
\leq 0 \sum_{j=-\infty}^{0} \left( \sum_{t=2}^{T} \left( \phi_{T,s,t}^{(A)} \otimes P_0(D_{m,s-j}^{(A)} \otimes D_{m,s-j}^{(A)}) \right) \right) \|_{S_{2,2}}^2
$$

the Cauchy-Schwarz inequality implies we obtain under Assumption 3.1

$$
\sum_{j=-\infty}^{0} \left( \sum_{s=1}^{T-1} \left( P_0(D_{m,s-j}^{(A)} \otimes D_{m,s-j}^{(A)}) \right) \right) \|_{S_{2,2}}^2 \leq \sum_{j=-\infty}^{0} \left( \sum_{s=1}^{T-1} \left( \left[ \sum_{t=2}^{T} \left( \phi_{T,s,t}^{(A)} \otimes P_0(D_{m,s-j}^{(A)} \otimes D_{m,s-j}^{(A)}) \right) \right] \right) \right) \right)^{1/2} \left( \sum_{s=1}^{T-1} \left( \left[ \sum_{t=2}^{T} \left( \phi_{T,s,t}^{(A)} \otimes P_0(D_{m,s-j}^{(A)} \otimes D_{m,s-j}^{(A)}) \right) \right] \right) \right)^{1/2}
\leq \sum_{s=1}^{T-1} \left( \left[ \sum_{t=2}^{T} \left( \phi_{T,s,t}^{(A)} \otimes P_0(D_{m,s-j}^{(A)} \otimes D_{m,s-j}^{(A)}) \right) \right] \right) \right)^{1/2} \left( \sum_{s=1}^{T-1} \left( \left[ \sum_{t=2}^{T} \left( \phi_{T,s,t}^{(A)} \otimes P_0(D_{m,s-j}^{(A)} \otimes D_{m,s-j}^{(A)}) \right) \right] \right) \right)^{1/2}
= o(\|\Phi_T\|_F^2).
$$

For the second term of (47), i.e.,

$$
\sum_{j=1}^{T-1} \mathbb{E}\|P_j \left( \sum_{t=2}^{T} M_{t-1}^{(A)} \otimes M_{t-1}^{(A)} \right)\|_2^2,
$$

28
we have to distinguish cases since $1 \leq j \leq T - 1$. Firstly observe that if $1 \leq t \leq j$, then $P_j(M^{(1)}_{t-1} \otimes M^{(2)}_{t-1}) = O_H$ since $M^{(1)}_{t-1} \otimes M^{(2)}_{t-1}$ is $\mathcal{G}_{t-1}$ measurable and hence $E[M^{(1)}_{t-1} \otimes M^{(2)}_{t-1}|G_j] = M^{(1)}_{t-1} \otimes M^{(2)}_{t-1}$. We can thus focus on $t > j$. To ease notation, denote $D^{(1)}_{m,s} := D_s$. Since expectation and tensor operator commute, we obtain for the various cases:

- if $s_1 \leq j - 1$:
  - $s_2 > j$: $E[D_{s_1} \otimes D_{s_2}|\mathcal{G}_j] = D_{s_1} \otimes E[D_{s_2}|\mathcal{G}_j] = O_H$ and similarly $E[D_{s_1} \otimes D_{s_2}|\mathcal{G}_{t-1}] = O_H$ and therefore $P_j(D_{s_1} \otimes D_{s_2}) = O_H$.
  - $s_2 = j$: We have $E[D_{s_1} \otimes D_{s_2}|\mathcal{G}_j] = D_{s_1} \otimes E[D_{s_2}|\mathcal{G}_j] = D_{s_1} \otimes D_{s_2}$ while $E[D_{s_1} \otimes D_{s_2}|\mathcal{G}_{t-1}] = O_H$. Hence, $P_j(D_{s_1} \otimes D_{s_2}) = D_{s_1} \otimes D_{s_2}$.
  - $s_2 > j - 1$: We have again $P_j(D_{s_1} \otimes D_{s_2}) = O_H$.

- if $s_1 > s_2 \geq j$: using the tower property, we have

$$E[D_{s_1} \otimes D_{s_2}|\mathcal{G}_j] = E[E[D_{s_1} \otimes D_{s_2}|\mathcal{G}_{s_2}]|\mathcal{G}_j] = E[E[D_{s_1}|\mathcal{G}_{s_2}] \otimes D_{s_2}|\mathcal{G}_j] = O_H.$$  

Hence, 

$$P_j(M_{t-1} \otimes M_{t-1}) = \sum_{s_1=1}^{j-1} (\phi_{T,s_1,t} \otimes \phi_{T,j,t})(D_{s_1} \otimes D_j) + \sum_{s_2=1}^{j-1} (\phi_{T,j,t} \otimes \phi_{T,s_2,t})(D_j \otimes D_{s_2}) + \sum_{s=j+1}^{T-1} (\phi_{T,s,t} \otimes \phi_{T,s,t}) P_j(D^{(1)}_{m,s} \otimes D^{(1)}_{m,s}) =: U_j + U_j^\top + V_j$$

and therefore

$$\sum_{j=1}^{T-1} \mathbb{E}\left[\left\|P_j\left(\sum_{t=2}^{T} M_{t-1} \otimes M_{t-1}\right)\right\|^2\right] = \sum_{j=1}^{T-1} \mathbb{E}\left[\left\|P_j\left(\sum_{t=j+1}^{T} M_{t-1} \otimes M_{t-1}\right)\right\|^2\right] \leq \sum_{j=1}^{T-1} \mathbb{E}\left[\sum_{t=j+1}^{T} U_j + U_j^\top + V_j\right]^2$$

For the first term, stationarity of $\{D^{(1)}_{m,s}\}$, the properties of $\otimes$ and Lemma A.1 yield

$$\sum_{j=1}^{T-1} \mathbb{E}\left[\sum_{t=j+1}^{T} U_j\right]^2 = \sum_{j=1}^{T-1} \mathbb{E}\left[\sum_{t=j+1}^{T} \sum_{s_1=1}^{j-1} \phi_{T,s_1,t} \otimes \phi_{T,j,t}(D_{s_1} \otimes D_j)\right]^2$$

$$= \sum_{j=1}^{T-1} \mathbb{E}\left[\sum_{s_1=1}^{j-1} \sum_{t=j+1}^{T} \phi_{T,s_1,t} \otimes \phi_{T,j,t}(D_{s_1} \otimes D_j)\right]^2$$

$$= \sum_{j=1}^{T-1} \sum_{s_1=1}^{j-1} \sum_{t=j+1}^{T} \phi_{T,s_1,t} \otimes \phi_{T,j,t}(D_{s_1} \otimes D_j)^2 \mathbb{E}[D_{s_1}]_H^2 \mathbb{E}[D_j]_H^2$$

$$= K^2 \sum_{j=1}^{T-1} \sum_{s_1=1}^{j-1} \sum_{t=j+1}^{T} \phi_{T,s_1,t} \otimes \phi_{T,j,t}(D_{s_1} \otimes D_j)^2 \mathbb{E}[D_{s_1}]_H^2 \mathbb{E}[D_j]_H^2$$

$$= o(\|\Phi_T\|_F^2)$$

which follows from Assumption 3.2(iv). The same order applies to the second term. For the third term we find

$$\sum_{j=1}^{T-1} \sum_{t=j+1}^{T} V_j^2 = \sum_{j=1}^{T-1} \sum_{t=j+1}^{T} \sum_{s=j+1}^{T} \sum_{t=s+1}^{T} (\phi_{T,s,t} \otimes \phi_{T,s,t}) P_0(D^{(1)}_{m,s-j} \otimes D^{(1)}_{m,s-j})^2$$

29
For convenience denote $A_s = \| \sum_{j=1}^{T-1} (\phi_{T,s,t} \Theta \Phi_{T,s,t}) \|_\infty$ and $\theta_{s-j} = \| P_0(D^{(A_1)}_{m,s-j} \otimes D^{(A_2)}_{m,s-j}) \|_{S_2.12}$. Then we split the sum over $j$ in a sum with terms $1, \ldots, T - 1 - k_T$ and with terms $T - k_T, \ldots, T - 1$ where $k_T = \lfloor T^\alpha \rfloor$, $\alpha \in (0, 1)$. Additionally, we split the inner sum of the first. We then find via tedious calculations and the Cauchy-Schwarz inequality

\[
\begin{align*}
\sum_{j=1}^{T-1} \left( \sum_{s=j+1}^{T-1} A_s \theta_{s-j} \right)^2 & \leq \sum_{j=1}^{T-1} \left( \sum_{s=j+1}^{T-1} A_s \theta_{s-j}^2 \right) + 2 \sum_{j=1}^{T-1} \left( \sum_{s=j+1}^{T-1} A_s \theta_{s-j} \right)^2 \\
& \leq \left( \sum_{j=1}^{T-1} \sum_{s=j+1}^{T-1} A_s^2 \right)^{\frac{1}{2}} \left( \sum_{j=1}^{T-1} \sum_{s=j+1}^{T-1} \theta_{s-j}^2 \right) + 2 \left( \sum_{j=1}^{T-1} \sum_{s=j+1}^{T-1} A_s \theta_{s-j} \right)^2 \\
& \leq C \left( \sum_{j=1}^{T-1} \left( \sum_{t=s+1}^{T-1} \| (\phi_{T,s,t} \Theta \Phi_{T,s,t}) \|_\infty \right)^2 k_T + 2C (T k_T \eta_T^2) + 2 \sum_{j=1}^{T-1} \sum_{s=j+1}^{T-1} \| (\phi_{T,s,t} \Theta \Phi_{T,s,t}) \|_\infty^2 \sum_{s=k_T}^{\infty} \theta_{s-j}^2 \right)
\end{align*}
\]

for some generic bounded constant $C$, which follows from Assumption 3.1 and Assumption 3.2 i) for any $k_T \to \infty$ such that $\frac{k_T}{T} \to 0$ as $T \to \infty$.

**Proof of Lemma B.1.** By Lemma A.1 and Jensen's inequality, we obtain for fixed $m$,

\[
\begin{align*}
\sum_{t=2}^{T} \left\| D^{(A)}_{m, t} \otimes \left( \sum_{s=-4m+1}^{t-1} \phi^{(A)}_{T,s-t} D^{(A)}_{m,s} \right) \right\|_{S_2.12}^2 & \leq K^2_2 \| D^{(A)}_{m,0} \|_{H_4}^2 \sum_{t=2}^{T} \sum_{s=t-4m+1}^{t-1} \| (\phi^{(A)}_{T,s-t} \Theta \Phi_{T,s-t}) \|_\infty^2 \| D^{(A)}_{m,s} \|_{H_4}^2 \\
& = 2K^2_2 \| D^{(A)}_{m,0} \|_{H_4}^4 \left( \sum_{t=2}^{T} \sum_{s=t-4m+1}^{t-1} \| (\phi^{(A)}_{T,s-t} \Theta \Phi_{T,s-t}) \|_\infty^2 \right) + \sum_{t=4m+1}^{T} \sum_{s=t-4m+1}^{t-1} \| (\phi^{(A)}_{T,s-t} \Theta \Phi_{T,s-t}) \|_\infty^2 \\
& \leq 2K^2_2 \| D^{(A)}_{m,0} \|_{H_4}^4 \max_t (4m) \eta_A^4 + T4m \sum_{t=T} \| (\Theta \Phi_{T,s-t} \Theta \Phi_{T,s-t}) \|_\infty^2 = o(\| \Phi_{T} \|_F^2) + T o(\eta_T^2) = o(\| \Phi_{T} \|_F^2).
\end{align*}
\]

**Proof of Lemma B.3.** Denote $\lambda = \lambda_1 \pm \lambda_2$ and recall that $N^{(A)}_{m,t} = \sum_{s=1}^{T-4m} \phi^{(A)}_{T,s-t} D^{(A)}_{m,s}$. Since the increments of $\{ D^{(A)}_{m,s} \}$ are uncorrelated, we have

\[
\begin{align*}
\sum_{t=1}^{T} \left\| E(N^{(A)}_{m,t}, u) \otimes (N^{(A)}_{m,t}, v) \right\|_{S_2.12} & = \sum_{t=1}^{T} \left\| \sum_{s=1}^{T-4m} e^{\lambda_1 (s-t)} \langle A_{T,s-t}(D^{(A_1)}_{m,s}), u \rangle \langle A_{T,s-t}(D^{(A_2)}_{m,s}), v \rangle \right\|_{S_2.12} \\
& = \sum_{t=1}^{T} \left\| \sum_{s=1}^{T-4m} e^{\lambda_1 (s-t)} \langle A_{T,s-t}(D^{(A_1)}_{m,s}) \otimes A_{T,s-t}(D^{(A_2)}_{m,s}), u \otimes v \rangle \right\|_{S_2.12} \\
& \leq \sum_{t=1}^{T} \left\| \sum_{s=1}^{T-4m} e^{\lambda_1 (s-t)} \langle A_{T,s-t}(D^{(A_1)}_{m,s}) \otimes A_{T,s-t}(D^{(A_2)}_{m,s}), u \otimes v \rangle \right\|_{S_2.12}.
\end{align*}
\]
To ease notation, set \( W_{s-t} = \left( (A_{T,s-t} \otimes \overline{A}_{T,s-t}) \mathbb{E}(D^{(\lambda_1)}_{m,0} \otimes D^{(\lambda_2)}_{m,0}), u \otimes \overline{\tau} \right) \) and write \( B_j = \sum_{k=1}^{j} e^{ikj} \).

Summation by parts, and Holder’s inequality for operators yield
\[
\sum_{t=1+4m}^{T} \sum_{s=1}^{t-4m} W_{s-t}(B_s - B_{s-1}) \leq \sum_{t=1+4m}^{T} \left| W_{4m} B_{t-4m} + \sum_{s=1}^{t-4m-1} B_s \right| \left( \mathbb{E}(D^{(\lambda_1)}_{m,0} \otimes D^{(\lambda_2)}_{m,0}) \| u \otimes \overline{\tau} \|_2 \right) + \sum_{s=1}^{t-4m-1} \| B_s \| \left( \mathbb{E}(D^{(\lambda_1)}_{m,0} \otimes D^{(\lambda_2)}_{m,0}) \| u \otimes \overline{\tau} \|_2 \right) + \sum_{s=1}^{t-4m-1} \| B_s \| \left( \mathbb{E}(D^{(\lambda_1)}_{m,0} \otimes D^{(\lambda_2)}_{m,0}) \| u \otimes \overline{\tau} \|_2 \right)
\]

Then, using Jensen’s inequality and the Cauchy-Schwarz inequality twice, we obtain
\[
\leq C_1 \| D^{(\lambda_1)}_{m,0} \|_{\mathbb{H}_2} \| D^{(\lambda_2)}_{m,0} \|_{\mathbb{H}_2} \| 1/ \sin(\lambda/2) \| \left( \sum_{t=1+4m}^{T} \| A_{4m} \|_{\mathbb{H}_2}^2 + \sum_{t=1+4m}^{T} \sum_{s=1}^{t-4m-1} \| A_{T,s-t} \|_{\mathbb{H}_2} \| \overline{A}_{T,s-t} - \overline{A}_{T,s-t+1} \|_{\mathbb{H}_2} \right)
\]

where we used that max_{1 \leq t \leq T} |B_t| \leq |1/(\sin(\lambda/2))|.

\[\square\]

C  Operator approximations

Proof of Lemma 3.3. We can decompose the quadratic form
\[
\hat{\Phi}_T^A = \mathcal{Y}_T^A + \mathcal{Y}_T^{\lambda A} + \sum_{1 \leq t \leq T} \phi_{T,t}(X_t \otimes X_t)
\]

Set \( \hat{\mathcal{C}}_T = \sum_{t=1}^{T} X_t \otimes X_t \). Then, using linearity of the operator \( \Phi_{T,t} \),
\[
\left\| \hat{\Phi}_T^A - \mathcal{E} \hat{\phi}_T^A \right\|_{\mathbb{S}_2} \leq \left\| \mathcal{Y}_T^A + \mathcal{Y}_T^{\lambda A} - \mathcal{E} \mathcal{Y}_T^A - \mathcal{E} \mathcal{Y}_T^{\lambda A} \right\|_{\mathbb{S}_2} + \left\| \Phi_{T,t}(\hat{\mathcal{C}}_T - \mathcal{E} \hat{\mathcal{C}}_T) \right\|_{\mathbb{S}_2}.
\]

For the last term, Holder’s inequality for operators and Lemma C.1 yield
\[
\frac{1}{\| \Phi_T \|_F} \left\| \Phi_{T,t}(\hat{\mathcal{C}}_T - \mathcal{E} \hat{\mathcal{C}}_T) \right\|_{\mathbb{S}_2} \leq \frac{1}{\| \Phi_T \|_F} \left\| \Phi_{T,t} \|_{\infty} \right\| \frac{1}{\| \Phi_T \|_F} \left\| \mathcal{C}_T - \mathcal{E} \mathcal{C}_T \right\|_{\mathbb{S}_2} = o(\mathcal{C}_T) O\left( \sqrt{T} \| \Phi_T \|_F \right) = o(1),
\]

For the first term, we find using Lemma C.2 and Lemma C.3 respectively
\[
\frac{1}{\| \Phi_T \|_F} \left\| \mathcal{Y}_T^A - \mathcal{Y}_T^{\lambda A} - \mathcal{M}_{T,m}^{(\lambda)} \right\|_{\mathbb{S}_2} \leq \frac{1}{\| \Phi_T \|_F} \left\| \mathcal{Y}_T^A - \mathcal{Y}_T^{\lambda A} - (\mathcal{Y}_T^{(m)} \mathcal{A} - \mathcal{Y}_T^{(m)} \mathcal{A}) \right\|_{\mathbb{S}_2}
\]
\[
\frac{1}{\|\phi_T\|_F} \left\| \mathcal{M}_T^{(m),\lambda} - \mathbb{E}\mathcal{M}_T^{(m),\lambda} \right\|_{S_2,2} \\
\leq K_4 \gamma_{4,m} \sum_{t=0}^{\infty} \nu_{H,4}(X_t) \\
+ \left\| X_0 \right\|_{H,4}^2 \sqrt{T} \left( \max_{t \to \infty} \|A_{T,t}\|_F^2 + m \sum_{s=1}^{T} \|A_{T,s-t} - A_{T,s-(t+1)}\|_F^2 \right)^{1/2}.
\]

The result therefore follows from Assumption 3.2.

**Lemma C.1.** Let \( X_t \) satisfy Assumption 3.1 with \( p = 4 \). Then

\[
\left\| \sum_{1 \leq t \leq T} (X_t \otimes X_t) - T\mathbb{E}(X_0 \otimes X_0) \right\|_{S_2,2} = O(\sqrt{T} \sum_{j=0}^{\infty} \nu_{H,4}(X_j)) = O(\sqrt{T}).
\]

**Proof of Lemma C.1** By stationarity, ergodicity and orthogonality of the projection operators

\[
\left\| \sum_{1 \leq t \leq T} (X_t \otimes X_t) - T\mathbb{E}(X_0 \otimes X_0) \right\|_{S_2,2}^2 = \left\| \sum_{t=1}^{T} \sum_{j=0}^{T} P_j(X_t \otimes X_t) \right\|_{S_2,2}^2 \leq \sum_{t=1}^{T} \sum_{j=0}^{T} \|P_j(X_t \otimes X_t)\|_{S_2,2}^2
\]

Minkowski's inequality, the Cauchy-Schwarz inequality and stationarity imply

\[
\left\| \sum_{t=1}^{T} P_j(X_t \otimes X_t) \right\|_{S_2,2} = \sum_{t=1}^{T} \sum_{j=0}^{T} \|P_j(X_t) \otimes X_j \|_{S_2,2} \\
\leq \sum_{t=1}^{T} \left\| X_t \otimes X_t \right\|_{S_2,2} \\
\leq \sum_{t=1}^{T} \left\| X_t \right\|_{S_2,2}^4 \leq 2 \|X_0\|_{H,4}^2 \sum_{t=1}^{T} \nu_{H,4}(X_t).
\]

Consequently,

\[
\sum_{t=1}^{T} \left\| X_t \otimes X_t \right\|_{S_2,2}^2 \leq 4 \|X_0\|_{H,4}^2 \sum_{t=1}^{T} \nu_{H,4}(X_t) \leq 4 \|X_0\|_{H,4}^2 T \sum_{j=0}^{\infty} \nu_{H,4}(X_j)^2.
\]

The result follows by taking the square root.

**Lemma C.2** (M-dependence approximation). Suppose (5) with \( 2p \) is satisfied for some \( p \geq 2 \). Then

\[
\frac{1}{\sqrt{T}\|\phi_T\|_F} \sum_{t=0}^{\infty} \nu_{H,2p}(X_t) \leq K_p \gamma_{2p,m}
\]

where \( \nu_{2p,m} = 2 \sum_{t=0}^{\infty} \min(\nu_{H,2p}(X_t), \Delta_{2p,m+1}^{1/2}) \) and

\[
\gamma_T^{(m)} := \sum_{s=1}^{T} \sum_{t=1}^{s-1} \Phi_{T,s-t}^{(m)}(X_s \otimes X_t) \quad \text{and} \quad \mathcal{M}_T^{(m),\lambda} := \sum_{s=1}^{T} \sum_{t=1}^{s-1} \Phi_{T,s-t}^{(m)}(X_{s-t}^{(m)} \otimes X_t^{(m)}).
\]

---

32
Proof of Lemma C.2. Let $N^{(m)}_{t,s} = \sum_{k=1}^{s-1} \phi^{(k)}_{T,s-1}(X^{(m)}_t)$ and $N_{T,s} = \sum_{k=1}^{s-1} \phi^{(k)}_{T,s-1}(X_t)$ and observe these are $q_s$-measurable. By orthogonality of the projections and Minkowski’s inequality

$$\|Y_T - \mathbb{E}Y_T - (Y^{(m),A}_T - \mathbb{E}Y^{(m),A}_T, \mathbb{E}Y^{(m),A}_T)\|_{S_2,p} \leq 2 \sum_{j=-\infty}^{T} \|P_j(Y^{(m),A}_T - \mathbb{E}Y^{(m),A}_T, \mathbb{E}Y^{(m),A}_T)\|_{S_2,p} + \|P_j(Y^{(m),A}_T - Y^{(m),A}_T)\|_{S_2,p},$$

where $\mathbb{E}Y^{(m),A}_T = \sum_{s=2}^{T} X_s \otimes N^{(m)}_{T,s}$. We shall focus on bounding the first term as the second is similar and has the same upper bound. A similar trick as in Lemma A.2 shows

$$\mathbb{E}[Y_T^{(m),A} - \mathbb{E}Y^{(m),A}_T|q_j] = \mathbb{E}[Y_T^{(m),A} - \mathbb{E}Y^{(m),A}_T|q_j] = \mathbb{E}[Y^{(m),A}_T - Y^{(m),A}_T|q_j].$$

By the contraction property of the conditional expectation

$$\|P_j(Y^{(m),A}_T - Y^{(m),A}_T)\|_{S_2,p} \leq \|Y^{(m),A}_T - Y^{(m),A}_T\|_{S_2,p} \leq \sum_{s=2}^{T} \|X_s - X_{s,(j)}\|_{\mathbb{H}^2} \otimes (N_{T,s} - N^{(m)}_{T,s})\|_{S_2,p} + \sum_{s=2}^{T} X_s \otimes (N_{T,s} - N^{(m)}_{T,s} - N^{(m)}_{T,s,(j)})\|_{S_2,p},$$

where we added and subtracted $X_s \otimes (N_{T,s} - N^{(m)}_{T,s})$ and applied Minkowski’s inequality. From Lemma A.2 (iii)

$$\|N_{T,s} - N^{(m)}_{T,s}\|_{\mathbb{H}^2} \leq (K_p^q \|\phi_T\|_{\ell_q}^q \Delta^q_{2p,1,m+1})^{1/q}$$

and $\|X_s - X_{s,(j)}\|_{\mathbb{H}^2} \leq \nu_{\mathbb{H}^2}(X_{s,(j)})$. Then, by the Cauchy-schwarz inequality and recalling that $q = \min(2,2p)$

$$\sum_{j=-\infty}^{T} f_1 \leq \sum_{j=-\infty}^{T} \left( \sum_{s=2}^{T} \nu_{\mathbb{H}^2}(X_{s,(j)}) \right)^2 \left( (K_p^q \|\phi_T\|_{\ell_q}^q \Delta^q_{2p,1,m+1})^{1/2} \right)^2 \leq K_p^q \|\phi_T\|_{\ell_q}^q \Delta^q_{2p,1,m+1} \sum_{s=2}^{T} \sum_{j=-\infty}^{T} \nu_{\mathbb{H}^2}(X_{s,(j)}) \sum_{s=2}^{T} \nu_{\mathbb{H}^2}(X_{s,(j)} \Delta^q_{2p,1,m+1} \Delta^q_{2p,1,m+1})$$

where we used that $\sum_{s=2}^{T} \nu_{\mathbb{H}^2}(X_{s,(j)}) \Delta^q_{2p,1,m+1}$ and Minkowski’s inequality, we obtain

$$\|X_t - X^{(m)}_t + X^{(m)}_s - X_{s,(j)}\|_{\mathbb{H}^2} \leq \Delta^q_{2p,1,m+1} \Delta^q_{2p,1,m+1} \Delta^q_{2p,1,m+1} \Delta^q_{2p,1,m+1}$$

Hence, changing the order of summation and using property (3) of $\phi^{(k)}_{T,s}$ yields

$$\sum_{j=-\infty}^{T} f_1 \leq \sum_{j=-\infty}^{T} \left( \sum_{s=2}^{T} \sum_{t=1}^{s-1} X_{s,(j)} \otimes \phi^{(k)}_{T,s-1}(X_t - X^{(m)}_t + X^{(m)}_s - X_{s,(j)})\right)^2_{S_2,p}$$

$$\leq \sum_{j=-\infty}^{T} \left( \frac{1}{T} \sum_{t=1}^{T} \phi^{(-k)}_{T,s-1}(X_{s,(j)} \otimes (X_t - X^{(m)}_t + X^{(m)}_s - X_{s,(j)}))\right)^2_{S_2,p}$$

$$\leq 2 \min \left( \left( \sum_{j=m+1}^{\infty} \|D_{t,j}\|_{\mathbb{H}^2}^2 \right)^{1/2}, \nu_{\mathbb{H}^2}(X_{s,(j)}) \right).$$
Firstly, we consider $M$.

Note that $D$ martingale difference sequence in $L^2$.

**Proof of Lemma C.3.** By construction $\mathcal{M}(\lambda)$ as in Lemma C.2. Under Assumption 3.1 with $p = 4$, we have

$$
\|\nabla_T - \nabla_T\|_{S^2} \leq C\left(\max_{t} \|A_{T,t}\|_{\infty} + \left(\sum_{t=1}^{\infty} \|A_{T,t} - A_{T,t-1}\|_{\infty}^2\right)^{1/2}\right).
$$

**Lemma C.3** (martingale approximation to the $m$-dependent process). Let $\mathcal{M}(\lambda)$ as defined in [1] and $\nabla(m)$ as in Lemma C.2. Under Assumption 3.1 with $p = 4$, we have

$$
\|\nabla_T - \mathcal{M}(\lambda)\|_{S^2} \leq C\left(\max_{t} \|A_{T,t}\|_{\infty} + \left(\sum_{t=1}^{\infty} \|A_{T,t} - A_{T,t-1}\|_{\infty}^2\right)^{1/2}\right).
$$

**Proof of Lemma C.3.**

By construction $D_{m,k} \in \mathcal{L}_m^m$ defines an $m$-dependent martingale difference and therefore we can write $D_{m,k} = \sum_{r=0}^{m} P_r(X_{t+k}^m)e^{-ir\lambda}$ since the terms $t > m$ are zero. We decompose the difference $\nabla_T - \mathcal{M}(\lambda)$ as follows

$$
\sum_{t=2}^{T} X_t^m \otimes \sum_{s=1}^{t-1} \phi_{T,s-t}(X_s^m - D_{m,s}) + \sum_{t=2}^{T} (X_t^m - D_{m,t}) \otimes \sum_{s=1}^{t} \phi_{T,s-t}D_{m,s} + \sum_{t=2}^{T} (X_t^m - D_{m,t}) \otimes \sum_{s=1}^{t-1} \phi_{T,s-t}D_{m,s} := \sum_{t} M_t^* + Y_t + Z_t.
$$

Note that $\nabla_T - \mathcal{M}(\lambda)$ is then a martingale difference sequence in $\mathcal{L}_m^2$. Let $W_k = \sum_{t=0}^{m} E[X_{t+k}^m|\mathcal{G}_k]e^{-it\lambda}$ and observe that

$$
X_t^m = \sum_{t=0}^{m} E[X_{t+k}^m|\mathcal{G}_k]e^{-it\lambda} - \sum_{t=0}^{m} E[X_{t+k}^m|\mathcal{G}_k]e^{-it\lambda} = \sum_{t=0}^{m} E[X_{t+k}^m|\mathcal{G}_k]e^{-it\lambda} - \sum_{t=0}^{m} E[X_{t+k+1}^m|\mathcal{G}_k]e^{-i(t+1)\lambda} = W_k - E[W_{k+1}|\mathcal{G}_k]e^{-i\lambda}
$$

and that $D_{m,k} = W_k - E[W_k|\mathcal{G}_{k-1}]$. Therefore,

$$
\|\sum_{s=1}^{t-1} \phi_{T,s-t}(X_s^m - D_{m,s})\|_{H_2} \leq \|\sum_{s=1}^{t-1} A_{T,s-t}e^{-i(t-1)\lambda}(E[W_s|\mathcal{G}_{s-1}] - E[W_{s+1}|\mathcal{G}_s]e^{-i\lambda})\|_{H_2}
$$

$$
\|\sum_{s=1}^{t-1} A_{T,s-t}e^{-i(t-1)\lambda}E[W_s|\mathcal{G}_{s-1}] - e^{i(t-1)\lambda}E[W_{s+1}|\mathcal{G}_s]\|_{H_2}
$$

34
Set $V_s = e^{(t-s)A}E\{W_s|\mathcal{G}_{s-1}\}$. Summation by parts, Holder's inequality for operators and Lemma A.1 together yield

$$
\left\| \sum_{s=1}^{t-4m} A_{T,s-t}(V_s - V_{s+1}) \right\|_{\mathcal{H}_2} \leq \max_{t} \|A_{T,t}\|_{\infty} \| V_{t-4m} \|_{\mathcal{H}_2} + \left\| \sum_{s=1}^{t-4m} (A_{T,s-t} - A_{T,s-t-1})(\sum_{l=1}^{m} P_{s-l} V_l) \right\|_{\mathcal{H}_2} \leq \max_{t} \|A_{T,t}\|_{\infty} \| V_{t-4m} \|_{\mathcal{H}_2} + \sqrt{\sum_{l=1}^{m} \| A_{T,s-t} - A_{T,s-t-1} \|_{\infty}^2 (P_{s-l} V_l) \|_{\mathcal{H}_2}^2} \leq 2m \| X_0 \|_{\mathcal{H}_2} \max_{t} \| A_{T,t} \|_{\infty} + \left( \sum_{t=1}^{t-4m} \| A_{T,s-t} - A_{T,s-t-1} \|_{\infty}^2 \right)^{1/2} C m^{3/2} \| X_0 \|_{\mathcal{H}_2}.
$$

Hence, for some finite constant $C$,

$$
\| M^*_t \|_{\mathcal{H}_2} \leq C m \| X_0 \|_{\mathcal{H}_2}^2 \left( \max_{t} \| A_{T,t} \|_{\infty} + \left( \sum_{t=1}^{t-4m} \| A_{T,s-t} - A_{T,s-t-1} \|_{\infty}^2 \right)^{1/2} m \right),
$$

where we used the contraction property and stationarity to find $\| V_{t-4m} \|_{\mathcal{H}_2} \leq \| W_k \|_{\mathcal{H}_2} \leq 2m \| X_0 \|_{\mathcal{H}_2}$, and where we used that $(\sum_{t=1}^{m} \| P_{t-1} V_0 \|_{\mathcal{H}_2}^2)^{1/2} \leq (\sum_{t=1}^{m} \| P_{t-1} V_0 \|_{\mathcal{H}_2}^2)^{1/2} \leq \sqrt{2m} \| X_0 \|_{\mathcal{H}_2}$ since the $P_j(\cdot)$ form martingale differences. Consequently, a martingale decomposition of the sum gives

$$
\left\| \sum_{t=1}^{T} M^*_t \right\|_{S_2,2} \leq \sum_{t=1}^{m} \left\| \sum_{s=0}^{m} M^*_{s+4m} \right\|_{S_2,2} \leq 4m^{3/2} T^{1/2} C \| X_0 \|_{\mathcal{H}_2}^2 \left( \max_{t} \| A_{T,t} \|_{\infty} + \left( \sum_{t=1}^{t-4m} \| A_{T,s-t} - A_{T,s-t-1} \|_{\infty}^2 \right)^{1/2} m \right).
$$

For $Y_t$, we note that $\sum_{s=t-4m+1}^{t} \phi_{T,s}^{(1)}(X_s^{(m)} - D_{m,s})$ is $\mathcal{G}_{t-1}$ measurable and that $Y_t$ is $5m$-dependent. Therefore, via Minkowski's inequality, the Cauchy-Schwarz inequality and a similar decomposition as above shows that

$$
\left\| \sum_{t=1}^{T} Y_t - E Y_t \right\|_{S_2,2} = C \sqrt{T} m^{3/2} \| X_0 \|_{S_2}^2 \left( \max_{t} \| A_{T,t} \|_{\infty} + \sum_{t=1}^{T} \| A_{T,s} - A_{T,s-1} \|_{\infty}^2 m^{1/2} \right)
$$

and similarly for $\sum_{t=1}^{T} Z_t - E Z_t \|_{S_2,2}$.

\section*{D Proofs of Section 4}

\textbf{Proof of Theorem 4.1.} We first prove part (i) and consider the following bias variance decomposition

$$
\left\| \hat{\mathcal{F}}_T^{(1)} - \mathcal{F}^{(1)} \right\|_{S_2,p} \leq \left\| \hat{\mathcal{F}}_T^{(1)} - E \hat{\mathcal{F}}_T^{(1)} \right\|_{S_2,p} + \left\| E \hat{\mathcal{F}}_T^{(1)} - \mathcal{F}^{(1)} \right\|_{S_2,p}.
$$

We start with the first term. We decompose the error as

$$
\hat{\mathcal{F}}_T^{(1)} - E \hat{\mathcal{F}}_T^{(1)} = \frac{1}{T} \left( \sum_{s=2}^{T} X_s \otimes N_{T,s}^{(1)} - \sum_{s=2}^{T} X_s \otimes N_{T,s}^{(1)} + \sum_{s=2}^{T} X_s \otimes N_{T,s}^{(1)\dagger} - \sum_{s=2}^{T} X_s \otimes N_{T,s}^{(1)\dagger} \right) + \frac{1}{T} \left( \sum_{1 \leq t \leq T} A_{T,0}(X_t \otimes X_t) - \sum_{1 \leq t \leq T} A_{T,0}(X_t \otimes X_t) \right),
$$

(49)

(50)
where $N_{t,s}^{(1)} = \sum_{j=1}^{s} \phi_{t,s,1}^{(1)} X_j$. We first derive the order of (49) in $L^2_{S_2(\mathcal{H})}$. Using ergodicity and orthogonality of the projection operator, we can write

$$\left\| \sum_{t=1}^{T} X_t \otimes N_{t,s}^{(1)} - \sum_{t=1}^{T} X_t \otimes \phi_{t,s}^{(1)} \right\|_{S_2,p}^2 = \sum_{j=-\infty}^{T} \left\| P_j \left( \sum_{t=1}^{T} (X_t \otimes \phi_{t,s}^{(1)} X_j) \right) \right\|_{S_2,p}^2.$$ 

The contraction property of the conditional expectation and the Cauchy-Schwarz inequality imply

$$\left\| P_j \left( \sum_{t=1}^{T} \phi_{t,s}^{(1)} X_t \otimes X_s \right) \right\|_{S_2,p} \leq \left\| \sum_{t=1}^{T} \phi_{t,s}^{(1)} (X_t \otimes X_s) \right\|_{S_2,p} + \left\| \sum_{t=1}^{T} \phi_{t,s}^{(1)} (X_t - X_s) \otimes X_s \right\|_{S_2,p}.$$

Hence, from Lemma A.2, we obtain

$$\frac{1}{T^2} \sum_{j=-\infty}^{T} \left\| P_j \left( \sum_{t=1}^{T} \phi_{t,s}^{(1)} X_t \otimes X_s \right) \right\|_{S_2,p}^2 \leq \frac{1}{T^2} \sum_{j=-\infty}^{T} \left( \sum_{t=1}^{T} \phi_{t,s}^{(1)} \right)^2 K_2 p \max_{1 \leq s \leq T-1} \left\| \phi_{s,T} \right\|_p \Delta_{2,p,1,0}^2 + \frac{1}{T^2} \sum_{j=-\infty}^{T} \left\| \phi_{t,s}^{(1)} \right\|_p \Delta_{2,p,1,0}^2 \leq \frac{1}{T^2} 4 K_2 p \left( \max_{1 \leq s \leq T} \left\| \phi_{t,s}^{(1)} \right\|_p \right)^2 T \Delta_{2,p,1,0}^2 = O((b_T T)^{-1}).$$

From Lemma C.1, it is immediate that (49) is of order $O(T^{-1})$ in $L^2_{S_2}$. It therefore follows by Minkowski’s inequality that

$$\left\| \mathbf{\hat{F}}_T^{(1)} - \mathbf{\hat{F}}_T^{(1)} \right\|_{S_2,p} = O((b_T T)^{-1}).$$

Let us then consider the bias. Observe that from (23), stationarity yields

$$E \hat{F}^{(1)}_T = \frac{1}{2T} \sum_{s,t=1}^{T} \mathbb{E}(X_s \otimes X_t) w(b_T(t-s)) e^{i\lambda(t-s)} = \frac{1}{2\pi} \sum_{|h| \leq T} w(b_T h) \frac{1}{T} \min(T,T-h) \text{Cov}(X_{t+h} \otimes X_t) e^{-i\lambda h}$$

$$= \frac{1}{2\pi} \sum_{|h| \leq T} w(bh)(1 - |h| T) C_h e^{-i\lambda h}.$$

Hence, using Minkowski’s inequality we can bound the error by

$$\left\| E \hat{F}_T^{(1)} - \mathcal{F}_T^{(1)} \right\|_2 \leq \left\| \frac{1}{2\pi} \sum_{|h| < T} (w(b_T h) - 1) C_h e^{-i\lambda h} \right\|_2 + \left\| \frac{1}{2\pi} \sum_{|h| > T} C_h e^{-i\lambda h} \right\|_2.$$
Since \( \sum h \| C_h \| < \infty \), it is immediate that

\[
\sup_{\lambda} \| R_{1, \lambda} \|_2 \leq \sum_{|h| \geq T} \| C_h \|_2 \to 0 \text{ as } T \to \infty.
\]

Since \( w(\cdot) \) is bounded, the final term satisfies \( \sum_{h \in \mathcal{Z}} w(bh) \| C_h \|_2 \leq \sup_x |w(x)| \sum_{h \in \mathcal{Z}} \| C_h \|_2 < \infty \).

Hence, by Kronecker's lemma \( \sup_{\lambda} \| R_{2, \lambda} \|_2 \to 0 \) as \( T \to \infty \). Finally, provided that \( b_T \to 0 \) as \( T \to \infty \) and since \( \lim_{x \to 0} w(x) = w(0) \), we obtain \( w(b_T h) \to w(0) = 1 \) as \( T \to \infty \). Thus, \( \sup_{\lambda} \| R_{0, \lambda} \|_2 \to 0 \) from which asymptotic unbiasedness follows. Next, we prove part (ii) for which it remains to derive the order of the bias under the additional assumptions. Firstly, from Proposition 3.1 the assumption \( \sum_{h \in \mathcal{Z}} h \| P_0(X_h) \|_{b1,2} < \infty \) implies that \( \sum_{h \in \mathcal{Z}} h \| C_h \|_{b1,2} < \infty \). Observe then that we can decompose the bias as

\[
\| E \hat{\mathcal{F}}^{(\lambda)} T - \mathcal{F}^{\lambda} T \|_2 \leq \sum_{|h| < 1/b_T} \| w(b_T h) - 1 \| C_h \|_2 + \sum_{|h| \geq 1/b_T} \| C_h \|_2.
\]

For the final term, we use that \( \sup_{x} |w(x)| = O(1) \) and that \( \sum_h |h| \| C_h \| < \infty \) in order to obtain

\[
\sup_{\lambda} \| R_{2, \lambda} \|_2 = \sum_{|h| < 1/b_T} \| w(b_T h) - 1 \| C_h \|_2 \leq \sup_x |w(x)| \sum_{|h| < 1/b_T} \| C_h \|_2 = O(\frac{1}{T}).
\]

Moreover, \( \sup_{\lambda} \| R_{1, \lambda} \|_2 = O(b_T) \). Finally, since \( w(x) - 1 = O(x) \) as \( x \to 0 \) and \( \sum_h |h| \| C_h \| < \infty \), we find for the first term

\[
\sup_{\lambda} \| R_{0, \lambda} \|_2 = \sum_{|h| < 1/b_T} \frac{1}{2\pi} \| w(b_T h) - 1 \| C_h \|_2 \leq \sum_h \| w(b_T h) \| \| C_h \|_2 = O(b_T).
\]

Proof of Theorem 4.2. From (23), we have \( \hat{\mathcal{F}}^{\lambda} = (2\pi T)^{-1/2} \hat{\mathcal{F}}^{\lambda} \) with \( \phi_{T,(t-s)}^{(\lambda)} = w(b_T(t - s)) e^{i\lambda(t-s)}. \)

Note that in this case \( \| \Phi_T \|_F^2 := \sum_{t=1}^{T} \sum_{s=1}^{T} w^2(b_T(t - s)) \) and that \( \phi_{T}^{2} = \sum_{t=1}^{T} w^2(b_T(t)). \) For weight functions satisfying Assumption 4.1 a change of variables and symmetry of the weight function in zero yield

\[
\sum_{t=1}^{T} \sum_{s=1}^{T} |\phi_{T,(t-s)}|^2 = \sum_{t=1}^{T} \sum_{s=1}^{T} |w(b_T(t - s))e^{-i\lambda(t-s)}|^2 = \sum_{t=1}^{T} \sum_{s=1}^{T} w^2(b_T(t - s)) \sim \frac{T}{b_T} \int w^2(x) dx = \frac{T}{b_T} \kappa.
\]

(52)

It therefore suffices to verify the conditions of Theorem 3.1 which is given here for completeness but follows a standard argument [see e.g., 27]. From Assumption 4.1 it is obvious that (ii) of Assumption 3.2 holds. Additionally, from (52), it is immediate that \( \phi_{T}^2 = \kappa O(b_T^{-1}) \) and \( \| \Phi_T \|_F^2 = O(T \phi_{T}^2) \) so that (i) is also satisfied. For (iii), observe that we can write

\[
\sum_{t=1}^{M/b_T} |w(b_T t) - w(b_T(t - 1))|^2 + \sum_{t=M/b_T}^{T} |w(b_T t) - w(b_T(t - 1))|^2
\]

37
By assumption, the function \( w \) is bounded and continuous except on a set of measure zero. Observe that there must exist \( \varepsilon > 0 \) such that \( |w(b_T t) - w(b_T (t - 1))| < \varepsilon \) uniformly for \( |t| \leq M/b_T \), except for a finite number of points, \( \varepsilon/b_T \), for some constant \( \varepsilon > 0 \). Because \( w(\cdot) \) is bounded, the first term will thus be of order \( O(1/b_T) \). For the second term, the length of the interval, \( I_M \), converges to zero for fixed \( b_T \), \( T \) as \( M \to \infty \). Since the summand is at most of order \( b_T^{-1} \) under the stated conditions, the second term is of order \( O(I_M/b_T) = O(b_T^{-2}) \). To verify (iv), we decompose the term as

\[
\sum_{j=1}^{T-1} \sum_{s=1}^{j-M/b_T} \left( \sum_{t=j+1}^{T} w(b_T(s-t)) w(b_T(j-t)) \right)^2 + \sum_{j=1}^{T-1} \sum_{s=1}^{j-M/b_T} \left( \sum_{t=j+1}^{T} w(b_T(s-t)) w(b_T(j-t)) \right)^2
\]

For the first term, the Cauchy-Schwarz inequality and the condition \( \sup_{0 \leq s < 1} b \sum_{h \geq M/b} w^2(bh) \to 0 \) as \( M \to \infty \) yield

\[
\sum_{j=1}^{T-1} \sum_{s=1}^{j-M/b_T} \sum_{t=j+1}^{T} w^2(b_T(s-t)) \sum_{t=j+1}^{T} w^2(b_T(j-t)) = \sum_{j=1}^{T-1} \sum_{s=1}^{j-M/b_T} \sum_{t=j+1}^{T} w^2(b_T(s-t)) \sum_{t=j+1}^{T} w^2(b_T(j-t)) = O(T b_T^{-1} I_M b_T^{-1}) = o(||\Phi_T||^2_F).
\]

Secondly,

\[
\sum_{j=1}^{T-1} \sum_{s=1}^{j-M/b_T} \sum_{t=j+1}^{T} w^2(b_T(s-t)) \sum_{t=j+1}^{T} w^2(b_T(j-t)) = O(T b_T^{-1} \eta_T^4) = O(T b_T^{-3} = o(||\Phi_T||^2_F)
\]

where we used that \( 1/b_T = o(T) \).

\[\square\]

**Proof of Lemma 4.1.** Note that we can write the spectral density operator as

\[
2\pi T \hat{\Phi}^{(\lambda)} = \sum_{s,t=1}^{T} \Phi_{t,s,T}^{A}((X_s - \mu + \mu - \hat{\mu}) \otimes (X_t - \mu + \mu - \hat{\mu}))
\]

\[
= 2\pi T \hat{\Phi}^{(\lambda)} + \sum_{s,t=1}^{T} \Phi_{t,s,T}^{A}((\mu - \hat{\mu}) \otimes (X_t - \mu)) + \sum_{s,t=1}^{T} \Phi_{t,s,T}^{A}((X_s - \mu) \otimes (\mu - \hat{\mu}))
\]

(53)

\[
+ \sum_{s,t=1}^{T} \Phi_{t,s,T}^{A}((\mu - \hat{\mu}) \otimes (\mu - \hat{\mu})).
\]

(54)

We therefore will show that the last two terms in (53) and the term in (54) are of lower order. For the second term of (53), a change of variables, the properties of the tensor product and the Cauchy-Schwarz inequality yield

\[
\left\| \sum_{|h| < T} \sum_{t = \max(1,1-h)}^{\min(T,T-h)} w(b_T(h)) e^{ih\lambda}(\mu - \hat{\mu}) \otimes (X_t - \mu) \right\|_{S_{2,2}}
\]

\[
= \left\| \sum_{|h| < T} w(b_T(h)) e^{ih\lambda}(\mu - \hat{\mu}) \sum_{t = \max(1,1-h)}^{\min(T,T-h)} (X_t - \mu) \right\|_{S_{2,2}}
\]

\[
\leq \sum_{|h| < T} w(b_T(h)) \left\| \|\mu - \hat{\mu}\|_{H,4} \sum_{t = \max(1,1-h)}^{\min(T,T-h)} (X_t - \mu) \right\|_{S_{2,2}}^{1/2}
\]

\[
\leq \sum_{|h| < T} w(b_T(h)) \left\| O(T^{-1}) O(T) \right\|^{1/2} = O\left(\frac{1}{b_T}\right),
\]

38
where we used Lemma A.2(i) in order to obtain \(|\mu - \hat{\mu}|^2_{H,4} = \|\sum_{s,t=1}^T (X_t - E X_T)\|^2_{H,4} \leq \frac{1}{T} T_4(2,0) = O(1)\) and to obtain \(\|\sum_{s,t=1}^T (X_t - \mu)\|_{H,4}^2 = O(T)\). The third term in (53) is similar. For (54), Holder’s inequality and Lemma A.2(ii) yield

\[
\left\| \sum_{s,t=1}^T \Phi^j_{t,s,T} ((\mu - \hat{\mu}) \otimes (\mu - \hat{\mu})) \right\|_{S_2,2} \leq \sum_{s,t=1}^T \|\Phi^j_{t,s,T}\|_\infty \left( E\|\mu - \hat{\mu}\| (\mu - \hat{\mu})^2 \right)^{1/2} \\
\leq \sum_{s,t=1}^T \|\Phi^j_{t,s,T}\|_\infty \|\mu - \hat{\mu}\|^2_{H,4} = O(T b_T^{-1/2} T) = O\left( \frac{1}{b_T} \right)
\]

where we used that \(\sum_{s,t=1}^T \|\Phi^j_{t,s,T}\|_\infty = \sum_{s,t=1}^T \omega(b_T(t - s)) = O(T b_T^{-1})\). Hence, \(\|\hat{\Phi}^j(\lambda) - \hat{\Phi}^j(\lambda)\|_{S_2,2} = O(T^{-1} b_T^{-1})\) and the exact argument shows that \(\|\hat{\Phi}^j(\lambda) - \hat{\Phi}^j(\lambda)\|_{S_2,2} = O(T^{-1} b_T^{-1})\) for \(p \geq 2\) provided the process is in \(L^p\).

Proof of Proposition 4.1. Consistency of the empirical eigenvalues follows immediately from Theorem 4.1(i) together with the fact that sup \(j |\beta^j - \hat{\beta}^j| \leq \|\hat{\Phi}^j - \Phi^j\|_\infty \leq \|\hat{\Phi}^j - \Phi^j\|_2\) [see e.g., 13]. To prove the consistency of the eigenprojectors, denote \(\delta_j = \inf_{l \neq k} |\lambda_{X,j,k}^*(\omega) - \lambda_{X,j}^*(\omega)|\) and let \(A_{j,T} = \{\|\hat{\Phi}_T - \Phi^j\|_2 < \delta_j / 4\}\) occurs and zero otherwise. From Proposition 3.1 of [28]

\(\hat{\Pi}^j - \Pi^j = \mathcal{F}_j(\hat{\Phi}_T - \Phi^j) + R_{j,T}\)

where \(\mathcal{F}_j\) is a bounded linear operator for each \(j\) and where \(R_{j,T}\) is a random \(S_2(H)\)-valued element which satisfies \(\|R_{j,T}\|_2 \leq 8 \delta_j^{-1} \|\hat{\Phi}_T - \Phi^j\|_2\). The result now follows again from Theorem 4.1(i) since \(P(\{\hat{\Phi}_T \to \Phi^j\}) \to 0\).

E Proofs of Section 5

Lemma E.1. Let \(X, Y \in L^p_H\) and set \(Z = \otimes_{i=1}^k Z_i\) with \(Z_i \in L^p_H\) for \(i = 1, \ldots, k\) for \(k \in \mathbb{N}\). If \(Z\) is \(\mathcal{G}_j\)-measurable, then

\[
E[ P_j(X) \otimes Y \otimes Z ] = E[ P_j(X) \otimes P_j(Y) \otimes Z ].
\]  

Proof. By linearity, \(\mathcal{G}_j\)-measurability of \(P_j(X)\) and \(Z\), respectively and the tower property

\[
E[ P_j(X) \otimes P_j(Y) \otimes Z ] = E[E[P_j(X) \otimes Y \otimes Z | \mathcal{G}_j]] = E[E[P_j(X) \otimes Y | \mathcal{G}_j-1] \otimes Z | \mathcal{G}_j-1]]
\]

The result follows then from noting that \(P_j(\cdot)\) is a martingale difference sequence with respect to \(\mathcal{G}_j\).

Proof of Theorem 5.1. It suffices to consider \(t_{k-1} \geq t_{k-2} \geq \ldots \geq t_1 \geq 0\) since

\[
\sum_{t_1, \ldots, t_{k-1} \in Z} \| \text{cum}(X_{t_1}, \ldots, X_{t_{k-1}}, X_0) \|_2 = k! \sum_{t_{k-1} \geq t_{k-2} \geq \ldots \geq t_1 \geq 0} \| \text{cum}(X_{t_1}, \ldots, X_{t_{k-1}}, X_0) \|_2.
\]

We use (26) and first consider the set \(|\nu_1| = k\). We may write

\[
E[ X_{t_1} \otimes \cdots \otimes X_{t_{k-1}} \otimes X_0 ] = \sum_{j_0 = -\infty}^{0} \sum_{j_1 = -\infty}^{t_1} \cdots \sum_{j_{k-1} = -\infty}^{t_{k-1}} E[ P_{j_0}(X_{t_1}) \otimes \cdots \otimes P_{j_{k-2}}(X_{t_{k-1}}) \otimes P_{j_{k-1}}(X_0) ],
\]  

(56)
where the equality holds in Hilbert-Schmidt norm. Observe that the term will be zero unless the two largest elements of the set \(J := \{j_0, \ldots, j_{k-1}\}\) are equal; otherwise we can apply to the tower property and condition on a filtration for which all but one element are measurable. Without loss of generality, we consider the case \(j_{k-2} = j_{k-3} \geq \ldots \geq j_1 \geq j_0\). In this case, \(65\) becomes

\[
\sum_{j_0 = -\infty}^{0} \sum_{j_1 = -\infty}^{t_1} \cdots \sum_{j_{k-3} = -\infty}^{t_{k-3}} \sum_{j_{k-2} = -\infty}^{t_{k-2}} \mathbb{E}[P_{j_1}(X_{t_1}) \otimes \cdots \otimes P_{j_{k-2}}(X_{t_{k-2}}) \otimes P_{j_{k-2}}(X_{t_{k-1}}) \otimes P_{j_0}(X_0)].
\]

Applying Lemma E.1 iteratively, we can write,

\[
\|\mathbb{E}[P_{j_1}(X_{t_1}) \otimes \cdots \otimes P_{j_{k-2}}(X_{t_{k-2}}) \otimes P_{j_{k-2}}(X_{t_{k-1}}) \otimes P_{j_0}(X_0)]\|_2
\]

\[
= \mathbb{E}[P_{j_0}(X_0) \otimes P_{j_1}(P_{j_2}(X_{t_2}) \otimes P_{j_3}(X_{t_3})) \otimes P_{j_4}(X_{t_4})] \|_2.
\]

To ease the exposition, we focus on \(k = 4\). Denote \(P_{j_{i \mid t}}(\cdot) = \mathbb{E}[\cdot | \mathcal{G}_{j_{i \mid t}}] - \mathbb{E}[\cdot | \mathcal{G}_{j_{i-1 \mid t}}]\). By Jensen’s inequality, the last term is then bounded by

\[
\leq \|P_{j_0}(X_0)\|_{H,4} = A \cdot \|P_{j_0}(X_0)\|_{H,4},
\]

where we used the notation \(\|\cdot\|_{H^{n, \ell}} = (\mathbb{E} \cdot \|\cdot\|_{H^{n, \ell}})^{1/\ell} \geq 0\). By Jensen’s inequality, \(A\) is bounded by

\[
\leq \|P_{j_1}(X_{t_1}) \otimes P_{j_2}(P_{j_3}(X_{t_2}) \otimes P_{j_4}(X_{t_3})) - P_{j_1}(P_{j_2}(X_{t_2}) \otimes P_{j_3}(X_{t_3})) \otimes P_{j_4}(X_{t_4})\|_{H,4}
\]

\[
\leq \|P_{j_1}(X_{t_1}) - P_{j_1,j_0}(X_{t_1})\|_{H,4} \|P_{j_2}(P_{j_3}(X_{t_2}) \otimes P_{j_4}(X_{t_3}))\|_{H,4}
\]

\[
+ \|P_{j_1,j_0}(X_{t_1})\|_{H,4} \|P_{j_2}(P_{j_3}(X_{t_2}) \otimes P_{j_4}(X_{t_3})) - P_{j_1,j_0}(P_{j_2}(X_{t_2}) \otimes P_{j_3}(X_{t_3}))\|_{H,4}
\]

\[
\leq \|P_{j_1,j_0}(X_{t_1})\|_{H,4} \cdot J + \|P_{j_1}(X_{t_1}) - P_{j_1,j_0}(X_{t_1})\|_{H,4} \|P_{j_2}(X_{t_2}) - P_{j_2,j_0,j_1}(X_{t_2})\|_{H,4}
\]

\[
+ \|P_{j_2,j_1}(X_{t_2})\|_{H,4} \|P_{j_3}(X_{t_3}) - P_{j_2,j_1,j_0,j_1}(X_{t_3})\|_{H,4},
\]

where \(J\) is given by

\[
\|P_{j_2}(X_{t_2}) \otimes P_{j_3}(X_{t_3}) - P_{j_2,j_1}(X_{t_2}) \otimes P_{j_2,j_1,j_0}(X_{t_3}) - P_{j_2,j_1,j_0}(X_{t_2}) \otimes P_{j_3}(X_{t_3}) + P_{j_2,j_0,j_1,j_0}(X_{t_3})\|_{H,4}
\]

Using that

\[
x_1y_1 - x_2y_2 - x_3y_3 + x_4y_4 = (x_1 - x_2 - x_3 + x_4)y_1 + x_4(y_4 - y_3 - y_2 + y_1)
\]

\[
+ (x_4 - x_2)(y_2 - y_1) + (x_3 - x_4)(y_1 - y_2),
\]

we can write \(J = J_1 + J_2 + J_3 + J_4\) with

\[
J_1 \leq \|P_{j_2}(X_{t_2}) - P_{j_2,j_1}(X_{t_2}) - P_{j_2,j_1,j_0}(X_{t_2}) + P_{j_2,j_0,j_1,j_0}(X_{t_2})\|_{H,4} \|P_{j_2}(X_{t_2})\|_{H,4}
\]

\[
J_2 \leq \|P_{j_2}(X_{t_2}) - P_{j_2,j_1}(X_{t_2}) - P_{j_2,j_1,j_0}(X_{t_2}) + P_{j_2,j_0,j_1,j_0}(X_{t_2})\|_{H,4} \|P_{j_2}(X_{t_2})\|_{H,4}
\]

\[
J_3 \leq \|P_{j_2}(X_{t_2}) - P_{j_2,j_1,j_0}(X_{t_2})\|_{H,4} \|P_{j_2}(X_{t_2}) - P_{j_2,j_1,j_0}(X_{t_2})\|_{H,4}
\]

\[
J_4 \leq \|P_{j_2,j_0}(X_{t_2}) - P_{j_2,j_0,j_1}(X_{t_2})\|_{H,4} \|P_{j_2}(X_{t_2}) - P_{j_2,j_1}(X_{t_2})\|_{H,4},
\]
By the contraction property of the conditional expectation

\[ \left\| P_{j_1}(X_t) + \sum_{i=1}^{k-1} (-1)^i \sum_{1 \leq t_1 < \ldots < t_i \leq k} P_{j_1 \ldots j_i}(X_t) \right\|_{\mathcal{H},p} \leq \left\| X_t + \sum_{i=1}^{k} (-1)^i \sum_{1 \leq t_1 < \ldots < t_i \leq k} \mathbb{E}[X_t|\mathcal{F}(j_1 \ldots j_i)] \right\|_{\mathcal{H},p}, \]

and therefore

\[ \left\| \mathbb{E}\left[ P_{j_1}(X_t) \otimes P_{j_2}(X_t) \otimes P_{j_3}(X_t) \otimes P_{j_0}(X_0) \right] \right\|_{S_2(S_2(H))} \leq v_{\mathcal{H},4}(j_0) v_{\mathcal{H},4}(t_1 - j_1) \left\{ v_{\mathcal{H},4}(t_2 - j_2, t_2 - j_1, t_2 - j_0) v_{\mathcal{H},4}(t_3 - j_3, t_3 - j_1, t_3 - j_0) v_{\mathcal{H},4}(t_2 - j_2) 
+ v_{\mathcal{H},4}(t_2 - j_0, t_2 - j_1) v_{\mathcal{H},4}(t_3 - j_3, t_3 - j_0) v_{\mathcal{H},4}(t_3 - j_1, t_3 - j_2) \right\} \]

Denote \( \Delta_{\mathcal{H},4}(k) := \sum_{j_0, \ldots, j_{k-1} = 0}^{\infty} v_{\mathcal{H},4}(j_0, \ldots, j_{k-1}). \) Tiedious calculations then show that, using (29), we obtain

\[ \sum_{t_1 \geq t_2 \geq t_1 \geq 0} \left\| \Pi_{1324} \mathbb{E}[X_t \otimes X_{t_2}] \otimes \mathbb{E}[X_{t_1} \otimes X_0] \right\|_2. \]

By orthogonality of the projections and a similar derivation as above yields that this is bounded by

\[ \leq \sum_{t_1 \geq t_2 \geq t_1 \geq 0} \left\| \Pi_{1324} \mathbb{E}[P_{j_0}(X_{t_2}) \otimes P_{j_0}(X_0)] \otimes \mathbb{E}[P_{j_1}(X_{t_1}) \otimes P_{j_1}(X_{t_1})] \right\|_2 \]

\[ \leq \sum_{t_1 \geq t_2 \geq t_1 \geq 0} \left\| \sum_{j_0 = 0}^{\infty} v_{\mathcal{H},2}(t_2 + j_0) v_{\mathcal{H},2}(j_0) v_{\mathcal{H},2}(t_1 + j_1) v_{\mathcal{H},2}(t_3 + j_1) \right\| \]

\[ \leq \Delta_{\mathcal{H},2}^4 \Delta_{\mathcal{H},2}(2) < \infty. \]

The other terms are similar and the proof is therefore omitted. \( \Box \)

**Proof of Proposition 5.7** Observe that for any \( k \in \{1, \ldots, n\} \), Minkowski’s inequality implies

\[ v_{\mathcal{H},p}(j_1, \ldots, j_n) := \left\| X_0 + \sum_{i=1}^{n} (-1)^i \sum_{1 \leq l_1 < \ldots < l_i \leq n} \mathbb{E}[X_0|\mathcal{G}_0(-j_1, \ldots, -j_i)] \right\|_{\mathcal{H},p} \]

\[ \leq \left\| X_0 + \sum_{i=1}^{n-1} (-1)^i \sum_{l_1 < \ldots < l_i \in \{1, \ldots, n\} \setminus \{k\}} \mathbb{E}[X_0|\mathcal{G}_0(-j_1, \ldots, -j_i)] \right\|_{\mathcal{H},p} \]

\[ + \left\| \mathbb{E}[X_0|\mathcal{G}_0(-j_k)] + \sum_{i=1}^{n-1} (-1)^i \sum_{l_1 < \ldots < l_i \in \{1, \ldots, n\} \setminus \{k\}} \mathbb{E}[X_0|\mathcal{G}_0(-j_1, \ldots, -j_i)] \right\|_{\mathcal{H},p}, \]

41
since the second term is equal to the first, it then easily follows that

$$\nu_{X,H,p}(j_1, \ldots, j_n) \leq 2 \min_{1 \leq l_1 < \cdots < l_{n-1} \leq n} \nu_{X,H,p}(j_{l_1}, \ldots, j_{l_{n-1}}).$$

(58)

Applying (58) consecutively gives the result. \qed