

# Rate of convergence for Hilbert space valued processes

MORITZ JIRAK

*Institut für Mathematische Stochastik, Technische Universität Braunschweig, Pockelsstr. 14, 38106 Braunschweig, Germany. E-mail: m.jirak@tu-braunschweig.de*

Consider a stationary, linear Hilbert space valued process. We establish Berry–Esseen type results with optimal convergence rates under sharp dependence conditions on the underlying coefficient sequence of the linear operators. The case of non-linear Bernoulli-shift sequences is also considered. If the sequence is  $m$ -dependent, the optimal rate  $(n/m)^{1/2}$  is reached. If the sequence is weakly geometrically dependent, the rate  $(n/\log n)^{1/2}$  is obtained.

*Keywords:* Berry–Esseen; Hilbert space; linear process; weak dependence

## 1. Introduction

Let  $\{X_k\}_{k \in \mathbb{Z}}$  be a zero mean process taking values in a separable Hilbert space  $\mathbb{H}$  with inner product  $\langle \cdot, \cdot \rangle$  and norm  $\|\cdot\|_{\mathbb{H}}$ . A fundamental issue in probability theory and statistics is whether or not the central limit theorem holds for the partial sum  $S_n(X) = \sum_{k=1}^n X_k$ , that is, if we have

$$\frac{1}{\sqrt{n}} S_n(X) \xrightarrow{w} Z_{\Lambda}, \quad (1.1)$$

where  $Z_{\Lambda}$  denotes a centered Gaussian random variable with associated covariance operator

$$\Lambda(\cdot) = \mathbb{E}[\langle Z_{\Lambda}, \cdot \rangle Z_{\Lambda}].$$

Going one step further, we can ask ourselves about a possible rate of convergence in (1.1), that is, if

$$\lim_{n \rightarrow \infty} d(P_{S_n(X)/\sqrt{n}}, P_{Z_{\Lambda}}) \tau_n < \infty \quad \text{for a sequence } \tau_n \rightarrow \infty, \quad (1.2)$$

where  $d(\cdot, \cdot)$  is a probability metric, and  $P_X$  denotes the probability measure induced by the random variable  $X$ . The rate  $\tau_n$  can be considered as a measure of reliability for statistical inference based on  $S_n(X)$ , and large rates are naturally preferred. In the context of general Hilbert space valued processes, the notion of “probability of hitting a ball” has turned out to be a convenient formulation. More precisely, we consider the uniform metric over Balls, that is,

$$\Delta_n(\mu) = \sup_{x \in \mathbb{R}} |P(\|n^{-1/2} S_n(X) + \mu\|_{\mathbb{H}} \leq x) - P(\|Z_{\Lambda} + \mu\|_{\mathbb{H}} \leq x)|, \quad \mu \in \mathbb{H}, \quad (1.3)$$

where  $Z_\Lambda$  is a zero mean Gaussian random variable with associated covariance operator  $\Lambda$ . If  $\{X_k\}_{k \in \mathbb{Z}}$  is i.i.d. and real valued ( $\mathbb{H} = \mathbb{R}$ ), a huge literature has evolved around (1.2) in the past decades, see, for instance, [34]. Interestingly, if  $X_k$  lies in a general infinite dimensional Hilbert space  $\mathbb{H}$ , much less can be found in the literature. To some extent, this can certainly be attributed to the significantly higher complexity of the problem. While the first optimal results about the rate of convergence in real valued cases appeared around 1940 (cf. [4,16]), it took more than another 30 years until analogue results were obtained if  $\mathbb{H}$  is a general, infinite dimensional Hilbert space. Notable contributions here among others are [3,19,31,32,42] and [44]. For a more detailed account on the historic development, see [45]. More recently, weakly dependent Hilbert space valued process have attracted more attention in the statistical context of functional principal component analysis, see [24] and [25]. In this note, we are concerned with possibly dependent, stationary processes that can be represented as

$$X_k = g_k(\{\varepsilon_j\}_{j \in \mathbb{Z}}), \quad k \in \mathbb{Z}, \tag{1.4}$$

for measurable functions  $g_k$  and i.i.d. random variables  $\{\varepsilon_k\}_{k \in \mathbb{Z}} \in \mathbb{S}$  for some measure space  $\mathbb{S}$ . Such processes are often also referred to as (non-causal) Bernoulli-shift processes. Special emphasis is devoted to non-causal linear processes, that is, we assume that  $X_k$  can be represented as

$$X_k = \sum_{j \in \mathbb{Z}} \alpha_j(\varepsilon_{k+j}), \quad k \in \mathbb{Z}, \tag{1.5}$$

where  $\{\varepsilon_k\}_{k \in \mathbb{Z}} \in \mathbb{H}$  is a centered i.i.d. sequence with  $\mathbb{E}[\|\varepsilon_k\|_{\mathbb{H}}^2] < \infty$ . Note that this implies existence of the associated covariance operator  $\mathbf{C}^\varepsilon$ . The sequence  $\{\alpha_j\}_{j \in \mathbb{N}}$  denotes linear operators, which we endow with the usual operator norm

$$\|\alpha_j\|_{\mathcal{H}} = \sup_{x \in \mathbb{H}: \|x\|_{\mathbb{H}}=1} \|\alpha_j(x)\|_{\mathbb{H}}.$$

For notational convenience, we assume here that  $\alpha_j$  maps from  $\mathbb{H}$  to  $\mathbb{H}$ , but also two different Hilbert spaces are possible. Linear processes are among the first (possibly weak dependent) generalizations from the i.i.d. case, but already constitute a relevant class of processes which contains important examples from the time series literature, for instance (functional) autoregressive processes (cf. [7,24]). The CLT for linear processes in Hilbert spaces was investigated, among others, in [29], where it was shown that a CLT is valid if and only if

$$\sum_{j \in \mathbb{Z}} \|\alpha_j\|_{\mathcal{H}} < \infty, \tag{1.6}$$

see below for some more comments on this result. It seems that the first results about the rate of convergence for linear processes were considered in [7], where the special case of Hilbert space valued AR(1) processes was treated, and a rate of  $\sqrt{n}$  was reached. Some extensions with possible suboptimal rates are obtained in [8], see also the correction in [9]. In [14], the rate  $\sqrt{n}$  was obtained if  $\sum_{j \in \mathbb{Z}} |j| \|\alpha_j\|_{\mathcal{H}} < \infty$  and the sequence  $\{\varepsilon_k\}_{k \in \mathbb{N}}$  has bounded support, that is,  $P(\|\varepsilon_k\|_{\mathbb{H}} > C) = 0$  for some  $C > 0$ . More recently, [30] considered random fields in Hilbert

and Banach spaces. In the special case of real-valued sequences  $\alpha_j \in \mathbb{R}$ , Berry–Esseen type bounds are established if (1.6) holds. However, unlike to the previous results, the approximating Gaussian measure depends on  $n$  in general, which is different from our results.

Regarding non-linear sequences, the problem becomes more difficult. Certain martingale difference sequences in Banach spaces have been investigated in [1,10]. In [36] and [37] (see also [6]),  $m$ -dependent sequences in Banach spaces are studied, whereas [46] considers  $\varphi(n)$ -mixing sequences with geometric decay. Though some of these results are optimal or close to optimality in a certain way, they lead to (significantly) inferior rates for Hilbert space valued sequences, as was pointed out in [41]. [41] is a notable exception, where a convergence rate of  $\tau_n = n^{1/2}(\log n)^{-2}$  is obtained if the sequence  $\{X_k\}_{k \in \mathbb{Z}}$  is geometrically  $\varphi(n)$ -mixing and satisfies some additional regularity assumptions (cf. Section 3).

The aim of this note is twofold. In case of linear processes, we first give a Berry–Esseen result with optimal rate under sharp moment assumptions ( $p \in (2, 3)$ ) and dependence conditions. We also show that the convergence rate may be arbitrarily slow. For non-linear processes, we first study one-dependent Bernoulli-shift sequences and establish the optimal rate. Based on this result, we then consider  $m$ -dependent Bernoulli-shift sequences and causal, weakly dependent Bernoulli-shift sequences with geometric decay in the dependence. In the latter, we obtain a convergence rate of  $(n/\log n)^{1/2}$ . For  $m$ -dependent processes, we obtain the optimal rate  $(n/m)^{1/2}$ .

This note is structured as follows. In Sections 2 and 3, the main results are presented and discussed. Proofs are given in Section 4. Throughout the remainder, we make the following convention. For  $p \geq 1$ , denote with  $\|\cdot\|_p$  the  $L^p$ -norm  $\mathbb{E}[|\cdot|^p]^{1/p}$ . We write  $\lesssim, \gtrsim, (\sim)$  to denote (two-sided) inequalities involving a multiplicative constant. Given a set  $\mathcal{A}$ , we denote with  $\mathcal{A}^c$  its complement.

## 2. Main results: Linear processes

Let us first introduce some additional necessary notation. Denote with  $\{\xi_k\}_{k \in \mathbb{Z}} \in \mathbb{H}$  an i.i.d. sequence of centered Gaussian random variables, where we require that the covariance operators of  $\varepsilon_k$  and  $\xi_k$  are equal. We then consider the Gaussian counter part of  $X_k$ , namely

$$Z_k = \sum_{j \in \mathbb{Z}} \alpha_j (\xi_{k+j}), \quad k \in \mathbb{Z}.$$

For  $k \in \mathbb{Z}$ , we also introduce the following (linear) operators, mapping from  $\mathbb{H}$  to  $\mathbb{H}$ .

$$\begin{aligned} \mathbf{A} &= \sum_{j \in \mathbb{Z}} \alpha_j, & \mathbf{A}_{n,k} &= \sum_{j=-n+k}^{k-1} \alpha_j, \\ \mathbf{A}_{n,k}^c &= \sum_{j>n-k} \alpha_j + \sum_{j<1-k} \alpha_j, & \mathbf{\Lambda} &= \mathbf{A}^c \mathbf{A}^*, \end{aligned} \tag{2.1}$$

where  $\mathbf{B}^*$  denotes the adjoint of an operator  $\mathbf{B}$ . Note that we assign  $\mathbf{\Lambda}$  a more concrete form here, and indeed one readily verifies that for  $x \in \mathbb{H}$

$$\mathbf{A}^c \mathbf{A}^*(x) = \mathbb{E}[\langle Z_{\mathbf{\Lambda}}, x \rangle Z_{\mathbf{\Lambda}}] \quad \text{where } Z_{\mathbf{\Lambda}} = \mathbf{A}(\xi_0).$$

One of the fundamental tools when working with linear processes is the elementary and well-known Beveridge and Nelson decomposition (BND) (cf. [5])

$$\begin{aligned} \sum_{k=1}^n X_k &= \sum_{k=1}^n \sum_{j=-n+k}^{k-1} \alpha_j(\varepsilon_k) + \sum_{k>n} \sum_{j=-n+k}^{k-1} \alpha_j(\varepsilon_k) + \sum_{k<1} \sum_{j=-n+k}^{k-1} \alpha_j(\varepsilon_k) \\ &= \sum_{k=1}^n \mathbf{A}(\varepsilon_k) - \sum_{k=1}^n \mathbf{A}_{n,k}^c(\varepsilon_k) + \sum_{k>n} \mathbf{A}_{n,k}(\varepsilon_k) + \sum_{k<1} \mathbf{A}_{n,k}(\varepsilon_k). \end{aligned}$$

It should be mentioned though that related, much more general martingale decompositions have already appeared earlier in the literature, see, [18] and [22]. For the CLT,  $S_n(\varepsilon) = \sum_{k=1}^n \mathbf{A}(\varepsilon_k)$  is the relevant part in (4.2), and indeed we have that

$$n^{-1/2} S_n(\varepsilon) \xrightarrow{w} Z_{\mathbf{A}} \quad \text{if } \sum_{j \in \mathbb{Z}} \|\alpha_j\|_{\mathcal{H}} < \infty,$$

see, [29]. Unlike to the real-valued case, condition  $\sum_{j \in \mathbb{Z}} \|\alpha_j\|_{\mathcal{H}} < \infty$  is sharp in the sense that if it fails, no CLT can hold, even not under a possibly different normalization [29]. The corresponding counter example itself is set in the Gaussian domain, i.e.  $\varepsilon_k = \xi_k$ , and solely relies on properties of the constructed sequence of linear operators  $\alpha_j$ . Thus, to a good proportion, the question of Berry–Esseen type results for linear processes is intimately connected to distributional properties of Gaussian random variables in Hilbert spaces.

Here we use results from [42] (cf. Lemma 4.3), and particularly Lemma 4.7 as our main tools for the linear case. This requires us to impose some conditions on the eigenvalues of  $\mathbf{A}$ , which we denote with  $\{\lambda_j\}_{j \in \mathbb{N}}$ . Throughout, we assume without loss of generality that  $\lambda_j \geq 0$ . We then derive our main results under the following assumptions.

**Assumption 2.1.** *For some  $2 < p \leq 3$  it holds that*

- (i)  $\mathbb{E}[\varepsilon_k] = 0$  and  $\mathbb{E}[\|\varepsilon_k\|_{\mathbb{H}}^p] < \infty$ ,
- (ii)  $\sum_{j \in \mathbb{Z}} \|\alpha_j\|_{\mathcal{H}} < \infty$ ,
- (iii)  $\min_{1 \leq j \leq 13} \lambda_j > 0$  for  $\mathbf{A}$  defined in (2.1).

Note that Assumption 2.1(ii) implies that  $\mathbf{A}_{n,k}$ ,  $\mathbf{A}$  and  $\mathbf{\Lambda}$  all exist and are of trace class. Our main result of this section is given below.

**Theorem 2.2.** *Grant Assumption 2.1 and let  $\mu \in \mathbb{H}$  with  $\|\mu\|_{\mathbb{H}} < \infty$ . Then*

$$\Delta_n(\mu) \lesssim n^{-\frac{p}{2}+1} (1 + \|\mu\|_{\mathbb{H}}^p) \mathbb{E}[\|\varepsilon_0\|_{\mathbb{H}}^p] + n^{-1} \sum_{j \in \mathbb{Z}} (|j| \wedge n) \|\alpha_j\|_{\mathcal{H}} \mathbb{E}[\|\varepsilon_0\|_{\mathbb{H}}^p].$$

The constant in  $\lesssim$  only depends on  $\sum_{j \in \mathbb{Z}} \|\alpha_j\|_{\mathcal{H}}$  and  $\min_{1 \leq j \leq 13} \lambda_j$ .

**Remark 2.3.** The primary objective of Theorem 2.2 is to provide tight bounds in terms of the rate, the sequence  $\{\alpha_j\}_{j \in \mathbb{N}}$  and the underlying moments  $p \in (2, 3]$ . Note however if  $\mu = 0$  and

$p > 3$  the results in [2] suggest that the rate can be improved. Observe also if Assumption 2.1(iii) is violated (or in fact if  $\lambda_j = 0$  for some finite  $j$ ), we are facing a multivariate problem, which has been the subject of intensive study (cf. [20,39]). Again results dealing with independent random variables suggest that Assumption 2.1(iii) may be weakened, see, for example, [45] and [35] for a general overview.

Unlike to the i.i.d. case, the rate of convergence is also governed by the additional component

$$\mathfrak{A}_n = n^{-1} \sum_{j \in \mathbb{Z}} (|j| \wedge n) \|\alpha_j\|_{\mathfrak{H}}.$$

Before discussing the bound  $\mathfrak{A}_n$  in more detail, we state optimality of the above result.

**Theorem 2.4.** *Grant Assumption 2.1 and let  $\mu \in \mathbb{H}$  with  $\|\mu\|_{\mathbb{H}} < \infty$ . Then the upper bound in Theorem 2.2 is sharp up to a constant, that is, there exist examples meeting Assumption 2.1 where the upper bound is reached up to a constant.*

Expression  $\mathfrak{A}_n$  is particularly interesting if it dominates the rate, that is,  $n^{-\frac{p}{2}+1} = o(\mathfrak{A}_n)$  and hence  $\tau_n = \mathfrak{A}_n^{-1}$ . In order to develop this a little further, let us consider real valued functions  $f$  where

$$f(x) \geq 0 \text{ is monotone and } \int_0^\infty f(x) dx < \infty. \tag{2.2}$$

Put  $b_j = j^2 f(j)$  for  $j \in \mathbb{N}$  and let  $\alpha_j = \frac{b_j - b_{j-1}}{j}$  for  $j \in \mathbb{N} \setminus \{0\}$  and  $\alpha_j = 0$  otherwise. Then

$$\sum_{j \in \mathbb{Z}} (|j| \wedge n) |\alpha_j| \geq \sum_{j=1}^n j \alpha_j = b_n = n^2 f(n).$$

On the other hand, using the monotonicity of  $f(n)$ , we have  $f(n)n \rightarrow 0$  as  $n \rightarrow \infty$ , and hence summation by parts yields

$$\sum_{j=1}^n \alpha_j \leq \frac{b_n}{n} + \sum_{j \in \mathbb{N}} \frac{b_j}{j^2} \leq o(1) + \sum_{j \in \mathbb{N}} f(j) < \infty, \quad n \in \mathbb{N}.$$

We thus obtain the following corollary.

**Corollary 2.5.** *Let  $f(x)$  be a function satisfying (2.2). Then there exist examples satisfying Assumption 2.1 where  $\tau_n \sim (nf(n))^{-1}$ .*

Corollary 2.5 gives a very simple method to provide upper bounds for the rate  $\tau_n$ . For example, setting

$$f(x) = (1+x)^{-\alpha} \quad \text{for } \alpha > 1 \text{ and } x \geq 0 \text{ and } f(x) = 1 \text{ if } x < 0$$

gives the upper bound  $n^{-\alpha+1}$ . Logarithmic rates are obtained by  $f(x) = (1+x)^{-1}(\log(1+x))^{-\alpha}$ ,  $\alpha > 1$ , and this can be continued in the obvious way. Let us mention here that for the real valued case it is shown already in [15] that the rate of convergence in the CLT can be arbitrarily slow, where a much more general framework is considered. Let us now address the question when the rate  $\tau_n = n^{\frac{p}{2}-1}$  persists. To this end, put  $\beta = \frac{p}{2} - 1$ . Since we have the bound

$$\sum_{j \in \mathbb{Z}} (|j| \wedge n) \|\alpha_j\|_{\mathcal{H}} \leq n^{1-\beta} \sum_{j \in \mathbb{Z}} |j|^\beta \|\alpha_j\|_{\mathcal{H}},$$

we obtain the following corollary.

**Corollary 2.6.** *Grant Assumption 2.1. If we have in addition*

$$\sum_{j \in \mathbb{Z}} |j|^{\frac{p}{2}-1} \|\alpha_j\|_{\mathcal{H}} < \infty, \quad p \in (2, 3],$$

then

$$\Delta_n(\mu) \lesssim n^{-\frac{p}{2}+1} (1 + \|\mu\|_{\mathbb{H}}^p) \mathbb{E}[\|\varepsilon_0\|_{\mathbb{H}}^p].$$

### 3. Main results: Non-linear processes

As mentioned earlier, it appears that the only result which obtains optimal rates up to logarithmic factors is [41], where  $\{X_k\}_{k \in \mathbb{Z}}$  is required to be geometrically  $\varphi(n)$ -mixing,  $X_i, X_j$  are uncorrelated for  $i \neq j$  and

$$\mathbb{E}[|\langle X_k, h \rangle|^3] \lesssim \|h\|_{\mathbb{H}} \mathbb{E}[|\langle X_k, h \rangle|^2], \quad h \in \mathbb{H}, k \in \mathbb{Z}.$$

In this section, we follow a different path and focus on Bernoulli-shift processes. We first consider the special case of one-dependent sequences. To this end, let  $\{\varepsilon_k\}_{k \in \mathbb{Z}}$  be a sequence of i.i.d. random variables in some measure space  $\mathbb{S}$ , and  $g : \mathbb{S} \rightarrow \mathbb{H}$  be a measurable map such that

$$X_k = g(\varepsilon_k, \varepsilon_{k-1}), \quad k \in \mathbb{Z}. \tag{3.1}$$

Regarding the method of proof, this special structure will allow us to redirect the problem to the independent case (subject to a special conditional probability measure), by employing a conditioning argument. Unfortunately, as the proof shows, setting this idea to work leads to some non-trivial technicalities that need to be dealt with. To overcome these obstacles, we need to impose slightly stronger moment assumptions on  $X_k$  than before.

For our main result, Theorem 3.2 below, we do not need to impose any additional conditions on  $\mathbb{S}$ , allowing for a large flexibility. This is demonstrated for instance by the subsequent Corollaries 3.3 and 3.4, where more general processes are considered. Our main assumptions are now the following.

**Assumption 3.1.** *For some  $p \geq 9/2$  it holds that*

- (i)  $\mathbb{E}[\varepsilon_k] = 0$  and  $\mathbb{E}[\|X_k\|_{\mathbb{H}}^p] < \infty$ ,
- (ii)  $\{X_k\}_{k \in \mathbb{Z}}$  satisfies (3.1),
- (iii)  $\min_{1 \leq j \leq 13} \lambda_j > 0$  for  $\Lambda(\cdot) = \sum_{|k| \leq 1} \mathbb{E}[\langle X_k, \cdot \rangle X_0]$ .

We then have the following result.

**Theorem 3.2.** *Grant Assumption 3.1. Then*

$$\Delta_n(\mu) \lesssim n^{-1/2} (1 + \|\mu\|_{\mathbb{H}}^3) \mathbb{E}[\|X_0\|_{\mathbb{H}}^{9/2}].$$

The constant in  $\lesssim$  only depends on  $\min_{1 \leq j \leq 13} \lambda_j$ .

Compared to the linear case, the moment condition  $p \geq 9/2$  appears to be suboptimal. On the other hand, for  $\mu \neq 0$ , the rate  $n^{1/2}$  is optimal also for  $p > 3$ , see [2].

The flexibility in the setup allows us to treat Hilbert space valued  $m$ -dependent *potential functions* (cf. [21] for the real valued analogue). More precisely, for  $m \in \mathbb{N}$ , let

$$X_k = g_m(\varepsilon_k, \varepsilon_{k-1}, \dots, \varepsilon_{k-m+1}), \quad k \in \mathbb{Z}, \tag{3.2}$$

for measurable functions  $g_m : \mathbb{S}^m \rightarrow \mathbb{H}$ . We explicitly allow that  $m = m_n$  with  $m = o(n)$  may depend on the sample size  $n$ . The crucial condition here is the non-degeneracy assumption

$$\liminf_{n \rightarrow \infty} \mathbb{E}[\|S_n(X)\|_{\mathbb{H}}^2] / (nm) > 0. \tag{3.3}$$

The underlying covariance operator is then given as

$$\Lambda_m(\cdot) = m^{-2} \sum_{|l| \leq 1} \mathbb{E}[\langle B_l, \cdot \rangle B_0], \quad B_l = \sum_{k=(l-1)m+1}^{lm} X_k. \tag{3.4}$$

We now modify Assumption 3.1(iii) to

$$\inf_m \min_{1 \leq j \leq 13} \lambda_{j,m} > 0 \quad \text{with } \Lambda_m(\cdot) \text{ as in (3.4),} \tag{3.5}$$

to obtain the following corollary.

**Corollary 3.3.** *Grant Assumption 3.1(i), and assume in addition the validity of (3.2), (3.3) and (3.5) with  $m = o(n)$ . Then*

$$\begin{aligned} & \sup_{x \in \mathbb{R}} |P(\|(nm)^{-1/2} S_n(X) + \mu\|_{\mathbb{H}} \leq x) - P(\|Z_{\Lambda_m} + \mu\|_{\mathbb{H}} \leq x)| \\ & \lesssim (n/m)^{-1/2} (1 + \|\mu\|_{\mathbb{H}}^3) \mathbb{E}[\|X_0\|_{\mathbb{H}}^{9/2}]. \end{aligned}$$

Recall that the rate  $(n/m)^{1/2}$  is optimal even for real-valued cases, see, for instance, [11,40] for analogue univariate and multivariate results (a reparametrization is necessary to obtain this explicit form of the rate), and [38] for a lower bound.

A different dependence setup is if  $\{X_k\}_{k \in \mathbb{Z}}$  exhibits weak dependence, the latter only coinciding with  $m$ -dependency in general if  $m$  is finite and independent of  $n$ . A huge variety of weak dependence concepts have been discussed in the literature, see for example [13] and [43]. In our context, the notion of Bernoulli-shift processes together with coupling coefficients is particularly useful (cf. [43]). For Hilbert space valued processes, a related concept is  $\mathbb{L}^p - m$  approximability, see [24]. To formalise the setup, consider

$$X_k = g(\varepsilon_k, \varepsilon_{k-1}, \dots), \quad k \in \mathbb{Z}, \tag{3.6}$$

for measurable functions  $g : \mathbb{S}^\infty \rightarrow \mathbb{H}$ . Let  $\{\varepsilon'_k\}_{k \in \mathbb{Z}}$  be an independent copy of  $\{\varepsilon_k\}_{k \in \mathbb{Z}}$ . We then define the ‘‘coupled’’ random variable  $X'_k$  as

$$X'_k = g(\varepsilon_k, \dots, \varepsilon_1, \varepsilon'_0, \varepsilon_{-1}, \dots), \quad k \in \mathbb{N},$$

see [43] for more details on this kind of coupling. Dependence measures can now be constructed by measuring the distance between  $X_k$  and  $X'_k$ , a popular measure being

$$\theta_p(k) = \mathbb{E}[\|X_k - X'_k\|_{\mathbb{H}}^p]^{1/p}, \quad p \geq 1, \tag{3.7}$$

which we use in the sequel. In the presence of infinite dependence, the underlying covariance operator is now (formally) defined as

$$\mathbf{\Lambda}(\cdot) = \sum_{k \in \mathbb{Z}} \mathbb{E}[(X_k, \cdot)X_0]. \tag{3.8}$$

Existence holds if  $\sum_{k \in \mathbb{N}} \theta_2(k) < \infty$ , see [12]. As before, we modify Assumption 3.1(iii) to

$$\min_{1 \leq j \leq 13} \lambda_j > 0 \quad \text{with } \mathbf{\Lambda}(\cdot) \text{ as in (3.8)}. \tag{3.9}$$

We then have the following result.

**Corollary 3.4.** *Grant Assumption 3.1(i), and assume the validity of (3.6) and (3.9). If in addition  $\theta_{9/2}(k) \lesssim \rho^k$ ,  $0 < \rho < 1$ , then*

$$\Delta_n(\mu) \lesssim (n/\log n)^{-1/2} (1 + \|\mu\|_{\mathbb{H}}^3) \mathbb{E}[\|X_0\|_{\mathbb{H}}^{9/2}].$$

The literature provides a huge variety of important examples of processes displaying a geometric decay in  $\theta_p(k)$ . A prominent example is the following.

**Example 3.5 (ARCH-processes).** Let  $\beta \in \mathbb{L}^2([0, 1]^2)$  be a non-negative Kernel and  $\{\varepsilon_k\}_{k \in \mathbb{Z}} \in \mathbb{L}^2([0, 1])$  be an i.i.d. sequence with  $\mathbb{E}[\varepsilon_k] = 0$ . If the function  $\mu \in \mathbb{L}^2([0, 1])$  is positive, then we call the process

$$X_k = X_k(t) = \varepsilon_k(t)\sigma_{k-1}(t), \quad k \in \mathbb{Z}, t \in [0, 1],$$

with

$$\sigma_k^2(t) = \mu(t) + \int_0^1 \beta(t, s) X_{k-1}^2(s) ds, \quad k \in \mathbb{Z}, t \in [0, 1],$$

the functional ARCH(1)-process. To see why  $X_k$  fits into our framework (satisfies representation (3.6)) is by formally iterating the recursion, yielding (with  $t = t_0$ )

$$X_k(t_0) = \varepsilon_k(t_0) \left( \mu(t_0) + \sum_{i=1}^{\infty} \prod_{j=1}^i \int_0^1 \beta(t_{j-1}, t_j) \varepsilon_{k-j}^2(t_j) \tilde{\mu}_i(j, t_j) dt_j \right)^{1/2},$$

where  $\tilde{\mu}_i(j, \cdot) = 1$  for  $i \neq j$  and  $\tilde{\mu}_i(i, \cdot) = \mu(\cdot)$ . This formal argument can be made rigorous by using Proposition 2.3 in [24] (see also [23], Theorem 2.1), provided that  $\mathbb{E}[K^p(\varepsilon_0^2)] < 1$  with  $p \geq 2$ , where

$$K^2(\varepsilon_0^2) = \int_0^1 \int_0^1 \beta^2(s, t) \varepsilon_0^4(s) ds.$$

Moreover, Proposition 2.3 in [24] also implies  $\theta_p(k) \lesssim \rho^k$  with  $0 < \rho < 1$  for  $p \geq 2$ . Hence if  $p \geq 9/2$  and (3.9) holds, Corollary 3.4 applies.

For additional examples with geometric decay, we refer to [24].

## 4. Proofs

We first deal with the results concerning linear processes, given in Section 2. This is then followed by the proofs of Section 3. We first collect and review some required results from the literature we make repeated use of.

Consider two compact operators  $\mathbf{K}, \mathbf{L}$  with singular value decompositions

$$\mathbf{K}(x) = \sum_{j \in \mathbb{N}} \lambda_j^K \langle x, u_j \rangle f_j, \quad \mathbf{L}(x) = \sum_{j \in \mathbb{N}} \lambda_j^L \langle x, v_j \rangle g_j. \quad (4.1)$$

The following lemma is proven in Section VI.1 of Gohberg *et al.* [17], see their Corollary 1.6 on page 99.

**Lemma 4.1.** *Let  $\mathbf{K}$  and  $\mathbf{L}$  be compact operators with singular value decompositions as in (4.1). Then*

$$|\lambda_j^L - \lambda_j^K| \leq \|\mathbf{K} - \mathbf{L}\|_{\mathcal{H}}.$$

The next lemma appears in some variants in the literature, see, for instance, [3,42] and [44].

**Lemma 4.2.** *Let  $Z \in \mathbb{H}$  be a zero mean Gaussian random variable and  $\mathbf{C}^Z$  its covariance operator. Let  $Y \in \mathbb{H}$  be another, independent random variable. Then*

$$|\mathbb{E}[\exp(it\|Z + Y\|_{\mathbb{H}}^2)]| \leq \prod_{k=1}^{\infty} (1 + 4t^2(\lambda_k^Z)^2)^{-1/4},$$

where  $\lambda_k^Z$  denotes the eigenvalues of  $\mathbf{C}^Z$ .

For the next lemma, we assume that  $\{Y_j\}_{j \in \mathbb{N}} \in \mathbb{H}$  is an independent sequence. For  $2 \leq p \leq 3$ , we introduce the quantities

$$\begin{aligned} \mathfrak{M}_{p,j} &= \mathbb{E}[\|Y_j \mathbf{1}_{\|Y_j\|_{\mathbb{H}} > 1}\|_{\mathbb{H}}^p], & \mathfrak{M}_p &= \sum_{j \in \mathbb{N}} \mathfrak{M}_{p,j}, \\ \mathfrak{L}_{p,j} &= \mathbb{E}[\|Y_j \mathbf{1}_{\|Y_j\|_{\mathbb{H}} \leq 1}\|_{\mathbb{H}}^p], & \mathfrak{L}_p &= \sum_{j \in \mathbb{N}} \mathfrak{L}_{p,j}. \end{aligned}$$

Let  $\Lambda_Y$  be the covariance operator of  $S(Y) = \sum_{j \in \mathbb{N}} Y_j$ . The following result is an adapted version of Theorem 1 in [42].

**Lemma 4.3.** *Assume that the first thirteen eigenvalues of  $\Lambda_Y$  are strictly positive. Then for any  $2 < p \leq 3$  and  $a \in \mathbb{H}$ , we have*

$$\sup_{x \in \mathbb{R}} |P(\|S(Y) - a\|_{\mathbb{H}} \leq x) - P(\|Z - a\|_{\mathbb{H}} \leq x)| \lesssim (1 + \|a\|_{\mathbb{H}}^p)(\mathfrak{M}_2 + \mathfrak{L}_p),$$

where  $Z \in \mathbb{H}$  is a Gaussian random variable with covariance operator  $\Lambda_Y$ .

## 4.1. Proofs of Section 2

We first state and prove the following auxiliary result.

**Theorem 4.4.** *Grant Assumption 2.1 and let  $\mu \in \mathbb{H}$  with  $\|\mu\|_{\mathbb{H}} < \infty$ . Then*

$$\begin{aligned} & \sup_{x \in \mathbb{R}} |P(\|n^{-1/2}S_n(X) + \mu\|_{\mathbb{H}} \leq x) - P(\|Z_{\Lambda_n} + \mu\|_{\mathbb{H}} \leq x)| \\ & \lesssim n^{-\frac{p}{2}+1} (1 + \|\mu\|_{\mathbb{H}}^p) \mathbb{E}[\|\varepsilon_0\|_{\mathbb{H}}^p], \end{aligned}$$

for an appropriate covariance operator  $\Lambda_n(\cdot)$ . The constant in  $\lesssim$  only depends on  $\sum_{j \in \mathbb{Z}} \|\alpha_j\|_{\mathfrak{H}}$  and  $\min_{1 \leq j \leq 13} \lambda_j$ .

Theorem 4.4 gives the optimal rates under sharp dependence condition Assumption 2.1(ii). Note that here the underlying covariance operator  $\Lambda_n$  depends on  $n$  (see the proof for the pre-

cise construction of  $\mathbf{\Lambda}_n$ ). Based on this result, we then obtain Theorem 2.2 by the comparison Lemma 4.7 for Hilbert space valued Gaussian random variables.

**Proof of Theorem 4.4.** For  $1 \leq k \leq n$ , put  $\sqrt{n}U_k = \mathbf{A}_{n,k}(\varepsilon_k)$  and  $\sqrt{n}U_k = \mathbf{A}_{n,k}(\varepsilon_k) + \mathbf{A}_{n,-k+n}(\varepsilon_{-k+n})$  for  $n+1 \leq k \leq \infty$ . We then use the abbreviations  $T_k = U_k$  for  $1 \leq k \leq n$  and  $T_{n+1} = \sum_{k=n+1}^{\infty} U_k$ . Moreover, we put  $S_1^n(U) = \sum_{k=1}^n U_k$  and  $S_1^n(T) = \sum_{k=1}^n T_k$ . Then

$$n^{-1/2} \sum_{k=1}^n X_k = S_1^{n+1}(T). \quad (4.2)$$

In addition, we denote with  $S_1^n(Z)$  the Gaussian counter parts, that is, every  $\varepsilon_i$  is replaced with  $\xi_i$  at the corresponding places. For  $x \in \mathbb{H}$  denote with

$$\begin{aligned} \mathbf{\Lambda}_n^{(0)}(x) &= n^{-1} \sum_{k=1}^n \mathbb{E}[\langle X_k, x \rangle X_k], \\ \mathbf{\Lambda}_{n,k}(x) &= \mathbf{A}_{n,k} \mathbf{C}^\varepsilon \mathbf{A}_{n,k}^*(x), \\ \mathbf{\Lambda}_n^{(1)}(x) &= n^{-1} \sum_{k=1}^n \mathbf{A}_{n,k} \mathbf{C}^\varepsilon \mathbf{A}_{n,k}^*, \end{aligned}$$

and with  $\{\lambda_{n,j}\}_{j \in \mathbb{N}}$  the eigenvalues of the covariance operator  $\mathbf{\Lambda}_n^{(0)}$ . The proof of Theorem 4.4 requires the following lemmas.

**Lemma 4.5.** *Grant Assumption 2.1. Then for any  $2 \leq p \leq 3$  we have*

$$\begin{aligned} \mathbb{E}[\|T_{n+1}\|_{\mathbb{H}}^p] &= \mathcal{O}(n^{-\frac{p}{2}+1} \mathbb{E}[\|\varepsilon_0\|_{\mathbb{H}}^p]) \quad \text{and} \\ \|\mathbf{\Lambda}_n^{(0)} - \mathbf{\Lambda}_n^{(1)}\|_{\mathcal{H}} &= \mathcal{O}(1). \end{aligned}$$

**Proof.** Due to the triangle inequality, it follows that for  $n+1 \leq i \leq \infty$  and  $p \geq 1$  we have

$$\begin{aligned} \mathbb{E}[\|\sqrt{n}U_i\|_{\mathbb{H}}^p] &\leq \left( \sum_{j=1-i}^{n-i} \|\alpha_j\|_{\mathcal{H}} + \sum_{j=1+i}^{n+i} \|\alpha_j\|_{\mathcal{H}} \right)^p \mathbb{E}[\|\varepsilon_0\|_{\mathbb{H}}^p] \\ &\lesssim \left( \sum_{j=1+i}^{n+i} \|\alpha_{-j}\|_{\mathcal{H}} + \sum_{j=1+i}^{n+i} \|\alpha_j\|_{\mathcal{H}} \right)^p \mathbb{E}[\|\varepsilon_0\|_{\mathbb{H}}^p]. \end{aligned} \quad (4.3)$$

Denote with  $a_{n,i}^+ = \sum_{j=1+i}^{n+i} \|\alpha_j\|_{\mathcal{H}}$  and  $a_{n,i}^- = \sum_{j=1+i}^{n+i} \|\alpha_{-j}\|_{\mathcal{H}}$ . Note that since  $\sum_{j \in \mathbb{Z}} \|\alpha_j\|_{\mathcal{H}} < \infty$  we have

$$a_{n,i}^+, a_{n,i}^- \rightarrow 0 \quad \text{as } i \rightarrow \infty, \text{ uniformly in } n. \quad (4.4)$$

Then by (4.3), it follows that for  $K \in \mathbb{N}$

$$\begin{aligned} \sum_{i=n+1}^{\infty} \mathbb{E}[\|\sqrt{n}U_i\|_{\mathbb{H}}^p] / \mathbb{E}[\|\varepsilon_0\|_{\mathbb{H}}^p] &\lesssim \sum_{i=1}^K (a_{n,i}^+ + a_{n,i}^-)^p + \sum_{i>K} (a_{n,i}^+ + a_{n,i}^-)^p \\ &\lesssim 2^p K \left( \sum_{i \in \mathbb{Z}} \|\alpha_i\|_{\mathcal{H}} \right)^p + \sum_{|i|>K} (n \wedge i) \|\alpha_i\|_{\mathcal{H}}. \end{aligned}$$

Selecting  $K = K_n \rightarrow \infty$  such that  $K_n = o(n)$ , we deduce that

$$\sum_{i=n+1}^{\infty} \mathbb{E}[\|\sqrt{n}U_i\|_{\mathbb{H}}^p] = o(n \mathbb{E}[\|\varepsilon_0\|_{\mathbb{H}}^p]). \tag{4.5}$$

Using a Rosenthal inequality for Hilbert spaces (cf. [33]), we get for  $p \geq 2$

$$\begin{aligned} \mathbb{E}[\|T_{n+1}\|_{\mathbb{H}}^p] &\lesssim n^{-\frac{p}{2}} \left\{ \left( \sum_{i=n+1}^{\infty} \mathbb{E}[\|\sqrt{n}U_i\|_{\mathbb{H}}^2] \right)^{\frac{p}{2}} + \sum_{i=n+1}^{\infty} \mathbb{E}[\|\sqrt{n}U_i\|_{\mathbb{H}}^p] \right\} \\ &= o(n^{-\frac{p}{2}+1} \mathbb{E}[\|\varepsilon_0\|_{\mathbb{H}}^p]). \end{aligned} \tag{4.6}$$

This gives the first claim. Next, denote with  $\lambda_j^{(T)}$  and  $e_j^{(T)}$  the eigenvalues and functions of the Covariance operator of  $T_{n+1}$ , which exists due to (4.6). Since  $T_{n+1}$  is independent of  $\{T_j\}_{1 \leq j \leq n}$  by construction, we get that for any  $x \in \mathbb{H}$

$$\begin{aligned} \Lambda_n^{(0)}(x) - \Lambda_n^{(1)}(x) &= \mathbb{E}[\langle S_1^n(T), x \rangle T_{n+1} + \langle T_{n+1}, x \rangle S_1^n(T) + \langle T_{n+1}, x \rangle T_{n+1}] \\ &= \mathbb{E}[\langle T_{n+1}, x \rangle T_{n+1}]. \end{aligned}$$

It then follows from Cauchy–Schwarz and Parseval’s identity that

$$\|\Lambda_n^{(0)} - \Lambda_n^{(1)}\|_{\mathcal{H}} \leq \|\mathbb{E}[\langle T_{n+1}, \cdot \rangle T_{n+1}]\|_{\mathcal{H}} \leq \left( \sum_{j=1}^{\infty} (\lambda_j^{(T)})^2 \right)^{1/2}.$$

By the triangle inequality and from (4.6), we deduce that

$$\left( \sum_{j=1}^{\infty} (\lambda_j^{(T)})^2 \right)^{1/2} \leq \sum_{j=1}^{\infty} \lambda_j^{(T)} = \mathbb{E}[\|T_{n+1}\|_{\mathbb{H}}^2] = o(1),$$

and hence

$$\|\Lambda_n^{(0)} - \Lambda_n^{(1)}\|_{\mathcal{H}} = o(1). \tag{4.7}$$

□

**Lemma 4.6.** *Assume that Assumption 2.1 holds. Then there exists an  $n_0 \in \mathbb{N}$  such that for  $n \geq n_0$  we have  $\lambda_{n,k} > 0$  for any fixed  $k \in \mathbb{N}$  where  $\lambda_k > 0$ .*

**Proof.** Note first that since  $\|\mathbf{A}\|_{\mathcal{H}}, \|\mathbf{C}^\varepsilon\|_{\mathcal{H}}, \|\mathbf{A}_{n,k}\|_{\mathcal{H}} < \infty$  for all  $i \in \mathbb{Z}$  and  $n \in \mathbb{N}$ ,  $\mathbf{A}_{n,k}$  and  $\mathbf{A}$  are bounded operators. By Lemma 4.1, we have

$$|\lambda_k - \lambda_{n,k}| \leq \|\mathbf{A} - \mathbf{A}_n^{(1)}\|_{\mathcal{H}},$$

hence it suffices to consider  $\|\mathbf{A} - \mathbf{A}_n^{(1)}\|_{\mathcal{H}}$ . Moreover, by the triangle inequality and Lemma 4.5, we only need to consider  $\|\mathbf{A} - \mathbf{A}_n^{(1)}\|_{\mathcal{H}}$ . Using the linearity of  $\mathbf{A}_{n,k}, \mathbf{A}, \mathbf{C}^\varepsilon$  and the fact that  $\|\mathbf{B}\|_{\mathcal{H}} = \|\mathbf{B}^*\|_{\mathcal{H}}$  for an operator  $\mathbf{B}$ , it follows that

$$\begin{aligned} \|\mathbf{A}_{n,k} - \mathbf{A}\|_{\mathcal{H}} &\leq \|\mathbf{A}_{n,k} \mathbf{C}^\varepsilon \mathbf{A}_{n,k}^* - \mathbf{A}_{n,k} \mathbf{C}^\varepsilon \mathbf{A}^*\|_{\mathcal{H}} + \|\mathbf{A} \mathbf{C}^\varepsilon \mathbf{A}^* - \mathbf{A}_{n,k} \mathbf{C}^\varepsilon \mathbf{A}^*\|_{\mathcal{H}} \\ &\leq 2\|\mathbf{A} - \mathbf{A}_{n,k}\|_{\mathcal{H}} \|\mathbf{C}^\varepsilon\|_{\mathcal{H}} (\|\mathbf{A}\|_{\mathcal{H}} + \|\mathbf{A}_{n,k}\|_{\mathcal{H}}). \end{aligned}$$

By the triangle inequality, we have  $\|\mathbf{A} - \mathbf{A}_{n,k}\|_{\mathcal{H}} \leq \sum_{j>n-k} \|\alpha_j\|_{\mathcal{H}} + \sum_{j<1-k} \|\alpha_j\|_{\mathcal{H}} = o(1)$  as  $k, n \rightarrow \infty$ . Since

$$\left\| n^{-1} \sum_{k=1}^n \mathbf{A}_{n,k} - \mathbf{A} \right\|_{\mathcal{H}} \leq n^{-1} \sum_{k=1}^{n-1} \|\mathbf{A}_{n,k} - \mathbf{A}\|_{\mathcal{H}} + 2l/n \|\mathbf{A}\|_{\mathcal{H}}, \quad (4.8)$$

the claim follows.  $\square$

We are now ready to proceed to the proof of Theorem 4.4. Due to Lemma 4.6, we have

$$\liminf_{n \rightarrow \infty} \min_{1 \leq k \leq 13} \lambda_{n,k} > 0. \quad (4.9)$$

Hence applying Lemma 4.3, it follows that

$$\begin{aligned} \Delta_n(\mu) &\leq \sup_{x \in \mathbb{R}} |P(\|S_1^{n+1}(T) + \mu\|_{\mathbb{H}} \leq x) - P(\|S_1^{n+1}(Z) + \mu\|_{\mathbb{H}} \leq x)| \\ &\lesssim (1 + \|\mu\|_{\mathbb{H}}^p) \left( \sum_{k=1}^{n+1} \|T_k\|_{\mathbb{H}} \mathbf{1}(\|T_k\|_{\mathbb{H}} \geq 1) \right)_2^2 + \sum_{k=1}^{n+1} \|T_k\|_{\mathbb{H}}^p \Big) \\ &\stackrel{\text{def}}{=} (1 + \|\mu\|_{\mathbb{H}}^p) (\mathbf{I}_n + \mathbf{II}_n). \end{aligned} \quad (4.10)$$

We first treat  $\mathbf{I}_n$ . For  $1 \leq k \leq n$ , we obtain

$$\begin{aligned} \|\|T_k\|_{\mathbb{H}} \mathbf{1}(\|T_k\|_{\mathbb{H}} \geq 1)\|_2^2 &= \|n^{-1/2} \|\mathbf{A}_{n,k}(\varepsilon_k)\|_{\mathbb{H}} \mathbf{1}(\|\mathbf{A}_{n,k}(\varepsilon_k)\|_{\mathbb{H}} \geq n^{1/2})\|_2^2 \\ &\lesssim n^{-1 - \frac{p-2}{2}} \|\|\mathbf{A}_{n,k}(\varepsilon_k)\|_{\mathbb{H}}\|_p^p \\ &\lesssim n^{-\frac{p}{2}} \mathbb{E}[\|\varepsilon_0\|_{\mathbb{H}}^p]. \end{aligned} \quad (4.11)$$

Similarly, using Lemma 4.5, we deduce that

$$\left\| \|T_{n+1}\|_{\mathbb{H}} \mathbf{1}(\|T_{n+1}\|_{\mathbb{H}} \geq 1) \right\|_2^2 \lesssim n^{-\frac{p}{2}+1} \mathbb{E}[\|\varepsilon_0\|_{\mathbb{H}}^p]. \quad (4.12)$$

Combining (4.11) and (4.12), we obtain

$$\mathbf{I}_n \lesssim n^{-\frac{p}{2}+1} \mathbb{E}[\|\varepsilon_0\|_{\mathbb{H}}^p]. \quad (4.13)$$

Next, we deal with  $\mathbf{II}_n$ . First note that for  $1 \leq k \leq n$ , we obtain via the triangle inequality

$$\| \|T_k\|_{\mathbb{H}} \| \| \lesssim n^{-\frac{1}{2}} \sum_{j \in \mathbb{Z}} \|\alpha_j\|_{\mathfrak{H}} \| \|\varepsilon_0\|_{\mathbb{H}} \| \|_p \lesssim n^{-1/2} \mathbb{E}[\|\varepsilon_0\|_{\mathbb{H}}^p].$$

Using Lemma 4.5, we thus deduce that

$$\mathbf{II}_n \lesssim \sum_{j=1}^{n+1} n^{-\frac{p}{2}} \mathbb{E}[\|\varepsilon_0\|_{\mathbb{H}}^p] \lesssim n^{-\frac{p}{2}+1} \mathbb{E}[\|\varepsilon_0\|_{\mathbb{H}}^p]. \quad (4.14)$$

Combining (4.13) with (4.14) completes the proof.  $\square$

**Proof of Theorem 2.2.** The proof is based on two main lemmas. The first one describes a comparison result for two Gaussian, Hilbert-space valued random variables in terms of perturbed covariance operators.

**Lemma 4.7.** *Let  $Y_1, Y_2 \in \mathbb{H}$  be two Gaussian random variables with covariance operators  $\mathbf{C}^{Y_1}$  and  $\mathbf{C}^{Y_2}$  of trace class and finite mean  $\mu$ . Suppose that for  $\delta > 0$*

$$\lambda_{13}^{Y_1} > \delta \quad \text{and} \quad \|\mathbf{C}^{Y_1} - \mathbf{C}^{Y_2}\|_{\mathfrak{H}} \leq \delta, \quad (4.15)$$

where  $\{\lambda_j^{Y_1}\}_{j \in \mathbb{N}}$  denotes the eigenvalues of  $\mathbf{C}^{Y_1}$ . Then

$$\sup_{x \in \mathbb{R}} |P(\|Y_1\|_{\mathbb{H}} \leq x) - P(\|Y_2\|_{\mathbb{H}} \leq x)| \lesssim |\text{tr}(\mathbf{C}^{Y_1}) - \text{tr}(\mathbf{C}^{Y_2})|,$$

where  $\text{tr}(\mathbf{B})$  denotes the trace of an operator  $\mathbf{B}$ .

**Proof.** We may argue similarly as in [44]. Due to the Gaussianity of  $Y_1$  and  $Y_2$

$$Y_1 - \mu \stackrel{d}{=} \sum_{k=1}^n \xi_{n,k}, \quad Y_2 - \mu \stackrel{d}{=} \sum_{k=1}^n \eta_{n,k}, \quad (4.16)$$

where  $\{\xi_{n,k}\}_{1 \leq k \leq n}$  and  $\{\eta_{n,k}\}_{1 \leq k \leq n}$  are i.i.d. Gaussian sequences with covariance operators  $n^{-1}\mathbf{C}^{Y_1}$  and  $n^{-1}\mathbf{C}^{Y_2}$ . For  $n \in \mathbb{N}$ , denote with

$$W_{n,k} = \mu + \sum_{j=1}^{k-1} \xi_{n,j} + \sum_{j=k+1}^n \eta_{n,j}, \quad 1 \leq k \leq n. \quad (4.17)$$

Following the proof in [44], a careful inspection reveals (cf. Equation 3.5 in [44]) that it suffices to reconsider the quantity

$$Q_{n,k}(t) = |t| |\mathbb{E}[\exp(it \|W_{n,k}\|_{\mathbb{H}}^2)] (\mathbb{E}[\|\xi_{n,k}\|_{\mathbb{H}}^2] - \mathbb{E}[\|\eta_{n,k}\|_{\mathbb{H}}^2])|, \quad t \in \mathbb{R}, \quad (4.18)$$

and establish that

$$\int_{\mathbb{R}} \frac{Q_{n,k}(t)}{t} dt \lesssim \frac{1}{n} |\operatorname{tr}(\mathbf{C}^{Y_1}) - \operatorname{tr}(\mathbf{C}^{Y_2})|. \quad (4.19)$$

Once we have (4.19), the results in [44] imply that

$$\begin{aligned} \sup_{x \in \mathbb{R}} |P(\|Y_1\|_{\mathbb{H}} \leq x) - P(\|Y_2\|_{\mathbb{H}} \leq x)| &\lesssim \frac{1}{n} \sum_{k=1}^n |\operatorname{tr}(\mathbf{C}^{Y_1}) - \operatorname{tr}(\mathbf{C}^{Y_2})| + \frac{1 + \|\mu\|_{\mathbb{H}}^3}{\sqrt{n}} \\ &\lesssim |\operatorname{tr}(\mathbf{C}^{Y_1}) - \operatorname{tr}(\mathbf{C}^{Y_2})| + \frac{1 + \|\mu\|_{\mathbb{H}}^3}{\sqrt{n}}. \end{aligned}$$

Selecting  $n$  sufficiently large, the claim follows. We proceed by showing (4.19). To this end, note that by Lemma 4.2 we have that

$$|\mathbb{E}[\exp(it \|W_{n,k}\|_{\mathbb{H}}^2)]| \leq \prod_{j=1}^{\infty} (1 + 4t^2 \tilde{\lambda}_{n,k,j}^2)^{-1/4}, \quad (4.20)$$

where  $\tilde{\lambda}_{n,k,j}$  denote the eigenvalues of the covariance operator

$$\tilde{\mathbf{C}}_{n,k}(x) = \mathbb{E}[\langle W_{n,k} - \mu, x \rangle \langle W_{n,k} - \mu \rangle]. \quad (4.21)$$

Exploiting the mutual independence of  $\xi_{n,k}$  and  $\eta_{n,k}$ , it follows that

$$\begin{aligned} \|\tilde{\mathbf{C}}_{n,k} - \mathbf{C}^{Y_1}\|_{\mathcal{H}} &\leq \frac{1}{n} \|\mathbf{C}^{Y_1}\|_{\mathcal{H}} + \frac{n-k}{n} \|\mathbf{C}^{Y_2} - \mathbf{C}^{Y_1}\|_{\mathcal{H}} \\ &\leq \delta + \mathcal{O}(n^{-1}). \end{aligned}$$

Hence an application of Lemma 4.1 yields that for  $1 \leq k \leq 13$

$$\begin{aligned} \tilde{\lambda}_{n,k,j} &\geq \lambda_k^{Y_1} - |\lambda_k^{Y_1} - \tilde{\lambda}_{n,k,j}| \\ &\geq \lambda_k^{Y_1} - \delta - \mathcal{O}(n^{-1}) > 0 \end{aligned} \quad (4.22)$$

for sufficiently large  $n$ . Thus  $\tilde{\lambda}_{n,k,j} > 0$  uniformly, and we conclude from (4.20) that

$$\int_{\mathbb{R}} |\mathbb{E}[\exp(it \|W_k\|_{\mathbb{H}}^2)]| dt \leq \int_{\mathbb{R}} \prod_{j=1}^{13} (1 + 4t^2 \tilde{\lambda}_{n,k,j}^2)^{-1/4} dt < \infty. \quad (4.23)$$

Note that here we actually only require  $\lambda_3^{Y_1} > 0$ . Since we have

$$\mathbb{E}[\|X\|_{\mathbb{H}}^2] = \text{tr}(\mathbf{C}^X)$$

for any  $X \in \mathbb{H}$  with covariance operator  $\mathbf{C}^X$  of trace class, the claim follows selecting  $n$  large enough.  $\square$

Next, recall

$$\mathbf{A}_{n,k}^c = \sum_{j>n-k} \alpha_j + \sum_{j<1-k} \alpha_j, \quad 1 \leq k \leq n,$$

and that we have the decomposition

$$\begin{aligned} \sum_{k=1}^n Z_k &= \sum_{k=1}^n \mathbf{A}(\xi_k) - \sum_{k=1}^n \mathbf{A}_{n,k}^c(\xi_k) + \sum_{k>n} \mathbf{A}_{n,k}(\xi_k) + \sum_{k<1} \mathbf{A}_{n,k}(\xi_k) \\ &\stackrel{\text{def}}{=} \sum_{k=1}^n \mathbf{A}(\xi_k) + \mathbf{III}_n + \mathbf{IV}_n + \mathbf{V}_n. \end{aligned} \quad (4.24)$$

We now have our second lemma.

**Lemma 4.8.** *Assume that Assumption 2.1 holds. Then*

$$\begin{aligned} &\left\| \mathbf{A}\mathbf{C}^\varepsilon \mathbf{A}^* - \frac{1}{n} \mathbb{E}[\langle S_1^{n+1}(Z), x \rangle S_1^{n+1}(Z)] \right\|_{\mathcal{H}} \\ &\lesssim \frac{1}{n} \sum_{j \in \mathbb{Z}} (|j| \wedge n) \|\alpha_j\|_{\mathcal{H}}. \end{aligned}$$

**Proof.** Observe that  $\sum_{k=1}^n \mathbf{A}(\xi_k) + \mathbf{III}_n$ ,  $\mathbf{IV}_n$  and  $\mathbf{V}_n$  are all mutually independent. It follows that for any  $x \in \mathbb{H}$  we have that

$$\mathbb{E}[\langle \mathbf{IV}_n, x \rangle \mathbf{V}_n] = \mathbb{E}[\langle \mathbf{V}_n, x \rangle \mathbf{IV}_n] = 0 \quad (4.25)$$

(with  $0 \in \mathbb{H}$ ), and this remains valid if we substitute  $\mathbf{IV}_n$  or  $\mathbf{V}_n$  with  $\sum_{k=1}^n \mathbf{A}(\xi_k) + \mathbf{III}_n$ . Similarly, if  $i \neq j$  one readily derives that for  $x \in \mathbb{H}$

$$\begin{aligned} \mathbb{E}[\langle \mathbf{A}(\xi_i), x \rangle \mathbf{A}_j(\xi_j)] &= \mathbb{E}[\langle \mathbf{A}(\xi_i), x \rangle \mathbf{A}_{n,j}^c(\xi_j)] \\ &= \mathbb{E}[\langle \mathbf{A}_{n,i}^c(\xi_i), x \rangle \mathbf{A}(\xi_j)] = 0 \end{aligned} \quad (4.26)$$

(with  $0 \in \mathbb{H}$ ), exploiting the independence of  $\xi_i$  and  $\xi_j$  and the linearity of the operators  $\mathbf{A}$  and  $\mathbf{A}_{n,j}^c$ . Denote with  $e_j$  the eigenfunctions of  $\mathbf{C}^\varepsilon$ , and with  $\lambda_j^\varepsilon$  its corresponding eigenvalues. If

$i = j$ , it then follows that

$$\begin{aligned} \|\mathbb{E}[\langle \mathbf{A}(\varepsilon_i), x \rangle \mathbf{A}_{n,i}^c(\varepsilon_i)]\|_{\mathcal{H}} &= \left\| \sum_{l=1}^{\infty} \lambda_l^\varepsilon \mathbb{E}[\langle \mathbf{A}(e_l), x \rangle \mathbf{A}_{n,i}^c(e_l)] \right\|_{\mathcal{H}} \\ &\leq \sum_{l=1}^{\infty} \lambda_l^\varepsilon \|\mathbf{A}\|_{\mathcal{H}} \|x\|_{\mathbb{H}} \|\mathbf{A}_{n,i}^c\|_{\mathcal{H}} \\ &\lesssim \|x\|_{\mathbb{H}} \|\mathbf{A}_{n,i}^c\|_{\mathcal{H}}, \end{aligned} \quad (4.27)$$

where we used Cauchy–Schwarz and  $\sum_{j=1}^{\infty} \|\alpha_j\|_{\mathcal{H}} < \infty$ . As before, the same bound also applies if we exchange  $\mathbf{A}_{n,i}^c$  and  $\mathbf{A}$ . Similarly, we also obtain that

$$\|\mathbb{E}[\langle \mathbf{A}_{n,i}(\varepsilon_i), x \rangle \mathbf{A}_{n,i}(\varepsilon_i)]\|_{\mathcal{H}} \lesssim \|x\|_{\mathbb{H}} \|\mathbf{A}_{n,i}\|_{\mathcal{H}} \quad (4.28)$$

and

$$\|\mathbb{E}[\langle \mathbf{A}_{n,i}^c(\varepsilon_i), x \rangle \mathbf{A}_{n,i}^c(\varepsilon_i)]\|_{\mathcal{H}} \lesssim \|x\|_{\mathbb{H}} \|\mathbf{A}_{n,i}^c\|_{\mathcal{H}}. \quad (4.29)$$

The treatment of  $\mathbf{IV}_n, \mathbf{V}_n$  follows as in Lemma 4.5. For the sake of completeness, we have that

$$\begin{aligned} \|\mathbb{E}[\langle \mathbf{IV}_n, x \rangle \mathbf{IV}_n]\|_{\mathcal{H}} &\leq \sum_{i>n} \sum_{j>n} \|\mathbb{E}[\langle \mathbf{A}_{n,i}(\xi_i), x \rangle \mathbf{A}_{n,j}(\xi_j)]\|_{\mathcal{H}} \\ &= \sum_{i>n} \|\mathbb{E}[\langle \mathbf{A}_{n,i}(\xi_i), x \rangle \mathbf{A}_{n,i}(\xi_i)]\|_{\mathcal{H}} \\ &\lesssim \sum_{i>n} \|\mathbf{A}_{n,i}\|_{\mathcal{H}} \|x\|_{\mathbb{H}} \\ &\lesssim \sum_{j \in \mathbb{Z}} (|j| \wedge n) \|\alpha_j\|_{\mathcal{H}} \|x\|_{\mathbb{H}}. \end{aligned} \quad (4.30)$$

The same bound applies to  $\mathbf{V}_n$ , that is, we have

$$\|\mathbb{E}[\langle \mathbf{V}_n, x \rangle \mathbf{V}_n]\|_{\mathcal{H}} \lesssim \sum_{j \in \mathbb{Z}} (|j| \wedge n) \|\alpha_j\|_{\mathcal{H}} \|x\|_{\mathbb{H}}. \quad (4.31)$$

By independence, we have

$$\begin{aligned} &\mathbb{E}[\langle \mathcal{S}_1^{n+1}(Z), x \rangle \mathcal{S}_1^{n+1}(Z)] \\ &= \mathbb{E} \left[ \left\langle \sum_{k=1}^n \mathbf{A}(\xi_k) + \mathbf{III}_n, x \right\rangle \left( \sum_{k=1}^n \mathbf{A}(\xi_k) + \mathbf{III}_n \right) \right] \\ &\quad + \mathbb{E}[\langle \mathbf{IV}_n, x \rangle \mathbf{IV}_n] + \mathbb{E}[\langle \mathbf{V}_n, x \rangle \mathbf{V}_n]. \end{aligned} \quad (4.32)$$

Using (4.27) and (4.29), it follows that

$$\begin{aligned} & \left\| \mathbb{E} \left[ \left\langle \sum_{k=1}^n \mathbf{A}(\xi_k) + \mathbf{III}_n, x \right\rangle \left( \sum_{k=1}^n \mathbf{A}(\xi_k) + \mathbf{III}_n \right) \right] - \sum_{k=1}^n \mathbb{E}[\langle \mathbf{A}(\xi_k), x \rangle \mathbf{A}(\xi_k)] \right\|_{\mathcal{H}} \\ & \lesssim \sum_{j \in \mathbb{Z}} (|j| \wedge n) \|\alpha_j\|_{\mathcal{H}}. \end{aligned} \quad (4.33)$$

Combining (4.30), (4.31) and (4.33), we finally obtain that

$$\left\| \mathbb{E}[(S_1^{n+1}(Z), x) S_1^{n+1}(Z)] - \sum_{k=1}^n \mathbb{E}[\langle \mathbf{A}(\xi_k), x \rangle \mathbf{A}(\xi_k)] \right\|_{\mathcal{H}} \lesssim \sum_{j \in \mathbb{Z}} (|j| \wedge n) \|\alpha_j\|_{\mathcal{H}}. \quad (4.34)$$

Since we have that

$$\mathbb{E}[\langle \mathbf{A}(\xi_k), x \rangle \mathbf{A}(\xi_k)] = \mathbf{A} \mathbf{C}^\varepsilon \mathbf{A}^*,$$

it follows that

$$\left\| \mathbf{A} \mathbf{C}^\varepsilon \mathbf{A}^* - \frac{1}{n} \mathbb{E}[(S_1^{n+1}(Z), x) S_1^{n+1}(Z)] \right\|_{\mathcal{H}} \lesssim \frac{1}{n} \sum_{j \in \mathbb{Z}} (|j| \wedge n) \|\alpha_j\|_{\mathcal{H}}. \quad \square$$

We are now ready to proceed to the actual proof. For  $\mu \in \mathbb{H}$  denote with

$$Y_1 = \frac{1}{\sqrt{n}} \sum_{k=1}^n \mathbf{A}(\xi_k) + \mu, \quad Y_2 = \frac{1}{\sqrt{n}} \sum_{k=1}^n Z_k + \mu,$$

and the corresponding covariance operators with  $\mathbf{C}^{Y_1}$ ,  $\mathbf{C}^{Y_2}$ . Note that  $\mathbf{C}^{Y_1} = \mathbf{A} \mathbf{C}^\varepsilon \mathbf{A}^*$ . The aim is to invoke Lemma 4.7. To this end, we need to establish the necessary bounds. Since

$$\sum_{j \in \mathbb{Z}} (|j| \wedge n) \|\alpha_j\|_{\mathcal{H}} = \mathcal{O}(n)$$

due to  $\sum_{j \in \mathbb{Z}} \|\alpha_j\|_{\mathcal{H}} < \infty$ , Lemma 4.8 yields that

$$\|\mathbf{C}^{Y_1} - \mathbf{C}^{Y_2}\|_{\mathcal{H}} = \mathcal{O}(1).$$

Hence condition (4.15) is valid by assumption. Next, note that by the independence of  $\sum_{k=1}^n \mathbf{A}(\xi_k) + \mathbf{III}_n$ ,  $\mathbf{IV}_n$  and  $\mathbf{V}_n$ , we have that

$$n \mathbb{E}[\|Y_1\|_{\mathbb{H}}^2] = \sum_{k=1}^n \mathbb{E}[\|\mathbf{A}(\xi_k) - \mathbf{A}_{n,k}^c(\xi_k)\|_{\mathbb{H}}^2] + \mathbb{E}[\|\mathbf{IV}_n\|_{\mathbb{H}}^2] + \mathbb{E}[\|\mathbf{V}_n\|_{\mathbb{H}}^2]. \quad (4.35)$$

Using the triangle inequality and  $a^2 - b^2 = (a - b)(a + b)$ , we get that

$$\begin{aligned} |\mathbb{E}[\|\mathbf{A}(\xi_k) - \mathbf{A}_{n,k}^c(\xi_k)\|_{\mathbb{H}}^2 - \|\mathbf{A}(\xi_k)\|_{\mathbb{H}}^2]| &\leq \mathbb{E}[\|\mathbf{A}_{n,k}^c(\xi_k)\|_{\mathbb{H}}(2\|\mathbf{A}(\xi_k)\|_{\mathbb{H}} + \|\mathbf{A}_{n,k}^c(\xi_k)\|_{\mathbb{H}})] \\ &\lesssim \|\mathbf{A}_{n,k}^c(\xi_k)\|_{\mathcal{H}} \mathbb{E}[\|\xi_0\|_{\mathbb{H}}^2]. \end{aligned}$$

Hence, we obtain the estimate

$$\left| \sum_{k=1}^n \mathbb{E}[\|\mathbf{A}(\xi_k) - \mathbf{A}_{n,k}^c(\xi_k)\|_{\mathbb{H}}^2] \right| \lesssim \sum_{j \in \mathbb{Z}} (|j| \wedge n) \|\alpha_j\|_{\mathcal{H}}. \quad (4.36)$$

Similarly, proceeding as in Lemma 4.5, one readily computes that for  $p \in (2, 3]$

$$\|\|\mathbf{IV}_n\|_{\mathbb{H}}\|_p^p, \|\|\mathbf{V}_n\|_{\mathbb{H}}\|_p^p \lesssim \sum_{j \in \mathbb{Z}} (|j| \wedge n) \|\alpha_j\|_{\mathcal{H}} \|\varepsilon_0\|_{\mathbb{H}}\|_p^p. \quad (4.37)$$

Combining (4.35) with (4.36) and (4.37), the claim then follows from Lemma 4.7.  $\square$

**Proof of Theorem 2.4.** For the proof, we construct an example where the upper bound is obtained, up to a constant. It suffices to consider the special case where  $\mathbb{H} = \mathbb{R}$  and  $X_k$  is “purely” non-causal, that is,  $X_k = \sum_{j=0}^{\infty} \alpha_j \varepsilon_{k+j}$ ,  $\alpha_j \in \mathbb{R}$ . Moreover, we assume throughout this section that  $\mathbb{E}[\varepsilon_k^2] = 1$ ,  $\alpha_j \geq 0$  and  $\mathbf{A} = 1$  to simplify matters. We first require the following lemma.

**Lemma 4.9.** *Assume that  $\alpha_j \geq 0$  such that  $\sum_{j \in \mathbb{N}} \alpha_j < \infty$  and  $\sum_{j \in \mathbb{N}} (j \wedge n) \alpha_j \rightarrow \infty$  as  $n \rightarrow \infty$ . Then*

$$\begin{aligned} \sum_{k>n} (\mathbf{A}_{n,k})^2 + \sum_{k<1} (\mathbf{A}_{n,k})^2 &= \mathcal{O}\left(\sum_{j \in \mathbb{N}} (j \wedge n) \alpha_j\right), \\ \sum_{k=1}^n (\mathbf{A}_{n,k}^c)^2 &= \mathcal{O}\left(\sum_{j \in \mathbb{N}} (j \wedge n) \alpha_j\right). \end{aligned}$$

**Proof.** We only show the first claim, the second follows in an analogue manner. Since  $\sum_{j>L} \alpha_j \rightarrow 0$  as  $L \rightarrow \infty$ , there exists  $\delta_n \rightarrow 0$  and  $m_n \rightarrow \infty$  as  $n \rightarrow \infty$ , such that

$$\begin{aligned} \sum_{k>n} (\mathbf{A}_{n,k})^2 &= \sum_{k>n} \left( \sum_{j=-n+k}^{k-1} \alpha_j \right)^2 = \sum_{l=1}^{\infty} \left( \sum_{j=l}^{n+l-1} \alpha_j \right)^2 \\ &\leq \sum_{l=1}^{m_n} \sum_{j=l}^{n+l-1} \alpha_j + \delta_n \sum_{l>m_n} \sum_{j=l}^{n+l-1} \alpha_j \\ &\leq \sum_{j=1}^{m_n} (j \wedge n) \alpha_j + \delta_n \sum_{j \in \mathbb{N}} (j \wedge n) \alpha_j, \end{aligned} \quad (4.38)$$

where we also used  $\sum_{j \in \mathbb{N}} \alpha_j = 1$ . Since  $\sum_{j \in \mathbb{N}} (j \wedge n) \alpha_j \rightarrow \infty$ , we can choose  $m_n$  such that

$$\sum_{j=1}^{m_n} (j \wedge n) \alpha_j = o\left(\sum_{j \in \mathbb{N}} (j \wedge n) \alpha_j\right),$$

and the claim follows for  $\sum_{k>n} (\mathbf{A}_{n,k})^2$ . Regarding expression  $\sum_{k<1} (\mathbf{A}_{n,k})^2$ , note that  $\mathbf{A}_{n,k} = 0$  for  $k \leq 0$  by assumption, hence the claim.  $\square$

It is known in the literature that the rate  $n^{p/2-1}$  is optimal (cf. [34]). Hence due to Theorem 4.4, it suffices to derive a lower bound for

$$\sup_{x \in \mathbb{R}} |P(Z_{\mathbf{A}_n} \leq x) - P(Z_{\mathbf{A}} \leq x)|.$$

Moreover, we may assume without loss of generality that

$$\sum_{j \in \mathbb{N}} (j \wedge n) \alpha_j \rightarrow \infty \quad \text{as } n \rightarrow \infty, \tag{4.39}$$

since otherwise the claim immediately follows. Denote with  $\sigma^2 = \|Z_{\mathbf{A}}\|_2^2$  and  $\sigma_n^2 = \|Z_{\mathbf{A}_n}\|_2^2$ . Note that  $\sigma^2 = \mathbf{A}^2$ , and by (4.35) we have  $\sigma_n^2 = n^{-1} \sum_{k \in \mathbb{N}} \mathbf{A}_{n,k}^2$ . The proof relies on the following lower bound. For large enough  $n$ , there exists a constant  $C_\alpha > 0$  (which can be chosen arbitrarily smaller than two) such that

$$\sigma^2 - \sigma_n^2 \geq \frac{C_\alpha}{n} \sum_{j \in \mathbb{N}} (j \wedge n) \alpha_j. \tag{4.40}$$

We first derive this lower bound, a simple application of the mean value theorem then yields the claim, see below. By Lemma 4.9, it follows that

$$\begin{aligned} \sigma_n^2 &= \frac{1}{n} \sum_{k=1}^n \mathbf{A}_{n,k}^2 + \frac{1}{n} \sum_{k>n} \mathbf{A}_{n,k}^2 + \frac{1}{n} \sum_{k<1} (\mathbf{A}_{n,k})^2 \\ &= \frac{1}{n} \sum_{k=1}^n \mathbf{A}_{n,k}^2 + o\left(\sum_{j \in \mathbb{N}} (j \wedge n) \alpha_j\right). \end{aligned}$$

On the other hand, we also have

$$\begin{aligned} \mathbf{A}^2 - \sum_{k=1}^n (\mathbf{A}_{n,k})^2 &= \sum_{k=1}^n \mathbf{A}_{n,k}^c (\mathbf{A} + \mathbf{A}_{n,k}) = \sum_{k=1}^n \mathbf{A}_{n,k}^c (2\mathbf{A} - \mathbf{A}_{n,k}^c) \\ &= 2 \sum_{k=1}^n \mathbf{A}_{n,k}^c - \sum_{k=1}^n (\mathbf{A}_{n,k}^c)^2. \end{aligned} \tag{4.41}$$

Hence another application of Lemma 4.9 yields the claim. We now finalize the proof. Let  $0 < x < \infty$  and denote with  $S = [x/\sigma^2, 2x/\sigma^2]$ . Then for large enough  $n$ , it follows from the mean value theorem that

$$P(Z_{\Lambda_n} \leq x) - P(Z_{\Lambda} \leq x) \geq x \left( \frac{1}{\sqrt{\sigma_n^2}} - \frac{1}{\sqrt{\sigma^2}} \right) \inf_{y \in S} \phi(y),$$

where  $\phi(y)$  denotes the density function of the Gaussian standard distribution. Using the fact that  $\sigma^2 = 1$ , we further obtain

$$P(Z_{\Lambda_n} \leq x) - P(Z_{\Lambda} \leq x) \geq C_x (\sigma^2 - \sigma_n^2)$$

for some  $C_x > 0$ . Employing the lower bound of (4.40) then yields the desired result. □

### 4.2. Proofs of Section 3

**Proof of Theorem 3.2.** The main idea of the proof is based on a conditioning argument, similar in spirit to the approach in [27]. To this end, we first require some notation. Put  $n = 2KL$  for  $L \sim n$  and  $3 \leq K < \infty$ ,  $K \in \mathbb{N}$  to be specified later. To simplify the exposition, we also assume that  $L \in \mathbb{N}$ , see the very last comment at the end of the proof on how to remove this assumption. We make the convention that  $X_k = 0$  for  $k \notin \{1, \dots, n\}$ . Put  $\mathcal{I}_l = \{k : (l-1)K < k \leq lK\}$ ,  $\mathcal{I}_l^* = \mathcal{I}_l \cup \{(l-1)K\}$ ,  $\mathcal{F} = \mathcal{F}_L^{(e)} = \sigma(\varepsilon_k, k \in \mathcal{I}_{2l}^*, 1 \leq l \leq L)$ , and introduce the block variables

$$V_l = \sum_{k \in \mathcal{I}_l} X_k, \quad \bar{V}_l = V_l - \mathbb{E}[V_l | \mathcal{F}_L^{(e)}].$$

The fact that  $\mathcal{I}_l^*$  also contains the left endpoint of the interval (unlike to  $\mathcal{I}_l$ ) is important in the sequel. We denote the corresponding even (*e*) and odd (*o*) partial sums with

$$\begin{aligned} S_L^{(o)}(\bar{V}) &= n^{-1/2} \sum_{l=0}^L \bar{V}_{2l+1}, & S_L^{(e)}(V) &= n^{-1/2} \sum_{l=1}^L V_{2l}, \\ R_L^{(o)}(V) &= n^{-1/2} \sum_{l=0}^L \mathbb{E}[V_{2l+1} | \mathcal{F}_L^{(e)}]. \end{aligned} \tag{4.42}$$

Hence we have the decomposition

$$n^{-1/2} S_n(X) = S_L^{(o)}(\bar{V}) + S_L^{(e)}(V) + R_L^{(o)}(V),$$

where we used  $X_k = 0$  for  $k \notin \{1, \dots, n\}$ . Next, consider the conditional probability measure  $P_{|\mathcal{F}}(\cdot) = P(\cdot | \mathcal{F}_L^{(e)})$ . Observe that  $\{\bar{V}_{2l+1}\}_{0 \leq l \leq L}$  is a sequence of centered, independent random variables under  $P_{|\mathcal{F}}$  since  $K \geq 3$  and due to the inclusion of the left endpoint in  $\mathcal{I}_l^*$ . Also note that  $S_L^{(e)}(V)$  and  $R_L^{(o)}(V)$  are  $\mathcal{F}_L^{(e)}$ -measurable. We make heavy use of these properties in the

sequel. Likewise, under the measure  $P_{|\mathcal{F}}$ , let  $Z_{L|\mathcal{F}}^{(o)}$  be a zero mean Gaussian random variable with (conditional) covariance operator

$$\mathbf{\Lambda}_{|\mathcal{F}}^{(o)}(\cdot) = n^{-1} \sum_{l=0}^L \mathbb{E}[\overline{V}_{2l+1} \langle \overline{V}_{2l+1}, \cdot \rangle | \mathcal{F}_L^{(e)}]. \quad (4.43)$$

Similarly, under  $P$ , let  $Z_L^{(o)}$  and  $Z_L^{(e,o)}$  be two mutually independent, zero mean Gaussian random variables, independent of  $\mathcal{F}_L^{(e)}$ , with covariance operators

$$\begin{aligned} \mathbf{\Lambda}^{(o)}(\cdot) &= \mathbb{E}[(S_L^{(o)}(\overline{V}), \cdot) S_L^{(o)}(\overline{V})] = \mathbb{E}[\mathbf{\Lambda}_{|\mathcal{F}}^{(o)}(\cdot)], \\ \mathbf{\Lambda}^{(e)}(\cdot) &= \mathbb{E}[(S_L^{(e)}(\overline{V}) + R_L^{(o)}(V), \cdot) (S_L^{(e)}(\overline{V}) + R_L^{(o)}(V))]. \end{aligned} \quad (4.44)$$

The proof relies on the following decomposition

$$\Delta_n(\mu) \leq \mathbb{E} \left[ \sup_{x \in \mathbb{R}} (\mathbf{I}_L(x) + \mathbf{II}_L(x)) \right] + \sup_{x \in \mathbb{R}} |\mathbb{E}[\mathbf{III}_L(x)]| + \sup_{x \in \mathbb{R}} \mathbf{IV}_L(x), \quad (4.45)$$

where

$$\begin{aligned} \mathbf{I}_L(x) &= |P_{|\mathcal{F}}(\|S_L^{(o)}(\overline{V}) + S_L^{(e)}(V) + R_L^{(o)}(V) + \mu\|_{\mathbb{H}} \leq x) \\ &\quad - P_{|\mathcal{F}}(\|Z_{L|\mathcal{F}}^{(o)} + S_L^{(e)}(V) + R_L^{(o)}(V) + \mu\|_{\mathbb{H}} \leq x)|, \\ \mathbf{II}_L(x) &= |P_{|\mathcal{F}}(\|Z_{L|\mathcal{F}}^{(o)} + S_L^{(e)}(V) + R_L^{(o)}(V) + \mu\|_{\mathbb{H}} \leq x) \\ &\quad - P_{|\mathcal{F}}(\|Z_L^{(o)} + S_L^{(e)}(V) + R_L^{(o)}(V) + \mu\|_{\mathbb{H}} \leq x)|, \\ \mathbf{III}_L(x) &= P_{|\mathcal{F}}(\|Z_L^{(o)} + S_L^{(e)}(V) + R_L^{(o)}(V) + \mu\|_{\mathbb{H}} \leq x) \\ &\quad - P_{|\mathcal{F}}(\|Z_L^{(o)} + Z_L^{(e,o)} + \mu\|_{\mathbb{H}} \leq x), \\ \mathbf{IV}_L(x) &= |P_{|\mathcal{F}}(\|Z_L^{(o)} + Z_L^{(e,o)} + \mu\|_{\mathbb{H}} \leq x) - P(\|Z_\Lambda + \mu\|_{\mathbb{H}} \leq x)|. \end{aligned} \quad (4.46)$$

Below, we derive separate bounds for all four quantities. The key step is dealing with  $\mathbf{II}_L(x)$ , where the dependence gets disentangled asymptotically to independence.

*Case  $\mathbf{I}_L(x)$ :* Here we apply Lemma 4.3 under the conditional probability  $P_{|\mathcal{F}}$ , which makes all involved quantities random. To ensure applicability, we need to rule out any pathologies in advance. In particular, we need to control the eigenvalues of the random operator  $\mathbf{\Lambda}_{|\mathcal{F}}^{(o)}$ . To this end, let  $\mathcal{C}_\delta = \{\|\mathbf{\Lambda}_{|\mathcal{F}}^{(o)} - \mathbf{\Lambda}\|_{\mathcal{H}} \leq \delta\}$ ,  $\delta > 0$ , and put  $\mathcal{I}'_l = \{k : (l-1)K + 1 < k \leq (lK - 1)\}$ ,  $1 \leq l \leq 2L$ . Then we have by independence

$$\mathbb{E}[X_k | \mathcal{F}_L^{(e)}] = \mathbb{E}[X_k], \quad k \in \mathcal{I}'_{2l+1}. \quad (4.47)$$

Using (4.47), routine calculations reveal that

$$\|\mathbb{E}[\overline{V}_{2l+1} \langle \overline{V}_{2l+1}, \cdot \rangle] - \mathbb{E}[V_{2l+1} \langle V_{2l+1}, \cdot \rangle]\|_{\mathcal{H}} < \infty$$

since  $K < \infty$ . Applying this bound then leads to

$$\|\Lambda^{(o)} - \Lambda\|_{\mathcal{H}} \lesssim \frac{L\mathbb{E}[\|X_0\|_{\mathbb{H}}^2]}{n} \lesssim \frac{\mathbb{E}[\|X_0\|_{\mathbb{H}}^2]}{K}. \tag{4.48}$$

Selecting  $K$  sufficiently large (but finite), we thus obtain

$$C_\delta^c = \{\|\Lambda_{|\mathcal{F}}^{(o)} - \Lambda\|_{\mathcal{H}} > \delta\} \subseteq \{\|\Lambda_{|\mathcal{F}}^{(o)} - \Lambda^{(o)}\|_{\mathcal{H}} > \delta/2\}, \tag{4.49}$$

and hence by Markov's inequality

$$P(C_\delta^c) \leq (\delta/2)^{-2}\mathbb{E}[\|\Lambda_{|\mathcal{F}}^{(o)} - \Lambda^{(o)}\|_{\mathcal{H}}^2].$$

Denote with  $\|\cdot\|_{\mathcal{S}}$  the Hilbert–Schmidt norm. Since  $\|\cdot\|_{\mathcal{H}} \leq \|\cdot\|_{\mathcal{S}}$  and the space of Hilbert–Schmidt operators forms again a Hilbert space, using a Rosenthal inequality for Hilbert space valued sequences (cf. [28]), we obtain

$$\begin{aligned} n\mathbb{E}[\|\Lambda_{|\mathcal{F}}^{(o)} - \Lambda^{(o)}\|_{\mathcal{H}}^2] &\lesssim \sum_{l=0}^L \mathbb{E}[\|\mathbb{E}[\bar{V}_{2l+1}(\bar{V}_{2l+1}, \cdot)|\mathcal{F}_L^{(e)}] - \mathbb{E}[\bar{V}_{2l+1}(\bar{V}_{2l+1}, \cdot)]\|_{\mathcal{S}}^2] \\ &\lesssim \sum_{l=0}^L \mathbb{E}[\|\bar{V}_{2l+1}\|_{\mathbb{H}}^4] \lesssim \sum_{l=0}^L \mathbb{E}\left[\left\|\sum_{k \in \mathcal{I}_l} X_k\right\|_{\mathbb{H}}^4\right] \lesssim L\mathbb{E}[\|X_0\|_{\mathbb{H}}^4]. \end{aligned}$$

This together with the above yields  $P(C_\delta^c) \lesssim (\delta^2 L)^{-1}\mathbb{E}[\|X_0\|_{\mathbb{H}}^4]$ . Next, let

$$T_l = \mathbb{E}[\|\bar{V}_{2l+1}\|_{\mathbb{H}}^2 \mathbf{1}(\|\bar{V}_{2l+1}\|_{\mathbb{H}} \geq n^{1/2}) + n^{-1/2}\|\bar{V}_{2l+1}\|_{\mathbb{H}}^3 |\mathcal{F}_L^{(e)}],$$

where routine calculations reveal that

$$\mathbb{E}[T_l] \lesssim n^{-1/2}\mathbb{E}[\|\bar{V}_{2l+1}\|_{\mathbb{H}}^3] \lesssim n^{-1/2}\mathbb{E}[\|X_0\|_{\mathbb{H}}^3]. \tag{4.50}$$

Note that  $\{T_l\}_{0 \leq l \leq L}$  is a sequence of independent, real valued random variables. Denote with

$$\mathcal{D} = \left\{ \left| \sum_{l=0}^L (T_l - \mathbb{E}[T_l]) \right| \leq L^{1/2} \right\}. \tag{4.51}$$

Using Burkholders, triangle and Jensens inequality, we get

$$\left\| \sum_{l=0}^L (T_l - \mathbb{E}[T_l]) \right\|_{3/2}^{3/2} \lesssim \sum_{l=0}^L \|T_l - \mathbb{E}[T_l]\|_{3/2}^{3/2} \lesssim Ln^{-3/4}\mathbb{E}[\|X_0\|_{\mathbb{H}}^{9/2}].$$

Hence we conclude via Markovs inequality

$$P(\mathcal{D}^c) \lesssim L^{1-3/4} n^{-3/4} \mathbb{E}[\|X_0\|_{\mathbb{H}}^{9/2}] \lesssim L^{-1/2} \mathbb{E}[\|X_0\|_{\mathbb{H}}^{9/2}]. \tag{4.52}$$

We are now in position to derive the actual bound. Observe that

$$\mathbb{E}\left[\sup_{x \in \mathbb{R}} \mathbf{I}_L(x)\right] \leq 2P(\mathcal{C}_\delta^c) + 2P(\mathcal{D}^c) + \mathbb{E}\left[\sup_{x \in \mathbb{R}} \mathbf{I}_L(x) \mathbf{1}_{\mathcal{C}_\delta \cap \mathcal{D}}\right]. \tag{4.53}$$

Since  $S_L^{(e)}(V), R_L^{(o)}(V) \in \mathcal{F}_L^{(e)}$  and  $\{\bar{V}_l\}_{0 \leq l \leq L}$  is a sequence of independent zero mean random variables under  $P_{|\mathcal{F}}$ , applying Lemma 4.3 leads to

$$\sup_{x \in \mathbb{R}} \mathbf{I}_L(x) \mathbf{1}_{\mathcal{C}_\delta \cap \mathcal{D}} \leq C_{|\mathcal{F}} \mathbf{1}_{\mathcal{C}_\delta \cap \mathcal{D}} (1 + \|S_L^{(e)}(V) + R_L^{(o)}(V) + \mu\|_{\mathbb{H}}^3) n^{-1} \sum_{l=0}^L T_l. \tag{4.54}$$

Since  $\lambda_{13} > 0$ , selecting  $\delta > 0$  sufficiently small (and  $K = K_\delta$  sufficiently large) and using Lemma 4.1, we may bound  $C_{|\mathcal{F}} \mathbf{1}_{\mathcal{C}_\delta} \leq C_\lambda$ , where  $C_\lambda$  only depends on  $\lambda_{13}$ . In addition, by construction of the set  $\mathcal{D}$ , we have

$$\begin{aligned} & \mathbb{E}\left[\mathbf{1}_{\mathcal{D}} (1 + \|S_L^{(e)}(V) + R_L^{(o)}(V) + \mu\|_{\mathbb{H}}^3) n^{-1} \sum_{l=0}^L T_l\right] \\ & \lesssim \mathbb{E}\left[1 + \|S_L^{(e)}(V) + R_L^{(o)}(V) + \mu\|_{\mathbb{H}}^3\right] \left(n^{-1} \sum_{l=0}^L \mathbb{E}[T_l] + L^{-1/2}\right). \end{aligned}$$

Using (4.50) and Rosenthals inequality, the above is further bounded by  $L^{-1/2} (1 + \|\mu\|_{\mathbb{H}}^3) \times \mathbb{E}[\|X_0\|_{\mathbb{H}}^3]$ , and hence we obtain

$$\mathbb{E}\left[\mathbf{1}_{\mathcal{D}} (1 + \|S_L^{(e)}(V) + R_L^{(o)}(V) + \mu\|_{\mathbb{H}}^3) n^{-1} \sum_{l=0}^L T_l\right] \lesssim L^{-1/2} (1 + \|\mu\|_{\mathbb{H}}^3) \mathbb{E}[\|X_0\|_{\mathbb{H}}^3]. \tag{4.55}$$

We thus conclude from (4.52)

$$\mathbb{E}\left[\sup_{x \in \mathbb{R}} \mathbf{I}_L(x)\right] \lesssim L^{-1/2} (1 + \|\mu\|_{\mathbb{H}}^3) \mathbb{E}[\|X_0\|_{\mathbb{H}}^{9/2}]. \tag{4.56}$$

*Case  $\mathbf{II}_L(x)$ :* Under the measure  $P_{|\mathcal{F}}$ , let  $\{\xi_{k|\mathcal{F}}\}_{1 \leq k \leq n}$  be a sequence of independent, zero mean Gaussian random variables, where each  $\xi_{k|\mathcal{F}}$  has covariance operator  $\mathbf{\Lambda}_{|\mathcal{F}}^{(o)}$  (instead of sample size  $n$  we could also select  $N > n$ , but  $n$  is sufficient). Similarly, let  $\{\xi_k\}_{1 \leq k \leq n}$  be i.i.d. Gaussian sequences independent of  $\mathcal{F}_L^{(e)}$ , with covariance operator  $\mathbf{\Lambda}^{(o)}$ . Next, introduce the mixed partial sums

$$\sqrt{n} W_{k|\mathcal{F}} = \sum_{j=1}^{k-1} \xi_{j|\mathcal{F}} + \sum_{j=k+1}^n \xi_j, \quad 1 \leq k \leq n, \tag{4.57}$$

and the corresponding conditional covariance operators

$$\begin{aligned}\tilde{\Lambda}_{k|\mathcal{F}}(\cdot) &= \mathbb{E}[\langle W_{k|\mathcal{F}}, \cdot \rangle W_{k|\mathcal{F}} | \mathcal{F}_L^{(e)}] \\ &= \frac{k-1}{n} \Lambda_{|\mathcal{F}}^{(o)}(\cdot) + \frac{n-k}{n} \Lambda^{(o)}(\cdot),\end{aligned}\tag{4.58}$$

with eigenvalues  $\tilde{\lambda}_{k,j|\mathcal{F}}$ ,  $j \in \mathbb{N}$ . Since these are random, we need to control them as in the previous case  $\mathbf{I}_L(x)$ . To this end, define the set

$$\tilde{\mathcal{C}}_\delta = \left\{ \max_{1 \leq k \leq n} \|\tilde{\Lambda}_{k|\mathcal{F}} - \Lambda\|_{\mathcal{H}} \geq \delta \right\}.\tag{4.59}$$

Using (4.47) and (4.48), yields

$$\max_{1 \leq k \leq n} \|\mathbb{E}[\tilde{\Lambda}_{k|\mathcal{F}}] - \Lambda\|_{\mathcal{H}} \leq \|\Lambda^{(o)} - \Lambda\|_{\mathcal{H}} + \|\Lambda\|_{\mathcal{H}}/n \lesssim \frac{L\mathbb{E}[\|X_0\|_{\mathbb{H}}^2]}{n} = \frac{\mathbb{E}[\|X_0\|_{\mathbb{H}}^2]}{K}.$$

Proceeding similarly as in the treatment of  $\mathcal{C}_\delta^c$  and using Burkholders inequality for Hilbert-space valued martingales (cf. [28]), it follows that for large enough  $K$

$$P(\tilde{\mathcal{C}}_\delta) \lesssim (\delta^2 L)^{-1} \mathbb{E}[\|X_0\|_{\mathbb{H}}^4], \quad \delta > 0 \text{ sufficiently small.}\tag{4.60}$$

Next, following the same approach as in Lemma 4.7 and using similar arguments as in the previous case  $\mathbf{I}_L(x)$ , we obtain from (4.60) that

$$\begin{aligned}\mathbb{E}\left[\sup_{x \in \mathbb{R}} \mathbf{II}_L(x)\right] &\leq 2P(\tilde{\mathcal{C}}_\delta^c) + \mathbb{E}\left[\sup_{x \in \mathbb{R}} \mathbf{II}_L(x) \mathbf{1}_{\tilde{\mathcal{C}}_\delta}\right] \\ &\lesssim L^{-1/2} (1 + \|\mu\|_{\mathbb{H}}^3) \mathbb{E}[\|X_0\|_{\mathbb{H}}^4] + \mathbb{E}\left[\sum_{k=1}^n \int_{\mathbb{R}} \frac{Q_{k|\mathcal{F}}(t)}{t} dt \mathbf{1}_{\tilde{\mathcal{C}}_\delta}\right],\end{aligned}\tag{4.61}$$

where

$$Q_{k|\mathcal{F}}(t) = |t| \left| \mathbb{E}\left[\exp(it \|W_{k|\mathcal{F}}\|_{\mathbb{H}}^2) | \mathcal{F}_L^{(e)}\right] \left( \mathbb{E}[\|\xi_{k|\mathcal{F}}\|_{\mathbb{H}}^2 | \mathcal{F}_L^{(e)}] - \mathbb{E}[\|\xi_k\|_{\mathbb{H}}^2 | \mathcal{F}_L^{(e)}] \right) \right|, \quad t \in \mathbb{R}.$$

By Lemma 4.2, we have the bound

$$\left| \mathbb{E}\left[\exp(it \|W_{k|\mathcal{F}}\|_{\mathbb{H}}^2) | \mathcal{F}_L^{(e)}\right] \right| \mathbf{1}_{\tilde{\mathcal{C}}_\delta} \leq \prod_{j=1}^{\infty} (1 + 4t^2 \tilde{\lambda}_{k,j|\mathcal{F}}^2)^{-1/4} \mathbf{1}_{\tilde{\mathcal{C}}_\delta},\tag{4.62}$$

where we recall that  $\tilde{\lambda}_{k,j|\mathcal{F}}$  denote the eigenvalues of the covariance operator  $\tilde{\Lambda}_{k|\mathcal{F}}$ . In particular, they are bounded away from zero uniformly on the set  $\tilde{\mathcal{C}}_\delta$  for  $\delta > 0$  sufficiently small due to Lemma 4.1. Observe next that by independence of  $\{\xi_k\}_{k \in \mathbb{Z}}$  from  $\mathcal{F}_L^{(e)}$

$$\left| \mathbb{E}[\|\xi_{k|\mathcal{F}}\|_{\mathbb{H}}^2 | \mathcal{F}_L^{(e)}] - \mathbb{E}[\|\xi_k\|_{\mathbb{H}}^2 | \mathcal{F}_L^{(e)}] \right| = n^{-1} \left| \sum_{l=0}^L (\mathbb{E}[\|\bar{V}_{2l+1}\|_{\mathbb{H}}^2 | \mathcal{F}_L^{(e)}] - \mathbb{E}[\|\bar{V}_{2l+1}\|_{\mathbb{H}}^2]) \right|.$$

Since  $\{\mathbb{E}[\|\bar{V}_{2l+1}\|_{\mathbb{H}}^2|\mathcal{F}_L^{(e)}]\}_{0 \leq l \leq L}$  is a sequence of independent, real-valued random variables, Jensen's and Rosenthal's inequality yield

$$\mathbb{E}[\|\mathbb{E}[\|\xi_{k|\mathcal{F}}\|_{\mathbb{H}}^2|\mathcal{F}_L^{(e)}] - \mathbb{E}[\|\xi_k\|_{\mathbb{H}}^2|\mathcal{F}_L^{(e)}]\|] \lesssim L^{-1/2} \mathbb{E}[\|X_0\|_{\mathbb{H}}^4]^{1/2}. \quad (4.63)$$

Combining (4.62) with (4.63), we thus obtain

$$\mathbb{E}\left[\sum_{k=1}^n \int_{\mathbb{R}} \frac{Q_{k|\mathcal{F}}(t)}{t} dt \mathbf{1}_{\tilde{\mathcal{C}}_{\delta}}\right] \lesssim L^{-1/2} \mathbb{E}[\|X_0\|_{\mathbb{H}}^4]. \quad (4.64)$$

Together with (4.61), this yields the final bound

$$\mathbb{E}\left[\sup_{x \in \mathbb{R}} \mathbf{III}_L(x)\right] \lesssim L^{-1/2} \mathbb{E}[\|X_0\|_{\mathbb{H}}^4]. \quad (4.65)$$

*Case  $\mathbf{III}_L(x)$ :* By independence, we have

$$\begin{aligned} \mathbb{E}[\mathbf{III}_L(x)] &= P(\|Z_L^{(o)} + S_L^{(e)}(V) + R_L^{(o)}(V) + \mu\|_{\mathbb{H}} \leq x) \\ &\quad - P(\|Z_L^{(o)} + Z_L^{(e,o)} + \mu\|_{\mathbb{H}} \leq x), \end{aligned}$$

hence we may directly appeal to Lemma 4.3. Routine calculations then reveal

$$\sup_{x \in \mathbb{R}} |\mathbb{E}[\mathbf{III}_L(x)]| \lesssim L^{-1/2} (1 + \|\mu\|_{\mathbb{H}}^3) \mathbb{E}[\|X_0\|_{\mathbb{H}}^3]. \quad (4.66)$$

*Case  $\mathbf{IV}_L(x)$ :* One readily verifies

$$|\mathrm{tr}(\mathbf{\Lambda}) - \mathrm{tr}(\mathbf{\Lambda}^{(o)} + \mathbf{\Lambda}^{(e)})| \lesssim L^{-1} \mathbb{E}[\|X_0\|_{\mathbb{H}}^2].$$

Since we have by independence

$$P_{|\mathcal{F}}(\|Z_L^{(o)} + Z_L^{(e,o)} + \mu\|_{\mathbb{H}} \leq x) = P(\|Z_L^{(o)} + Z_L^{(e,o)} + \mu\|_{\mathbb{H}} \leq x),$$

an application of Lemma 4.7 then yields

$$\sup_{x \in \mathbb{R}} \mathbf{IV}_L(x) \lesssim L^{-1} \mathbb{E}[\|X_0\|_{\mathbb{H}}^2]. \quad (4.67)$$

Since  $n \sim L$ , combining all four bounds completes the proof. As a final remark, let us elaborate on the case where  $L \notin \mathbb{N}$ . In this case, we may have a slightly smaller additional remainder term  $\tilde{R}_{L+1}$  in the decomposition

$$n^{-1/2} S_n(X) = S_L^{(o)}(\bar{V}) + S_L^{(e)}(V) + R_L^{(o)}(V) + \tilde{R}_{L+1}, \quad (4.68)$$

which we can always add to the last summand, be it even or odd. This just results in more complicated notation, but the proof remains the same.  $\square$

**Proof of Corollary 3.3.** If  $L = n/m \in \mathbb{N}$ , writing

$$(nm)^{-1/2} S_n(X) = (n/m)^{-1/2} \sum_{l=1}^L B_l/m, \quad B_l = \sum_{k=(l-1)m+1}^{lm} X_k,$$

we may directly apply Theorem 3.2. If  $L \notin \mathbb{N}$ , we have an additional remainder part, which however does not require any particular different treatment, see also the remark around (4.68).  $\square$

**Proof of Corollary 3.4.** By virtue of Theorem 3.2, we may almost identically repeat the proof of Theorem 2.1 together with Corollary 2.2 in [26].  $\square$

## Acknowledgements

I would like to thank the Associate Editor and the anonymous Reviewers for the constructive comments and suggestions. The generous help has been a major benefit.

FOR 1735 *Structural Inference in Statistics: Adaptation and Efficiency* is gratefully acknowledged.

## References

- [1] Basu, A.K. (1988). Uniform and nonuniform estimates in the CLT for Banach valued dependent random variables. *J. Multivariate Anal.* **25** 153–163. [MR0940537](#)
- [2] Bentkus, V. and Götze, F. (1996). Optimal rates of convergence in the CLT for quadratic forms. *Ann. Probab.* **24** 466–490. [MR1387646](#)
- [3] Bentkus, V.Y. (1984). Asymptotic expansions for distributions of sums of independent random elements in a Hilbert space. *Lith. Math. J.* **24** 305–319.
- [4] Berry, A.C. (1941). The accuracy of the Gaussian approximation to the sum of independent variates. *Trans. Amer. Math. Soc.* **49** 122–136. [MR0003498](#)
- [5] Beveridge, S. and Nelson, C.R. (1981). A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the “business cycle”. *Journal of Monetary Economics* **7** 151–174.
- [6] Bjarnotas, V. and Paulauskas, V.I. (1980). A nonuniform estimate in the central limit theorem in some Banach spaces. *Lith. Math. J.* **19** 177–190.
- [7] Bosq, D. (2000). *Linear Processes in Function Spaces: Theory and Applications. Lecture Notes in Statistics* **149**. New York: Springer. [MR1783138](#)
- [8] Bosq, D. (2003). Berry–Esseen inequality for linear processes in Hilbert spaces. *Statist. Probab. Lett.* **63** 243–247. [MR1986323](#)
- [9] Bosq, D. (2004). Erratum and complements to: “Berry–Esseen inequality for linear processes in Hilbert spaces” [*Statist. Probab. Lett.* **63** (2003), no. 3, 243–247; [MR1986323](#)]. *Statist. Probab. Lett.* **70** 171–174. [MR2100824](#)
- [10] Butzer, P.L., Hahn, L. and Roeckerath, M.T. (1983). Central limit theorem and weak law of large numbers with rates for martingales in Banach spaces. *J. Multivariate Anal.* **13** 287–301. [MR0705552](#)
- [11] Chen, L.H.Y. and Shao, Q.-M. (2004). Normal approximation under local dependence. *Ann. Probab.* **32** 1985–2028. [MR2073183](#)

- [12] Dedecker, J. and Merlevède, F. (2003). The conditional central limit theorem in Hilbert spaces. *Stochastic Process. Appl.* **108** 229–262. [MR2019054](#)
- [13] Dedecker, J. and Prieur, C. (2005). New dependence coefficients. Examples and applications to statistics. *Probab. Theory Related Fields* **132** 203–236. [MR2199291](#)
- [14] El Machkouri, M. (2010). Berry–Esseen’s central limit theorem for non-causal linear processes in Hilbert space. *Afr. Diaspora J. Math.* **10** 81–86. [MR2774260](#)
- [15] El Machkouri, M. and Volný, D. (2004). On the central and local limit theorem for martingale difference sequences. *Stoch. Dyn.* **4** 153–173. [MR2069687](#)
- [16] Esseen, C.-G. (1945). Fourier analysis of distribution functions. A mathematical study of the Laplace–Gaussian law. *Acta Math.* **77** 1–125. [MR0014626](#)
- [17] Gohberg, I., Goldberg, S. and Kaashoek, M.A. (2003). *Basic Classes of Linear Operators*. Basel: Birkhäuser. [MR2015498](#)
- [18] Gordin, M.I. (1969). The central limit theorem for stationary processes. *Sov. Math., Dokl.* **10** 1174–1176.
- [19] Götze, F. (1979). Asymptotic expansions for bivariate von Mises functionals. *Z. Wahrsch. Verw. Gebiete* **50** 333–355. [MR0554550](#)
- [20] Götze, F. and Hipp, C. (1983). Asymptotic expansions for sums of weakly dependent random vectors. *Z. Wahrsch. Verw. Gebiete* **64** 211–239. [MR0714144](#)
- [21] Götze, F. and Hipp, C. (1989). Asymptotic expansions for potential functions of i.i.d. random fields. *Probab. Theory Related Fields* **82** 349–370. [MR1001518](#)
- [22] Hannan, E.J. (1973). Central limit theorems for time series regression. *Z. Wahrsch. Verw. Gebiete* **26** 157–170. [MR0331683](#)
- [23] Hörmann, S., Horváth, L. and Reeder, R. (2013). A functional version of the ARCH model. *Econometric Theory* **29** 267–288. [MR3042756](#)
- [24] Hörmann, S. and Kokoszka, P. (2010). Weakly dependent functional data. *Ann. Statist.* **38** 1845–1884. [MR2662361](#)
- [25] Horváth, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications*. Springer Series in Statistics. New York: Springer. [MR2920735](#)
- [26] Jirak, M. (2015). On non-uniform Berry–Esseen bounds for time series. *Probab. Math. Statist.* **35** 1–14. [MR3355579](#)
- [27] Jirak, M. (2016). Berry–Esseen theorems under weak dependence. *Ann. Probab.* **44** 2024–2063. [MR3502600](#)
- [28] Junge, M. and Xu, Q. (2003). Noncommutative Burkholder/Rosenthal inequalities. *Ann. Probab.* **31** 948–995. [MR1964955](#)
- [29] Merlevède, F., Peligrad, M. and Utev, S. (1997). Sharp conditions for the CLT of linear processes in a Hilbert space. *J. Theoret. Probab.* **10** 681–693. [MR1468399](#)
- [30] Mielkaitis, E. and Paulauskas, V. (2011). Rates of convergence in the CLT for linear random fields. *Lith. Math. J.* **51** 233–250. [MR2805741](#)
- [31] Nagaev, S.V. (1984). Estimates of Berry–Esseen type for sums of random variables with values in a Hilbert space. *Sov. Math., Dokl.* **29** 692–693.
- [32] Nagaev, S.V. and Chebotarev, V.I. (2005). On the accuracy of Gaussian approximation in Hilbert space [Translation of *Mat. Tr.* **7** (2004), no. 1, 91–152; [MR2068278](#)]. *Siberian Adv. Math.* **15** 11–73. [MR2141789](#)
- [33] Osekowski, A. (2012). A note on the Burkholder–Rosenthal inequality. *Bull. Pol. Acad. Sci. Math.* **60** 177–185. [MR2914545](#)
- [34] Petrov, V.V. (1995). *Limit Theorems of Probability Theory: Sequences of Independent Random Variables*. Oxford Studies in Probability **4**. New York: Oxford Univ. Press. [MR1353441](#)

- [35] Prokhorov, Y.V. and Ulyanov, V.V. (2013). Some approximation problems in statistics and probability. In *Limit Theorems in Probability, Statistics and Number Theory. Springer Proc. Math. Stat.* **42** 235–249. Heidelberg: Springer. [MR3079145](#)
- [36] Rhee, W. and Talagrand, M. (1981). On Berry–Esseen type bounds for  $m$ -dependent random variables valued in certain Banach spaces. *Z. Wahrsch. Verw. Gebiete* **58** 433–451. [MR0638030](#)
- [37] Rhee, W. and Talagrand, M. (1986). Uniform bound in the central limit theorem for Banach space valued dependent random variables. *J. Multivariate Anal.* **20** 303–320. [MR0866077](#)
- [38] Sunklodas, J. (1997). On a lower bound of the rate of convergence in the central limit theorem for  $m$ -dependent random variables. *Lith. Math. J.* **37** 291–299.
- [39] Sweeting, T.J. (1977). Speeds of convergence for the multidimensional central limit theorem. *Ann. Probab.* **5** 28–41. [MR0428400](#)
- [40] Tihomirov, A.N. (1981). Convergence rate in the central limit theorem for weakly dependent random variables. *Theory Probab. Appl.* **25** 790–809.
- [41] Tikhomirov, A.N. (1991). On the accuracy of the normal approximation of the probability of sums of weakly dependent Hilbert-space-valued random variables hitting a ball. I. *Theory Probab. Appl.* **36** 738–751.
- [42] Ulyanov, V.V. (1986). Asymptotic expansions for distributions of sums of independent random variables in  $H$ . *Theory Probab. Appl.* **31** 25–39.
- [43] Wu, W.B. (2005). Nonlinear system theory: Another look at dependence. *Proc. Natl. Acad. Sci. USA* **102** 14150–14154 (electronic). [MR2172215](#)
- [44] Yurinskiĭ, V.V. (1982). On the accuracy of normal approximation of the probability of hitting a ball. *Theory Probab. Appl.* **27** 280–289.
- [45] Zaleskiĭ, B.A., Sazonov, V.V. and Ulyanov, V.V. (1988). A sharp estimate for the accuracy of the normal approximation in a Hilbert space. *Theory Probab. Appl.* **33** 700–701.
- [46] Zuparov, T.M. (1983). Estimates for the rate of convergence in the central limit theorem for absolutely regular random variables with values in some Banach spaces. *Sov. Math., Dokl.* **28** 475–478.

*Received October 2014 and revised May 2016*