On the Breuer–Major Theorem on Hilbert spaces, functional time series, and feature learning[⋆]

Marie-Christine Düker[a], Pavlos Zouboulogloub[b,*]

[a]Technical University of Munich, Department of Mathematics, Boltzmannstraße 3, Garching n. Munich, 85748, Germany
[b]University of Münster, Institute for Mathematical Stochastics, Orléans-Ring 10, Münster, 48149, Germany

ARTICLE INFO

Keywords:
Breuer–Major theorem
functional time series
sample autocovariance operators
continuum random features
kernel autocovariance operators

ABSTRACT

The Hilbert space-valued Breuer–Major theorem of [4] holds under either of two sets of sufficient conditions. The one that is the focus of this work requires summability of the qth powers of the trace norms of the autocovariance operators of the underlying Gaussian process, together with a regularity condition on the transformation G, where q is its Hermite rank. We demonstrate that these regularity conditions are satisfied for a broad class of functionals in statistics and machine learning, yielding both known and new central limit theorems.

1. Introduction

Consider a centered stationary Gaussian sequence $\{X_k\}_{k\in\mathbb{Z}}$, defined on a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and taking values in a real separable Hilbert space $\mathcal{H}_1$, with nondegenerate covariance operator Q, meaning $\ker Q = \{0\}$. Let $G : \mathcal{H}_1 \to \mathcal{H}_2$ be a measurable map into another real separable Hilbert space. Throughout, suppose that $\mathrm{E}\, G[X_0] = 0$ and $G[X_0] \in L^2(\Omega : \mathcal{H}_2)$, that is, $\mathrm{E}\, \|G[X_0]\|^2_{\mathcal{H}_2} < \infty$. We study central limit theorems for the normalized partial sums of the subordinated process $\{G[X_k]\}_{k\in\mathbb{Z}}$

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} G[X_k] \xrightarrow{d} \mathcal{N}(0, \mathcal{T}_Z) \quad \text{in } \mathcal{H}_2, \qquad \mathcal{T}_Z \doteq \sum_{v\in\mathbb{Z}} \mathrm{E}[G[X_0] \otimes G[X_v]]. \tag{1}$$

Here $\mathcal{N}(0, \mathcal{T}_Z)$ is a centered Gaussian law with *long-run covariance* $\mathcal{T}_Z$, assuming it converges in trace-norm.

Limits as in (1) characterize the asymptotic distributions of function- or operator-valued statistics and can enable downstream statistical tasks, such as the design of hypothesis tests and confidence regions in functional time series analysis. In machine learning, analogous limits characterize the asymptotic behavior of empirical kernel operators and random-feature approximations, including those arising from randomly initialized neural operators; see also the applications in [4, Section 7].

Following the seminal work of Breuer and Major [1], such results are referred to as Breuer–Major theorems. Our companion paper [4, Theorem 3.1] establishes (1) under either of two sets of sufficient conditions, using the fourth-moment theorem on Hilbert spaces [3]. Let $q = \mathrm{rank}(G) \geq 1$ denote the Hermite rank of G, namely the least positive order of a nonzero Hermite component; see [4, Definition 4.1].

For Hilbert spaces E, F, let $\mathcal{L}(E, F)$ denote the space of bounded linear operators, equipped with the operator norm $\|\cdot\|_{\mathrm{op}}$. For $1 \leq r < \infty$, write $\mathcal{S}_r(E, F)$ for the Schatten class with norm $\|T\|_{\mathcal{S}_r(E,F)} = (\sum_j s_j(T)^r)^{1/r}$, where $s_j(T)$ are the singular values of T, and abbreviate $\mathcal{S}_r(E, E) = \mathcal{S}_r(E)$. In particular, $\mathcal{S}_1$ and $\mathcal{S}_2$ are the trace-class and Hilbert–Schmidt classes. We write Tr for the trace and T^* for the adjoint of T, and omit spaces from norms when unambiguous. All tensor products are completed Hilbert tensor products, and $E^{\otimes p}$ denotes the pth tensor power of E.

For $v \in \mathbb{Z}$, let $\Gamma(v)$ be the autocovariance operator (ACO) defined by $\Gamma(v)[h] = \mathrm{E}[\langle X_0, h\rangle_{\mathcal{H}_1} X_v]$. Let $R(v)$ be the autocorrelation operator associated with the white-noise coordinates of X_k, so that $\Gamma(v) = Q^{1/2} R(v) Q^{1/2}$. With this notation, we can state the two alternative sets of sufficient conditions.

Assumption A. *The autocorrelation operators satisfy* $\sum_{v\in\mathbb{Z}} \|R(v)\|^q_{\mathrm{op}} < \infty$.

[⋆] 2020 Mathematics Subject Classification: 60F05, 60G15, 62M10, 62M15, 62R10.
*Corresponding author
marie.dueker@tum.de (M. Düker); p.zouboulogloou@uni-muenster.de (P. Zouboulogloou)

For Assumption B, fix an orthonormal eigenbasis $\{e_j\}_{j\geq 1}$ of Q and an orthonormal basis $\{v_i\}_{i\geq 1}$ of $\mathcal{H}_2$. The Hilbert-valued Wiener–Itô decomposition [4, Section 4] gives

$$G[X_k] = \sum_{p=q}^{\infty} \sum_{i\geq 1} I_p \left[\sum_{\boldsymbol{j}\in\mathbb{N}^p} b_{i,\boldsymbol{j}}\, \varepsilon_{j_1 k} \otimes \cdots \otimes \varepsilon_{j_p k} \right] v_i \quad \text{in } L^2(\Omega : \mathcal{H}_2),$$

where $b_{i,\boldsymbol{j}}$ is symmetric in $\boldsymbol{j} = (j_1, \ldots, j_p)$ and $\varepsilon_{jk} \in \mathfrak{H}$ are the isonormal coordinate vectors. Following [4, Eq. (3.2)], define $C_p \in \mathcal{L}(\mathcal{H}_2, \mathcal{H}_1^{\otimes p})$ by

$$C_p[v_i] = \sum_{\boldsymbol{j}\in\mathbb{N}^p} b_{i,\boldsymbol{j}}\, e_{j_1} \otimes \cdots \otimes e_{j_p}, \qquad \text{and set } Q_p^{1/2} = (Q^{1/2})^{\otimes p}. \tag{2}$$

Assumption B. *Assume that the following two conditions hold:* [1]

$$\textstyle\sum_{v\in\mathbb{Z}} \|\Gamma(v)\|_{\mathcal{S}_1}^q < \infty \text{ and} \tag{3}$$

$$\text{For every } p \geq q, \text{ there exists } D_p \in \mathcal{L}(\mathcal{H}_2, \mathcal{H}_1^{\otimes p}) \text{ such that } C_p = Q_p^{1/2} D_p \text{ and } \textstyle\sum_{p=q}^{\infty} p!\, \|D_p\|_{\mathrm{op}}^2 < \infty. \tag{4}$$

Note that, from the Schatten ideal property, we have that $\|\Gamma(v)\|_{\mathcal{S}_1} \leq \mathrm{Tr}(Q)\, \|R(v)\|_{\mathrm{op}}$. For that reason, we call (A) *strong short-range dependence at rank q*, and (3) *weak short-range dependence at rank q*. The weak short-range dependence condition (3) alone does not suffice to ensure that the candidate long-run covariance operator $\mathcal{T}_Z$ in (1) is trace-class, and so the additional coefficient conditions in (4) are needed.

In finite dimensions, assuming Q is positive definite yields $\|R(v)\|_{\mathrm{op}} \leq \|Q^{-1/2}\|_{\mathrm{op}}^2\, \|\Gamma(v)\|_{\mathcal{S}_1}$, so (3) implies Assumption A without further regularity on G. In contrast, $Q^{-1/2}$ is always unbounded in infinite dimensions and so $\Gamma(v)$ can be easier to control than $R(v)$. This observations makes the trace-norm formulation more useful in, among other models, Hilbert space-valued linear and autoregressive processes; see [4, Sections 3.3–3.4].

In this article, we shed light on the regime described by Assumption B. Although (3) is often verifiable, the requirements involving C_p in (4) are considerably more technical. We identify tractable conditions on G that imply the coefficient requirements of (4) (Proposition 2.1), thereby bypassing the computation of the full array of Hermite coefficients. In Sections 3–6, we verify these conditions for four examples in statistics and machine learning: smoothed kurtosis operators, joint fixed-lag sample ACOs, continuum random-feature maps, and kernel ACOs. In particular:

Let G be any of the maps considered in Sections 3–6, satisfying the corresponding hypotheses therein. Let $\mathrm{rank}(G) = q$ and assume that $\sum_{v\in\mathbb{Z}} \|\Gamma(v)\|_{\mathcal{S}_1}^q < \infty$. Then the convergence in (1) holds.

2. A class of operators G satisfying (4)

Suppose that the orthonormal basis $\{e_j\}_{j\geq 1}$ are eigenfunctions of Q, i.e., $Q[e_j] = \lambda_j e_j$, where $\lambda_j > 0$ and $\sum_j \lambda_j = \mathrm{Tr}(Q) < \infty$. For $p \geq 1$, let $\mathcal{H}_1^{\odot p}$ denote the closed symmetric subspace of $\mathcal{H}_1^{\otimes p}$, and let $\mathrm{sym}_p : \mathcal{H}_1^{\otimes p} \to \mathcal{H}_1^{\odot p}$ be the canonical symmetrization (see [8, Eq. (B.3.1)]).

Proposition 2.1. *Let $A_0 \in \mathcal{H}_2$ and $A_r \in \mathcal{L}(\mathcal{H}_1^{\otimes r}, \mathcal{H}_2)$ for $1 \leq r \leq m$. Define, for $x \in \mathcal{H}_1$,*

$$P[x] = A_0 + \textstyle\sum_{r=1}^{m} A_r[x^{\otimes r}], \qquad G[x] = P[x] - \mathrm{E}\, P[X_0]. \tag{5}$$

Then $G[X_1] \in L^2(\Omega : \mathcal{H}_2)$ and, moreover, G satisfies condition (4).

Proof. For every $1 \leq r \leq m$, we have $\mathrm{E}\, \|A_r[X_1^{\otimes r}]\|_{\mathcal{H}_2}^2 \leq \|A_r\|_{\mathrm{op}}^2\, \mathrm{E}\, \|X_1\|_{\mathcal{H}_1}^{2r}$, where finiteness follows from Fernique's theorem. Hence $G[X_1] \in L^2(\Omega : \mathcal{H}_2)$. Moreover, $x^{\otimes r} = \mathrm{sym}_r[x^{\otimes r}]$, so replacing A_r by $A_r \circ \mathrm{sym}_r$ does not change G. We may therefore assume that $A_r = A_r \circ \mathrm{sym}_r$, and hence $A_r^*(\mathcal{H}_2) \subseteq \mathcal{H}_1^{\odot r}$.

[1] For each p, the bounded factorization in (4) is equivalent to $C_p(\mathcal{H}_2) \subseteq \mathrm{Ran}(Q_p^{1/2})$, by the closed graph theorem.

Fix $k \in \mathbb{Z}$ and set $e_{\boldsymbol{j}} = e_{j_1} \otimes \cdots \otimes e_{j_p}$ and $\varepsilon_{\boldsymbol{j},k} = \varepsilon_{j_1 k} \otimes \cdots \otimes \varepsilon_{j_p k}$. Let J_p denote the orthogonal projection onto the pth real-valued Wiener–Itô chaos. Write

$$G[x] = \sum_{r=1}^{m} G_r[x], \quad G_r[x] \doteq A_r[x^{\otimes r}] - \mathrm{E}(A_r[X_0^{\otimes r}]); \quad J_p\Big[\langle G_r[X_k], v_i\rangle_{\mathcal{H}_2}\Big] = I_p\Bigg[\sum_{\boldsymbol{j}\in\mathbb{N}^p} \Big\langle C_p^{(r)}[v_i], e_{\boldsymbol{j}}\Big\rangle_{\mathcal{H}_1^{\otimes p}} \varepsilon_{\boldsymbol{j},k}\Bigg]. \tag{6}$$

The identities in (6) allow us to identify the coefficient operators $C_p^{(r)}$ of G_r, and then those of G by taking the sum over r. Thus, identifying the symmetric kernel of this projection determines $C_p^{(r)}[v_i]$.

To introduce a candidate kernel, we need some further notation. Set $q_Q = \sum_{j\geq 1} \lambda_j e_j^{\otimes 2}$ and define, for $a \in \{0, \ldots, \lfloor r/2 \rfloor\}$, $p = r - 2a$, the operator $K_{r,a,Q} : \mathcal{H}_1^{\odot r} \to \mathcal{H}_1^{\odot p}$ by successive contraction with q_Q, i.e., $\langle K_{r,a,Q}[w], u\rangle_{\mathcal{H}_1^{\otimes p}} = \langle w, q_Q^{\otimes a} u\rangle_{\mathcal{H}_1^{\otimes r}}$, $w \in \mathcal{H}_1^{\odot r}$, $u \in \mathcal{H}_1^{\odot p}$ and with $K_{r,0,Q} = \mathrm{Id}$. Fix $1 \leq p \leq r$ with $r - p$ even, and set $a = (r-p)/2$ and $\gamma_{r,a} \doteq r!/(2^a a! p!)$. For $w \in \mathcal{H}_1^{\odot r}$, define the candidate kernel

$$f_{p,k}[w] \doteq \sum\nolimits_{\boldsymbol{j}\in\mathbb{N}^p} \Big\langle \gamma_{r,a} Q_p^{1/2}[K_{r,a,Q}[w]], e_{\boldsymbol{j}}\Big\rangle_{\mathcal{H}_1^{\otimes p}} \varepsilon_{\boldsymbol{j},k} \in \mathfrak{H}^{\odot p}, \tag{7}$$

where the dependence of $f_{p,k}$ on the fixed degree r is suppressed. Since $\{\varepsilon_{jk}\}_{j\geq 1}$ is orthonormal for fixed k, the series in (7) converges in $\mathfrak{H}^{\otimes p}$. Its kernel is symmetric because $Q_p^{1/2} K_{r,a,Q}[w]$ is symmetric. The contraction inequality and the definition of q_Q give

$$\|q_Q\|_{\mathcal{H}_1^{\otimes 2}}^2 = \sum\nolimits_{j\geq 1} \lambda_j^2 < \infty, \qquad \|K_{r,a,Q}\|_{\mathrm{op}} \leq \|q_Q\|_{\mathcal{H}_1^{\otimes 2}}^a, \qquad K_{r,a,Q}[h^{\otimes r}] = \langle Q[h], h\rangle_{\mathcal{H}_1}^a\, h^{\otimes (r-2a)}. \tag{8}$$

Since centering affects only the zeroth chaos, for every $p \geq 1$ and $i \geq 1$,

$$J_p \langle G_r[X_k], v_i\rangle_{\mathcal{H}_2} = J_p \Big\langle A_r[X_k^{\otimes r}], v_i\Big\rangle_{\mathcal{H}_2} = J_p \Big\langle X_k^{\otimes r}, w_i\Big\rangle_{\mathcal{H}_1^{\otimes r}}, \qquad w_i = A_r^*[v_i] \in \mathcal{H}_1^{\odot r}. \tag{9}$$

By the polarization identity, finite linear combinations of pure powers are dense in $\mathcal{H}_1^{\odot r}$. Thus, for each $n \geq 1$, we may choose $N_n \in \mathbb{N}$, coefficients $c_{\ell,n} \in \mathbb{R}$, and vectors $h_{\ell,n} \in \mathcal{H}_1$, $1 \leq \ell \leq N_n$, such that

$$w_{i,n} \doteq \sum\nolimits_{\ell=1}^{N_n} c_{\ell,n} h_{\ell,n}^{\otimes r}, \qquad \|w_{i,n} - w_i\|_{\mathcal{H}_1^{\otimes r}} \to 0. \tag{10}$$

For fixed r, p, i, k, and with further explanations given below, we obtain

$$J_p\Big[\langle G_r[X_k], v_i\rangle_{\mathcal{H}_2}\Big] - I_p\big[f_{p,k}[w_i]\big] = J_p\Big[\Big\langle X_k^{\otimes r}, w_i - w_{i,n}\Big\rangle_{\mathcal{H}_1^{\otimes r}}\Big] + J_p\Big[\Big\langle X_k^{\otimes r}, w_{i,n}\Big\rangle_{\mathcal{H}_1^{\otimes r}}\Big] - I_p\big[f_{p,k}[w_i]\big] \tag{11}$$

$$= J_p\Big[\Big\langle X_k^{\otimes r}, w_i - w_{i,n}\Big\rangle_{\mathcal{H}_1^{\otimes r}}\Big] + I_p\big[f_{p,k}[w_{i,n}] - f_{p,k}[w_i]\big] \xrightarrow[n\to\infty]{L^2(\Omega)} 0, \tag{12}$$

where (11) is due to (7) and (9). We get the equality in (12), since by linearity of J_p, I_p, and $f_{p,k}$,

$$J_p \Big\langle X_k^{\otimes r}, w_{i,n}\Big\rangle_{\mathcal{H}_1^{\otimes r}} = \sum\nolimits_{\ell=1}^{N_n} c_{\ell,n} J_p \Big\langle X_k^{\otimes r}, h_{\ell,n}^{\otimes r}\Big\rangle_{\mathcal{H}_1^{\otimes r}} = \sum\nolimits_{\ell=1}^{N_n} c_{\ell,n} I_p\big(f_{p,k}[h_{\ell,n}^{\otimes r}]\big) = I_p\big(f_{p,k}[w_{i,n}]\big), \tag{13}$$

where for the second equality in (13), introduce $\sigma_h^2 \doteq \langle Q[h], h\rangle_{\mathcal{H}_1} > 0$ for $h \in \mathcal{H}_1 \setminus \{0\}$, such that

$$J_p \Big\langle X_k^{\otimes r}, h^{\otimes r}\Big\rangle_{\mathcal{H}_1^{\otimes r}} = J_p \langle X_k, h\rangle_{\mathcal{H}_1}^r = \gamma_{r,a} \sigma_h^r H_p\Big(\langle X_k, h\rangle_{\mathcal{H}_1}/\sigma_h\Big) = I_p\Big(\gamma_{r,a}\sigma_h^{2a} \xi_{h,k}^{\otimes p}\Big) = I_p\big(f_{p,k}[h^{\otimes r}]\big), \tag{14}$$

where we use the Hermite expansion of $z \mapsto z^r$ and the relation between Hermite polynomials and multiple Wiener–Itô integrals [8, Proposition 1.4.2(v) and Theorem 2.7.7] characterized by $\xi_{h,k}^{\otimes p} = \sum_{\boldsymbol{j}\in\mathbb{N}^p} \Big\langle Q_p^{1/2}[h^{\otimes p}], e_{\boldsymbol{j}}\Big\rangle_{\mathcal{H}_1^{\otimes p}} \varepsilon_{\boldsymbol{j},k}$. For the last equality, we used (7) and (8). The identity (14) also holds for $h = 0$. For $p > r$ or $r - p$ odd, the corresponding chaos projection is zero. For the two summands in (12),

$$\left\| J_p \left\langle X_k^{\otimes r}, w_{i,n} - w_i \right\rangle_{\mathcal{H}_1^{\otimes r}} \right\|_{L^2(\Omega)} \leq \left(\mathrm{E} \left\| X_k \right\|_{\mathcal{H}_1}^{2r} \right)^{1/2} \left\| w_{i,n} - w_i \right\|_{\mathcal{H}_1^{\otimes r}} \to 0,$$

$$\left\| I_p \left[f_{p,k}[w_{i,n}] - f_{p,k}[w_i] \right] \right\|_{L^2(\Omega)} \leq \sqrt{p!}\, \gamma_{r,a} \| Q_p^{1/2} \|_{\mathrm{op}} \left\| q_Q \right\|_{\mathcal{H}_1^{\otimes 2}}^a \left\| w_{i,n} - w_i \right\|_{\mathcal{H}_1^{\otimes r}} \to 0.$$

The first bound uses that J_p is an orthogonal projection and Cauchy–Schwarz; the second uses (7), (8), the Wiener–Itô isometry, and orthonormality of the coordinate tensors. Both limits follow from (10).

Since the left-hand side of (12) does not depend on n, it is zero in $L^2(\Omega)$. Then, a.s.,

$$J_p \Big[\langle G_r[X_k], v_i \rangle_{\mathcal{H}_2} \Big] = I_p \big[f_{p,k}[w_i] \big] = I_p \left[\sum\nolimits_{j \in \mathbb{N}^p} \left\langle \gamma_{r,a} Q_p^{1/2} [K_{r,a,Q}[A_r^*[v_i]]], e_j \right\rangle_{\mathcal{H}_1^{\otimes p}} \varepsilon_{j,k} \right]. \tag{15}$$

Comparing (15) with (2), and using uniqueness of the symmetric Wiener–Itô kernel, gives

$$C_p^{(r)}[v_i] = \sum\nolimits_{j \in \mathbb{N}^p} \left\langle \gamma_{r,a} Q_p^{1/2} [K_{r,a,Q}[A_r^*[v_i]]], e_j \right\rangle_{\mathcal{H}_1^{\otimes p}} e_j = \gamma_{r,a} Q_p^{1/2} [K_{r,a,Q}[A_r^*[v_i]]].$$

Since this holds for every i, we have the following operator factorization. For $1 \leq p \leq r$ with $r - p$ even,

$$C_p^{(r)} = Q_p^{1/2} D_p^{(r)}, \qquad D_p^{(r)} \doteq \gamma_{r,a} K_{r,a,Q} A_r^*, \quad a = (r-p)/2,$$

and $C_p^{(r)} = D_p^{(r)} = 0$ for all other $p \geq 1$. Each $D_p^{(r)}$ is bounded by (8). By (6), the coefficient operators of G satisfy $C_p = \sum_{r=1}^m C_p^{(r)} = Q_p^{1/2} D_p$, $D_p \doteq \sum_{r=1}^m D_p^{(r)}$ and $D_p = 0$ for $p > m$. We get (4) by Cauchy–Schwarz and (8) since

$$\sum_{p=1}^{\infty} p! \| D_p \|_{\mathrm{op}}^2 \leq m \sum_{r=1}^{m} \sum_{p=1}^{r} p! \| D_p^{(r)} \|_{\mathrm{op}}^2 \leq m \sum_{r=1}^{m} \left\| A_r \right\|_{\mathrm{op}}^2 \sum_{1 \leq p \leq r, r-p \text{ even}} p!\, \gamma_{r,(r-p)/2}^2 \left\| q_Q \right\|_{\mathcal{H}_1^{\otimes 2}}^{r-p} < \infty.$$

□

3. Functional kurtosis operators

Let $\mathcal{H}_1 = \mathcal{H}$ be infinite-dimensional and $X \sim \mathcal{N}(0, Q)$. The Gaussian-centered functional Fourth-Order Blind Identification (FOBI) map, with values in $\mathcal{H}_2 = \mathcal{S}_2(\mathcal{H}) \cong \mathcal{H} \otimes \mathcal{H}$, is given by the map

$$G_K[x] = \|x\|_{\mathcal{H}}^2 (x \otimes x) - \left((\operatorname{Tr} Q) Q + 2Q^2 \right). \tag{16}$$

Note that G_K is centered and $G_K[X] \in L^2(\Omega : \mathcal{H}_2)$ since $\left\| \, \|X\|_{\mathcal{H}}^2 (X \otimes X) \right\|_{\mathcal{S}_2(\mathcal{H})} = \|X\|_{\mathcal{H}}^4$. Proposition 2.1 does not apply to (16), since the factor $\|x\|_{\mathcal{H}}^2 = \operatorname{Tr}(x \otimes x)$ entails a trace contraction, which is unbounded with respect to the Hilbert tensor norm in infinite dimensions. For that reason, a smoothed version of the FOBI operator was proposed in [10]. We impose an additional Schatten condition. Let $S \in \mathcal{S}_4(\mathcal{H})$ and

$$G : \mathcal{H} \to \mathcal{S}_2(\mathcal{H}), \quad G[x] = \|S[x]\|_{\mathcal{H}}^2 (S[x] \otimes S[x]) - \mathrm{E} \left[\|S[X]\|_{\mathcal{H}}^2 (S[X] \otimes S[X]) \right]. \tag{17}$$

Lemma 3.1. *The operator G defined in* (17) *satisfies* (4). *Moreover, if $S \neq 0$, then* $\operatorname{rank}(G) = 2$.

Proof. Since $S \in \mathcal{S}_4$, we have that $S^* S \in \mathcal{S}_2$. Define the operator $A_{4,S}$ such that

$$A_{4,S}[u_1 \otimes u_2 \otimes u_3 \otimes u_4] = \langle S[u_1], S[u_2] \rangle_{\mathcal{H}} (S[u_3] \otimes S[u_4]) \quad \text{with } \left\| A_{4,S} \right\|_{\mathrm{op}} \leq \left\| S^* S \right\|_{\mathcal{S}_2(\mathcal{H})} \|S\|_{\mathrm{op}}^2 < \infty.$$

Under $\mathcal{S}_2(\mathcal{H}) \cong \mathcal{H} \otimes \mathcal{H}$, the first factor is the bounded functional represented by $S^* S \in \mathcal{S}_2$ and the second is $S \otimes S$. Applying Proposition 2.1 to $A_{4,S}\, \mathrm{sym}_4$ proves both coefficient conditions.

For the rank, we first see that $J_0[G] = 0$ since we have centered, and $J_1[G] = 0$ since $A_{4,s}$ is of fourth order, so odd chaoses vanish. For the second chaos, denote $Y = S[X]$, $K = SQS^*$, and choose u with $\sigma^2 = \langle K[u], u \rangle_{\mathcal{H}} > 0$. If $Z = \langle Y, u \rangle_{\mathcal{H}} / \sigma$, then $Y = aZ + W$, where $a = K[u]/\sigma$ and W is independent of Z. Hence

$$\mathrm{E} \left[\langle A_{4,S}[X^{\otimes 4}][u], u \rangle_{\mathcal{H}} H_2(Z) \right] = \mathrm{E} \left[\|Y\|_{\mathcal{H}}^2 \langle Y, u \rangle_{\mathcal{H}}^2 H_2(Z) \right] = \sigma^2 \left(12 \|a\|_{\mathcal{H}}^2 + 2 \, \mathrm{E} \|W\|_{\mathcal{H}}^2 \right) > 0.$$

Thus a scalar coordinate has a nonzero second chaos, so $\operatorname{rank}(G) = 2$. □

4. Joint sample autocovariance operators

Let $\{X_k\}_{k\in\mathbb{Z}}$ be centered stationary Gaussian in $\mathcal{H}$ and fix $L \in \mathbb{N}_0$. Define the ACO and its empirical counterpart

$$\Gamma_X(h) = \mathrm{E}(X_h \otimes X_0), \quad \mathbf{\Gamma}_L = (\Gamma_X(h))_{h=0}^L, \quad \widehat{\Gamma}_n(h) = n^{-1}\sum_{k=1}^n X_{k+h} \otimes X_k, \quad \widehat{\mathbf{\Gamma}}_{n,L} = (\widehat{\Gamma}_n(h))_{h=0}^L.$$

Gaussian fluctuations for the sample ACOs in Hilbert spaces were studied in [2, 6].

For our setting, we define $Y_k = (X_k, \dots, X_{k+L}) \in \mathcal{H}^{L+1}$, and write $\widetilde{Q} = \mathrm{Cov}(Y_0)$ and $\widetilde{Q}_2^{1/2} = (\widetilde{Q}^{1/2})^{\otimes 2}$. Then, we take $\mathcal{H}_1 = \mathcal{H}^{L+1}$, $\mathcal{H}_2 = \mathcal{S}_2(\mathcal{H})^{L+1}$. Let $P_h : \mathcal{H}_1 \to \mathcal{H}$ be the hth coordinate projection and define $A \in \mathcal{L}(\mathcal{H}_1^{\otimes 2}, \mathcal{H}_2)$ by $[A[w]]_h = (P_h \otimes P_0)[w]$ for $0 \le h \le L$. Then, set

$$G : \mathcal{H}_1 \to \mathcal{H}_2, \qquad G[Y_k] = (X_{k+h} \otimes X_k - \Gamma_X(h))_{h=0}^L = A\left[Y_k^{\otimes 2} - \mathrm{E}[Y_k^{\otimes 2}]\right]. \tag{18}$$

Lemma 4.1. *The operator G defined in* (18) *satisfies* (4) *and, moreover* $\mathrm{rank}(G) = 2$.

Proof. Since $\|P_h \otimes P_0\|_{\mathrm{op}} \le 1$, the map A is bounded, with $\|A\|_{\mathrm{op}}^2 \le L+1$, and its definition gives (18). Proposition 2.1, applied with $A_2 = A$, shows that the centered quadratic polynomial G has only a second-chaos component and satisfies the coefficient assumptions. This component is nonzero when $Q \neq 0$, since for some $a \in \mathcal{H}$, $\mathrm{Var}\left(\langle X_k, a\rangle_{\mathcal{H}}^2 - \langle Q[a], a\rangle_{\mathcal{H}}\right) = 2\,\langle Q[a], a\rangle_{\mathcal{H}}^2 > 0$. Thus G has Hermite rank two. □

Remark 4.2. Here $Y_k = (X_k, \dots, X_{k+L}) \in \mathcal{H}^{L+1}$ satisfies (3), since $\sum_v \|\Gamma_Y(v)\|_{\mathcal{S}_1}^q \le (L+1)^{2q} \sum_v \|\Gamma_X(v)\|_{\mathcal{S}_1}^q$, $q \ge 1$.

5. Continuum random-feature maps

Let $(\mathcal{T}, \mathcal{A}, \nu)$ be a probability space, $\mathcal{H}_1$ a general, separable Hilbert space, and set $\mathcal{H}_2 = L^2(\mathcal{T}, \nu)$. Let $t \mapsto h_t \in \mathcal{H}_1$ be strongly measurable, let $t \mapsto b_t \in \mathbb{R}$ be measurable, and set $\sigma_t^2 = \langle Q[h_t], h_t\rangle_{\mathcal{H}_1} > 0$ for ν-almost every $t \in \mathcal{T}$.

Fix a bounded continuous function $\phi : \mathbb{R} \to \mathbb{R}$ and, for $Z \sim \mathcal{N}(0,1)$, set $m(t) \doteq \mathrm{E}[\phi(\sigma_t Z + b_t)]$, $\phi_t(z) \doteq \phi(\sigma_t z + b_t) - m(t)$. Suppose that $\int_{\mathcal{T}} \mathrm{E}\,|\phi_t(Z)|^2 \nu(dt) < \infty$. For ν-almost every t, write

$$\phi_t(z) = \sum\nolimits_{p=1}^\infty a_p(t) H_p(z), \quad a_p(t) = \frac{1}{p!}\mathrm{E}[\phi_t(Z) H_p(Z)], \quad q = \min\left\{p \ge 1 : \int_{\mathcal{T}} |a_p(t)|^2 \nu(dt) > 0\right\}. \tag{19}$$

Define the centered random-feature map $G : \mathcal{H}_1 \to \mathcal{H}_2$ and the empirical mean by

$$G[x](t) = \phi(\langle x, h_t\rangle_{\mathcal{H}_1} + b_t) - m(t), \quad \widehat{m}_n(t) = \frac{1}{n}\sum\nolimits_{k=1}^n \phi(\langle X_k, h_t\rangle_{\mathcal{H}_1} + b_t) \tag{20}$$

We seek a simultaneous $L^2(\mathcal{T},\nu)$-CLT for a continuum of dependent random features. The model includes the centered random Fourier kernel estimator of [9] with $\mathcal{T} = M \times M$, $M \subset \mathbb{R}^d$ compact, $h_{(s,t)} = s - t$, and $\phi = \cos$. For $t = (v, s)$ and $h_{v,s} = \int_D \eta_{s,y} v(y)\, dy$, (20) is the centered output of a one-layer neural operator with linearly parameterized kernel $\kappa_x(s,y) = \left\langle x, \eta_{s,y}\right\rangle_{\mathcal{H}_1}$; see the integral-operator architecture of [5]. Finite $\mathcal{T}$ gives finite features.

Lemma 5.1. *Assume that $q < \infty$ and*

$$\sum_{p=q}^\infty p! \int_{\mathcal{T}} |a_p(t)|^2 \left(\|h_t\|_{\mathcal{H}_1}^2 / \sigma_t^2\right)^p \nu(dt) < \infty. \tag{21}$$

Then G has Hermite rank q and satisfies (4).

Proof. Set $u_t = Q^{1/2}[h_t]/\sigma_t$ and $\eta_{t,k} = \sum_j \left\langle u_t, e_j\right\rangle_{\mathcal{H}_1} \varepsilon_{jk}$. Then $\|u_t\|_{\mathcal{H}_1} = \|\eta_{t,k}\|_{\mathfrak{H}} = 1$ and $W_{u_t}[X_k] = \langle X_k, h_t\rangle_{\mathcal{H}_1}/\sigma_t$. Using Fubini's theorem, the Wiener-Ito isometry and (19) give, in $L^2(\Omega : \mathcal{H}_2)$,

$$G[X_k](t) = \sum\nolimits_{p=q}^\infty a_p(t) I_p[\eta_{t,k}^{\otimes p}], \qquad \mathrm{E}\left\|a_p(\cdot) I_p[\eta_{\cdot,k}^{\otimes p}]\right\|_{\mathcal{H}_2}^2 = p! \int_{\mathcal{T}} |a_p(t)|^2 \nu(dt).$$

Thus $\operatorname{rank}(G) = q$. Comparing kernels with [4, Eqs. (3.2), (4.8)] gives, for $v \in \mathcal{H}_2$,

$$C_p[v] = \int_{\mathcal{T}} v(t) a_p(t) \left(\frac{Q^{1/2}[h_t]}{\sigma_t} \right)^{\otimes p} \nu(dt) = Q_p^{1/2}[D_p[v]], \quad \text{where } D_p[v] = \int_{\mathcal{T}} v(t) a_p(t) \left(\frac{h_t}{\sigma_t} \right)^{\otimes p} \nu(dt).$$

Moreover, (4) follows, since (21) implies

$$\sum_{p=q}^{\infty} p! \left\| Q_p^{-1/2} C_p \right\|_{\mathrm{op}}^2 = \sum_{p=q}^{\infty} p! \left\| D_p \right\|_{\mathrm{op}}^2 \leq \sum_{p=q}^{\infty} p! \left\| D_p \right\|_{\mathcal{S}_2(\mathcal{H}_2, \mathcal{H}_1^{\otimes p})}^2 = \sum_{p=q}^{\infty} p! \int_{\mathcal{T}} |a_p(t)|^2 \left(\|h_t\|_{\mathcal{H}_1}^2 / \sigma_t^2 \right)^p \nu(dt) < \infty.$$

□

6. Kernel autocovariance operators

Kernel ACOs are proposed as a non-linear alternative to the usual ACOs; see [7]. The setting is as follows. We choose the polynomial kernel of degree m, and define the feature map $\Phi_m : \mathcal{H} \to \mathcal{H}^{\otimes m}$, $\Phi_m[x] = x^{\otimes m}$ with associated kernel $k_m[x, z] = \langle \Phi_m[x], \Phi_m[z] \rangle_{\mathcal{H}^{\otimes m}} = \langle x, z \rangle_{\mathcal{H}}^m$. In this case $\mu_m = \mathrm{E}[X_0^{\otimes m}] \in \mathcal{H}^{\otimes m}$, and the kernel ACO and its empirical counterpart are given by

$$C_{\Phi_m, \mu_m}(h) = \mathrm{E}[(\Phi_m[X_h] - \mu_m) \otimes (\Phi_m[X_0] - \mu_m)] \in \mathcal{S}_2(\mathcal{H}^{\otimes m}), \quad \widehat{C}_{\Phi_m, n}(h) = \frac{1}{n} \sum_{k=1}^{n} (\Phi_m[X_{k+h}] - \mu_m) \otimes (\Phi_m[X_k] - \mu_m).$$

This gives rise to $\mathcal{H}_1 = \mathcal{H} \oplus \mathcal{H}$, $\mathcal{H}_2 = \mathcal{S}_2(\mathcal{H}^{\otimes m})$, and $G : \mathcal{H}_1 \to \mathcal{H}_2$, with

$$G[y_1, y_0] = (y_1^{\otimes m} - \mu_m) \otimes (y_0^{\otimes m} - \mu_m) - C_{\Phi_m, \mu_m}(h). \tag{22}$$

Our CLT for Gaussian inputs covers unbounded polynomial kernels, unlike [7, Theorem 17].

Lemma 6.1. *The map G given in* (22) *satisfies* (4).

Proof. Let $P_1, P_0 : \mathcal{H} \oplus \mathcal{H} \to \mathcal{H}$ be the coordinate projections as in Section 4. For $z = (y_1, y_0) \in \mathcal{H} \oplus \mathcal{H}$,

$$\begin{aligned} G[z] &= (P_1[z])^{\otimes m} \otimes (P_0[z])^{\otimes m} - (P_1[z])^{\otimes m} \otimes \mu_m - \mu_m \otimes (P_0[z])^{\otimes m} + \mu_m \otimes \mu_m - C_{\Phi_m, \mu_m}(h) \\ &= (\mu_m \otimes \mu_m - C_{\Phi_m, \mu_m}(h)) - \left(\left(P_1^{\otimes m}[z^{\otimes m}] \right) \otimes \mu_m + \mu_m \otimes \left(P_0^{\otimes m}[z^{\otimes m}] \right) \right) + \left(P_1^{\otimes m} \otimes P_0^{\otimes m} \right)[z^{\otimes 2m}], \end{aligned}$$

which is the form (5) with only the terms of order $0, m, 2m$ present. Thus the conclusion follows by Proposition 2.1. □

Declaration of competing interest. The authors declare no competing financial interests or personal relationships that could have influenced this work.

Data availability. No data were used in this research.

Declaration of generative AI use. The authors used ChatGPT Astra 6 (OpenAI) to assist with identifying the applications of Sections 3 and 5, revising text, standardizing notation, and reviewing mathematical arguments and proofs. The authors take full responsibility for the correctness of the arguments of the manuscript.

References

[1] Breuer, P. and Major, P. Central limit theorems for non-linear functionals of Gaussian fields. *J. Multivar. Anal.*, 13(3):425–441, 1983.

[2] Düker, M.-C. and Zoubouloglou, P. Scaling limits for sample autocovariance operators of Hilbert space-valued linear processes. *Preprint, [arXiv]*, 2025.

[3] Düker, M.-C. and Zoubouloglou, P. The fourth-moment theorem on Hilbert spaces. *Preprint, [arXiv]*, 2025.

[4] Düker, M.-C. and Zoubouloglou, P. Breuer–Major theorems for Hilbert space-valued random variables. *Preprint, [arXiv] (v2)*, 2026.

[5] Kovachki, N., Li, Z., Liu, B., Azizzadenesheli, K., Bhattacharya, K., Stuart, A., and Anandkumar, A. Neural operator: Learning maps between function spaces with applications to PDEs. *J. Mach. Learn. Res.*, 24(89):1–97, 2023.

[6] Mas, A. Weak convergence for the covariance operators of a hilbertian linear process. *Stoch. Process. Their Appl.*, 99(1):117–135, 2002.

[7] Mollenhauer, M., Klus, S., Schütte, C., and Koltai, P. Kernel autocovariance operators of stationary processes: Estimation and convergence. *J. Mach. Learn. Res.*, 23(327):1–34, 2022.

[8] Nourdin, I. and Peccati, G. *Normal Approximations with Malliavin Calculus: From Stein's Method to Universality*, volume 192 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 2012.

[9] Rahimi, A. and Recht, B. Random features for large-scale kernel machines. In *Advances in Neural Information Processing Systems 20*, pages 1177–1184, 2007.

[10] Vidal, M., Leman, M., and Aguilera, A. M. Functional independent component analysis by choice of norm: A framework for near-perfect classification. *Adv. Data Anal. Classif.*, 20:521–549, 2026.