Received: 12 December 2022 | Revised: 20 March 2023 | Accepted: 5 April 2023

DOI: 10.1002/mma.9301

RESEARCH ARTICLE

WILEY

Adaptive Fourier decomposition-type sparse representations versus the Karhunen–Loève expansion for decomposing stochastic processes

Tao Qian[1] | **Ying Zhang**[2] | **Wanquan Liu**[3] | **Wei Qu**[4]

[1]Macau Center for Mathematical Sciences, Macau University of Science and Technology, Macau

[2]Department of Engineering Science, Faculty of Innovation Engineering, Macau University of Science and Technology, Macau

[3]School of Intelligent Systems Engineering, Sun Yat-sen University, Shenzhen, China

[4]College of Sciences, China Jiliang University, Hangzhou, China

Correspondence
Wei Qu, College of Sciences, China Jiliang University, Xueyuan Street No. 258, Qiantang District, Hangzhou, Zhejiang 310018, China.
Email: quwei2math@qq.com

Communicated by: F. Colombo

Funding information
Zhejiang Provincial Natural Science Foundation of China, Grant/Award Number: LQ23A010014

The present study devotes to introducing and analyzing the use of the adaptive Fourier decomposition (AFD)-type methods in the area of stochastic processes and random fields. We involve two types of algorithms, namely, stochastic AFD (SAFD) and stochastic pre-orthogonal AFD (SPOAFD), for, respectively, the Hardy space format and non-Hardy space ones, as may be regarded. We provide both their theoretical results and practical algorithms and compare them with the well adopted and, in fact, dominating Karhunen–Loève (KL)-type expansions. The AFD methods involve a finite or infinite sequence of optimally chosen parameters; they, in contrast, do not rely on and hence not have to compute the eigenpairs of the second type Fredholm integral equation with the covariance function as the kernel. Apart from such computational conveniences, they with the same convergence rate have flexibility of choosing best suitable dictionaries for doing the specific task in the practice. We include a number of experiments showing that in terms of effectiveness, the AFD methods give better approximations before all the positive eigenvalues running out in the case the integral operator being of a finite rank or before the KL iteration step becomes excessively large: The AFD methods normally give better approximations from the very beginning of the iterations.

KEYWORDS

reproducing kernel Hilbert space, sparse representation by dictionary elements, stochastic adaptive Fourier decomposition, stochastic process

MSC CLASSIFICATION

62L20, 60G99, 30B99, 60J65

1 | INTRODUCTION

For the self-containing purpose, this section will introduce adaptive Fourier decomposition (AFD)-type sparse representations with emphasis on the stochastic AFDs (SAFDs) [1]. We are based on a dictionary $\mathcal{D}$ of a complex Hilbert space $\mathcal{H}$. By definition, a dictionary of $\mathcal{H}$ consists of a class of unimodular elements whose linear span is dense in $\mathcal{H}$. The formulation we adopt is that $\mathcal{H}$ is the L^2-space of complex-valued functions on a manifold, $\partial\mathbf{D}$ is the boundary of $\mathbf{D}$, where $\mathbf{D}$ itself is an open and connected domain, called a region, in an Euclidean space. A process in a finite time interval may be well formulated as defined on the unit circle $\partial \mathbf{B}_1 = \{e^{it} \,|\, t \in [0, 2\pi)\}$, where in the defined notation $\mathbf{D} = \mathbf{B}_1$, the unit disc. If a process is defined in the entire time range, we use the format that $\mathbf{D} = \mathbf{C}^+$, the upper-half plane, and $\partial \mathbf{D} = \partial \mathbf{C}^+ \triangleq \mathbb{R}^1$ the whole real line. For random functions with the space variable on the n-sphere S_n, we may fit the problem to the format where $\mathbf{D}$ being the $(n+1)$-dimensional solid ball $\mathbf{B}_{n+1}, S_n = \partial \mathbf{B}_{n+1}$ in $\mathbb{R}^{n+1}$. For random functions defined on the Euclidean space $\mathbb{R}^n$, we may take the parameter set $\mathbf{D} = \mathbb{R}^{n+1}_+$, where $\partial \mathbf{D} = \partial \mathbb{R}^{n+1}_+ = \mathbb{R}^n$. The concerned contexts may be classified into the following two structures: the coarse and the fine structure, as described below.

(i) The coarse structure: $\mathcal{H} = L^2(\partial \mathbf{D})$ and the elements of the dictionary $\mathcal{D}$ are indexed by all q in $\mathbf{D}$. Examples of such model can include $\mathcal{D}$ being the collection of the Poisson kernels in the unit disc or the unit ball or those in the upper-half space. In the upper-half space, case $\mathcal{D}$ can be collections of the heat kernels or various kinds of dilated and translated convolution kernels [2].

(ii) The fine structure: Certain functions defined on a region $\mathbf{D}$ may constitute a reproducing kernel Hilbert space (RKHS), being denoted as H_K, or $H^2(\mathbf{D})$, and called the Hardy space of the context, where $K : \mathbf{D} \times \mathbf{D} \to \mathbf{C}$ is the reproducing kernel, satisfying $K_q(p) = K(p, q)$ for any pair $p, q \in \mathbf{D}$. The related theory and examples may be found in previous studies [2–4], as well as in Yang [5]. The relevant literature address various types of RKHSs in such setting, including, for instance, the $H^2(\mathbf{B}_1)$ space of complex holomorphic functions in the unit disc $\mathbf{B}_1$, and the $h^2\left(\mathbb{R}^{n+1}_+\right)$ space of harmonic functions in the upper-half Euclidean space, where, precisely,

$$H^2(\mathbf{B}_1) = \left\{ f : \mathbf{B}_1 \to \mathbf{C} \,|\, f(z) = \sum_{k=0}^{\infty} c_k z^k, \|f\|^2 = \sum_{k=0}^{\infty} |c_k|^2 < \infty \right\}$$

and

$$h^2\left(\mathbb{R}^{n+1}_+\right) = \left\{ u : \mathbb{R}^{n+1}_+ \to \mathbf{C} \,|\, \Delta u = 0 \text{ on } \mathbb{R}^{n+1}_+, \sup_{(t,y)\in \Gamma^\alpha_x} |u(t, y)| \in L^2(\mathbb{R}^n) \right\},$$

where Γ^α_x is the orthogonal α-cone in $\mathbb{R}^{1+n}_+$ with its tip at $x \in \mathbb{R}^n$. Similarly, there exist the harmonic h^2-space of the unit n-ball in $\mathbb{R}^{1+n}$ and the heat kernel Hardy space $H^2_{heat}(\mathbb{R}^{1+n}_+)$. In the present paper, all such RKHSs are denoted by $H^2(\mathbf{D})$.

As a property of reproducing kernel, in each of the fine cases, the parameterized reproducing kernels $K_q, q \in \mathbf{D}$, form a dictionary. To simplify the terminology, we also call a dictionary element as a kernel. We now briefly review a few of sparse representation models belonging to the AFD type. Unless otherwise specified, the norm $\|\cdot\|$ and the inner product notation $\langle \cdot, \cdot \rangle$ will always refer to those of the underlying complex Hilbert space $\mathcal{H} = L^2(\partial \mathcal{D})$.

(a) AFD, or Core AFD: In the $\mathcal{H} = H^2(\mathbf{B}_1)$ context, the inner product is defined through the one on the boundary $L^2(\partial \mathbf{B}_1)$ owing to existence, as fundamental property, of boundary limits of the Hardy space functions, namely,

$$\langle f, g \rangle_{H^2(\mathbf{B}_1)} = \frac{1}{2\pi} \int_0^{2\pi} f\left(e^{it}\right) \bar{g}\left(e^{it}\right) dt.$$

We work with the Szegö kernel (or the reproducing kernel) dictionary

$$\mathcal{D}(\mathbf{B}_1) \triangleq \{k_a(z)\}_{a\in \mathbf{B}_1} \triangleq \left\{ \frac{1}{1 - \bar{a}z} \right\}_{a\in \mathbf{B}_1}.$$

For any given $f \in H^2(\mathbf{B}_1)$, we have a greedy selections of the parameters

$$a_k = \arg\max \{ |\langle f_k, e_a \rangle| \,|\, a \in \mathbf{B}_1 \}, \tag{1}$$

where with $f_1 = f$, the f_k are the *reduced remainders*, obtained through inductively using the *generalized backward shift*:

$$f_k(z) = \frac{f_{k-1}(z) - \langle f_{k-1}, e_{a_{k-1}} \rangle e_{a_{k-1}}(z)}{\frac{z - a_{k-1}}{1 - \bar{a}_{k-1} z}}, \tag{2}$$

where $e_a = k_a/\|k_a\|, \|k_a\| = 1/\sqrt{1-|a|^2}$, is the normalized Szegö kernel, and the validity of the maximal selection principle (1) (MSP) is a consequence of the boundary vanishing condition (BVC)

$$\lim_{|a|\to 1} \langle f_k, e_a \rangle = 0.$$

Then, there follows (see Qian and Wang [6])

$$f(z) = \sum_{k=1}^{\infty} \langle f_k, e_{a_k} \rangle B_k(z),$$

where $\{B_k\}$ is automatically an orthonormal system, called the Takenaka–Malmquist (TM) system, given by

$$B_k(z) = e_{a_k}(z) \prod_{l=1}^{k-1} \frac{z - a_l}{1 - \bar{a}_l z}. \tag{3}$$

Whether the associated TM system is a basis depends on whether $\sum_{k=1}^{\infty}(1 - |a_k|) = \infty$. In both the basis and the non-basis cases, the above infinite series converges in the $L^2(\partial \mathbf{B}_1)$ norm sense on the boundary with the convergence rate

$$\left\| f - \sum_{k=1}^{n} \langle f_k, e_{a_k} \rangle B_k \right\| \le \frac{M}{\sqrt{n}}, \tag{4}$$

where

$$M = \inf \left\{ \sum_{k=1}^{\infty} |c_k| \mid f(z) = \sum_{k=1}^{\infty} c_k e_{b_k}, b_k \in \mathbf{B}_1 \right\}$$

(see Qian and Wang [6] or Temlyakov [7]). Being compared with the usual greedy algorithm, Core AFD was originated by the purpose of finding meaningful positive frequency decomposition of a signal. It was found that analytic phase derivatives of inner functions provide the source of such meaningful positive frequencies [8, 9]. Apart from this background, AFD addresses attainability of globally maximal energy matching pursuit. The attainability in various contexts of kernel approximation further motivated the multiple kernel concept (see below). For Core AFD, the reader is referred to Qian and Wang [6] and Qian [10]. For general greedy algorithms and convergence rate results, the reader is referred to DeVore and Temlyakov [11] and Temlyakov [7]. Core AFD has the following variations and generalizations.

(b) Unwinding AFD: The inductive steps in (a) give rise to the relation

$$f(z) = \sum_{k=1}^{n-1} \langle f_k, e_{a_k} \rangle B_k + g_n(z),$$

where

$$g_n(z) = f_n(z) \prod_{l=1}^{n-1} \frac{z - a_l}{1 - \bar{a}_l z}$$

is the *standard orthogonal remainder of degree n*. If, before doing the optimal energy matching pursuit for f_n, one first factorizes out the inner function factor of g_n, and next perform a maximal parameter selection, one obtains what is called *maximal unwinding AFD* [12], which converges considerably faster than Core AFD and, again, with positive frequencies. Coifman et al. studied a direct unwinding method named as Blaschke unwinding expansion in 2000 [13–15].

(c) Based on generalizations of Blaschke products, the AFD theory is extended to several complex variables contexts and to matrix-valued functions (see Alpay et al. [16, 17]).

(d) The generalized backward shift defined by (2) is crucial in Core AFD. In a general Hilbert space context, there usually do not exist a Blaschke product theory, and the Gram–Schmidt (GS) orthogonalization of parameterized kernels do not have explicit formulas as those for the TM system (3). It is the relations

$$\langle f_n, e_{a_n}\rangle = \langle g_n, B_n\rangle = \langle f, B_n\rangle \tag{5}$$

in the Core AFD context that gave inspiration to generalize the latter algorithm to the context of general Hilbert space with a dictionary, called *pre-orthogonal AFD* (*POAFD*) [1, 10, 18]. Precisely, the above relations hint the notion B_n^a within the n-orthonormal system $(B_1, \cdots, B_{n-1}, B_n^a)$, where the latter being the GS orthonormalization of the non-orthogonal system $(B_1, \cdots, B_{n-1}, k_a)$, where $a \in \mathbf{D}$ is to be determined.

POAFD is formulated as follows. Changing to the general Hilbert space $\mathcal{H}$ setting, below, we will use the notations $\mathbf{D}, q, K_q, E_q, E_n$ in place of, for the Core AFD case, $\mathbf{B}_1, a \in \mathbf{B}_1, k_a$, the Szegö or reproducing kernel, e_a, the normalized reproducing kernel or dictionary element, and B_n, the member of TM system. A class of dictionaries with good boundary behavior is first specified: A dictionary $\mathcal{D}$ of the Hilbert space $\mathcal{H}$ is said to satisfy BVC (or called a *BVC dictionary*), if for any $f \in \mathcal{H}$,

$$\lim_{q\to\partial\mathbf{D}} |\langle f, E_q\rangle| = 0.$$

The Szegö kernel enjoys the BVC property. Apart from some exceptional cases (see Qu and Dang [19]), most commonly used dictionaries are of BVC. Under the BVC assumption (or one has to first verify), there exists

$$q_n = \arg\sup\left\{\left|\left\langle f, \overset{q}{\underset{n}{\mathrm{E}}}\right\rangle\right| \mid q \in \mathbf{D}\right\}, \tag{6}$$

where $\{E_1, \cdots, E_{n-1}, E_n^q\}$ is the GS orthogonalization of $\{E_1, \cdots, E_{n-1}, \tilde{K}_q\}$, where for each $j : 1 \le j \le n$,

$$\tilde{K}_{q_j} = \left[\left(\frac{\partial}{\partial\bar{q}}\right)^{(l(j)-1)} K_q\right]_{q=q_j}, \ j = 1, 2, \cdots, n,$$

where $\frac{\partial}{\partial\bar{q}}$ is a directional derivative with respect to $\bar{q}$ and $l(j)$ is the multiple of q_j in the j-tuple $(q_1, \cdots, q_j)$, $1 \le j \le n$. $\tilde{K}_{q_j}$, sometimes denoted as $\tilde{K}_j$, is called the *multiple kernel* with respect to q_j in the j-tuple $(q_1, \cdots, q_j)$. As a particular case of the notation, we have $E_q = E_1^q$, as used in (8). In order to guarantee attainability of the supreme value at each of the parameter selections, one must allow repeating selections of the parameters that induces the multiple kernel notion [20, 21]. Recall that the basic functions in a TM system correspond to the GS orthogonalizations of the involved multiple Szegö kernels [1]. Due to the relations (5), POAFD in the classical Hardy space is identical with AFD. This shows that POAFD for a Hilbert space with a BVC dictionary has the same power as AFD in the Hardy space, while the former does not have the fine structure such as TM system in relation to Blaschke product. We remark that the Hardy space seems to be the only case in which the GS process on reproducing kernels gives rise to nice and practical formulas as in the TM system ([1, 16, 17]).

To perform GS orthogonalization in POAFD is to compute

$$E_n = E_n^{q_n} = \frac{\tilde{K}_{q_n} - \sum_{k=1}^{n-1}\langle\tilde{K}_{q_n}, E_k\rangle E_k}{\sqrt{\|\tilde{K}_{q_n}\|^2 - \sum_{k=1}^{n-1}|\langle\tilde{K}_{q_n}, E_k\rangle|^2}}. \tag{7}$$

With such formulation, Core AFD is extended to contexts of great variety in which a practical Blaschke product theory may not be known or may not exist. Significant generalizations include POAFD for product dictionary [10], POAFD for quaternionic space [22], POAFD for multivariate real variables in the Clifford algebra setting [23], POAFD for weighted Bergman and weighted Hardy spaces [19, 24], and most recently sparse representations for the Dirac δ function [2]. We note that the MSP of POAFD is the greediest one-step parameter selection and, in particular, more greedy than what is called *orthogonal greedy algorithm* [7] due to the relation: For $q \neq q_1, \cdots, q_{n-1}$,

$$\left|\left\langle g_n, E_n^q \right\rangle\right| = \frac{1}{\sqrt{1 - \sum_{k=1}^{n-1} |\langle E_q, E_k \rangle|^2}} \left|\left\langle g_n, E_1^q \right\rangle\right| \geq \left|\left\langle g_n, E_q \right\rangle\right|, \tag{8}$$

the ending term being recognized to be the objective function for the orthogonal greedy algorithm.

(e) n-Best AFD: A manipulation of a single optimal parameter selection with Core AFD or POAFD in (1) or (6), respectively, is n-best AFD, also called *n-best kernel approximation*, formulated as finding $(q_1, \cdots, q_n)$ such that

$$\left\| P_{\mathrm{span}\{\tilde{K}_{q_1}, \cdots, \tilde{K}_{q_n}\}}(f) \right\| = \sup \left\{ \left\| P_{\mathrm{span}\{\tilde{K}_{p_1}, \cdots, \tilde{K}_{p_n}\}}(f) \right\| \,|\, p_1, \cdots, p_n \in \mathbf{D} \right\}, \tag{9}$$

where we use $P_X(f)$ for the projection of f into the linear subspace X. In the classical Hardy space case $\mathbf{D} = \mathbf{B}_1$, the problem is equivalent with one of finding best approximation by rational functions of degree not exceeding n. Existence of a solution of such n-best rational approximation problem has long been solved (see Wang and Qian [25] and the references therein), a practical and mathematical rigorous algorithm, in contrast, has been left open until now [26]. To the authors' knowledge, none of the published algorithms can prevent from falling into local minimal distance. In general, for a Hilbert space with a BVC dictionary asserting existence of a solution for the n-best problem is by no means easy. The existence result for the classical complex Hardy space case has a number of proofs and recently been re-proved by using the maximal module principle of complex analytic functions as a new approach [25]. This progress allows to generalize existence of an n-best solution to weighted Bergman spaces and further to a wide class of RKHSs [27] for analytic functions in the unit disc. In the upper-half of the complex plane, there is a parallel theory. Reproofs of existence of the n-best approximation are also for the algorithm purpose whose progress will be reported in a separate paper.

(f) The most up-to-date developments of AFD is SAFD and stochastic pre-orthogonal AFD (SPOAFD) [1]. The former is precisely in the classical complex Hardy space context which, as mentioned, with the convenience of the TM system, and the latter is for general Hilbert spaces with a BVC dictionary. The purpose of the present paper is to introduce SAFD and SPOAFD to the study and practice with stochastic processes and random fields.

Definition 1. ([1, 28]). Suppose $\mathbf{D} \subset \mathbb{R}^n$ and $(\Omega, d\mathbb{P})$ is a probability space. The Bochner space $L^2(\Omega, L^2(\mathbf{D}))$ is the Hilbert space consisting of all the $L^2(\mathbf{D})$-valued random functions f satisfying

$$\begin{aligned} \|f\|^2_{L^2(\Omega, L^2(\mathbf{D}))} &\triangleq \int_\Omega \int_{\mathbf{D}} |f(x,\omega)|^2 dx d\mathbb{P} \\ &= E_\omega \|f_\omega\|^2_{L^2(\mathbf{D})} < \infty, \end{aligned} \tag{10}$$

where $f_\omega(x) = f(x,\omega)$.

For brevity, we also write $\mathcal{N} = L^2(\Omega, L^2(\mathbf{D}))$. The theory developed in Qian [1] is for the same space in the unit disc $\mathbf{B}_1$ but in terms of the Fourier expansion with random coefficients, being equivalent with the above defined due to the Plancherel Theorem. Besides SAFD, Qian [1] also develops SPOAFD for general stochastic Hilbert spaces with a BVC dictionary.

SAFD (identical with SAFD2 in the terminology of Qian [1]), concerning the complex Hardy space $H^2(\mathbf{B}_1)$ with the Szegö kernel dictionary $\mathcal{D}(\mathbf{B}_1)$, precisely corresponds to $\mathcal{N} = L^2(\Omega, H^2(\mathbf{B}_1))$. For $f \in \mathcal{N}$, there holds, for a.s. $\omega \in \Omega$, $f_\omega \in H^2(\mathbf{B}_1)$, and the stochastic MSP (SMSP) reads as

$$a_k = \arg\max \left\{ \mathbb{E}_\omega \left|\left\langle f_\omega, B_k^a \right\rangle\right|^2 | a \in \mathbf{B}_1 \right\}, \tag{11}$$

where for the previously known $a_1, \cdots, a_{k-1}$,

$$B_k^a(z) = e_a(z) \prod_{l=1}^{k-1} \frac{z - a_l}{1 - \bar{a}_l z} \tag{12}$$

is the kth term of the TM system with the undetermined parameter $a \in \mathbf{B}_1$, to be optimally determined according to (11). Existence of an optimal a_k is proved in Qian [1]. Then, with such optimally chosen a_k and $B_k = B_k^{a_k}$, the consecutively determined TM system gives rise to an expansion of f in the $\mathcal{N}$ norm sense [1]:

$$f(e^{it}, \omega) \overset{\mathcal{N}}{=} \sum_{k=1}^{\infty} \langle f_\omega, B_k \rangle B_k \left(e^{it}\right).$$

SPOAFD (identical with SPOAFD2 in the terminology of Qian [1]) is for a general Bochner space $\mathcal{N} = L^2(\Omega, L^2(\mathbf{D}))$, where the space $L^2(\mathbf{D})$ has a BVC dictionary. We have the same result except that the TM system $\{B_k\}$ is replaced by the orthonormal system $\{E_k\}$, as composed in (7), using the multiple kernels $\tilde{K}_{q_k}$, where q_k is selected according to the *stochastic pre-orthogonal MSP* (SPOMSP):

$$q_k = \arg\max \left\{ \mathbb{E}_\omega \left| \left\langle f_\omega, E_k^q \right\rangle \right|^2 | \, q \in \mathbf{D} \right\}. \tag{13}$$

In the case, there holds in the $\mathcal{N}$ norm sense [1]:

$$f(x, \omega) \overset{\mathcal{N}}{=} \sum_{k=1}^{\infty} \langle f_\omega, E_k \rangle E_k(x). \tag{14}$$

The strength of SPOAFD is that the optimally selected parameters q_ks generate an orthonormal system that gives rise to sparse representation of $f(\omega, \cdot)$ in the $L^2(\mathbf{D})$ norm sense for a.s. ω. The expansion, as the Karhunen–Loève (KL) decomposition does, enjoys the optimal convergence rate $O\left(\frac{1}{\sqrt{n}}\right)$. SPOAFD, as an extension of SAFD, can be associated with any BVC dictionary. As with the deterministic case [2], such flexibility makes SPOAFD a convenient tool to solve boundary value and initial boundary value problems. Dirichlet boundary value and Cauchy initial value problems with random data are studied in Yang [5]: We use the dictionaries of the shifted and dilated Poisson and heat kernels, respectively. After obtaining sparse representations of the random boundary or initial values, we perform the lifting up technology based on the semigroup properties of the kernels to get the random function solutions. The same method can be used to solve a wide type random boundary and initial value problems ([29]). SAFD and SPOAFD are remarkably convenient in the computation respect: The computation only uses the covariance function but not the eigenvalues and eigenfunctions of the related integral operator. In contrast, the KL expansion crucially relies on the eigenpairs that have to be firstly computed out in applications.

(g) Stochastic n-best (SnB) POAFD: The related SnB approximation problems are first formulated and studied in Qu et al. [30] (specially for the stochastic complex Hardy spaces) and further in Qian [27] (for a wide class of stochastic RKHSs). The general formulation in Qian [27] is as follows. For any n-tuple $\mathbf{p} = (p_1, \cdots, p_n) \in \mathbf{D}^n$, where multiplicity is allowed, there exists an n-orthonormal system $\{E_k^{\mathbf{p}}\}_{k=1}^n$, generated by the corresponding, possibly multiple, kernels $\{\tilde{K}_{p_k}\}_{k=1}^n$ through the GS process. The associated objective function to be maximized is

$$A(f, \mathbf{p}) = \mathbb{E}_\omega \left(\sum_{k=1}^{n} \left| \left\langle f_\omega, E_k^{\mathbf{p}} \right\rangle \right|^2 \right). \tag{15}$$

In other words, the SnB problem amounts to finding $\mathbf{q} \in \mathbf{D}^n$ such that

$$A(f, \mathbf{q}) = \sup\{A(f, \mathbf{p}) \, | \, \mathbf{p} \in \mathbf{D}^n\}.$$

The just formulated SnB is so far the state of the art among the existing variations of the POAFD-type models. The main goal of the present paper is to compare SPOAFD and SnB with the KL decomposition method for decomposing stochastic processes, the latter, methodology-wise, being of the same kind as POAFD in contrast with the Wiener chaos type.

In Section 2, we give an account of KL expansion with regards to the points concerned in this study. We cite brief proofs of the known results for the self-containing purpose. Some of the results and the proofs in Theorem 2 may be new. In Section 3, we prove some convergence rate results, analyze and compare the AFD type and the KL expansions, and specify, in both the theory and computational aspects, particular properties of the two types, respectively. Illustrative experiments are contained in Section 4. Conclusions are drawn in Section 5.

2 | AN ACCOUNT TO KL EXPANSION

Either of an orthonormal basis of $L^2(\Omega)$ or one in $L^2(\partial\mathbf{D})$ can be used to induce a decomposition of $f \in L^2(\Omega, L^2(\partial\mathbf{D}))$. The Fourier–Hermite (Wiener–Chaos) expansion and KL expansion, respectively, correspond to a basis in $L^2(\Omega)$ and $L^2(\partial\mathbf{D})$. Since KL is of the same type as the AFD ones, and they concern the same type of problems, in this study, we restrict ourselves to only analyze the KL and the AFD-type decompositions.

The following material of the KL decomposition is standard [28]. We involve a compact set $\mathcal{T}$ of the Euclidean space as time or space domain of the random function in the Bochner space $L^2(\Omega, L^2(\mathcal{T}))$, where $\mathcal{T}$ is in place of $\partial\mathbf{D}$ of the proceeding context. Let $f(t,\omega)$ be given in $L^2(\Omega, L^2(\mathcal{T}))$ and fixed throughout the rest of this section, and $\mu(t) = \mathbb{E}_\omega f(t,\cdot)$. Denote by $\mathbf{C}$ the covariance function:

$$\mathbf{C}(s,t) = \mathbb{E}_\omega\left[(f(s,\cdot) - \mu(s))(\bar{f}(t,\cdot) - \bar{\mu}(t))\right],$$

and T the integral operator using $\mathbf{C}(s,t)$ as its kernel, $T : L^2(\mathcal{T}) \to L^2(\mathcal{T})$,

$$TF(s) = \int_{\mathcal{T}} \mathbf{C}(s,t)F(t)dt.$$

We denote by $R(T)$ the range of the operator T, and $R(L)$ the range of the operator $L : L^2(\Omega) \to L^2(\mathcal{T})$, defined as, for any $g \in L^2(\Omega)$,

$$L(g)(t) = E_\omega(g\overline{[f^t - \mu(t)]}) = \int_\Omega g(\omega)\overline{[f(t,\omega) - \mu(t)]}d\mathbb{P}(\omega),$$

where $f^t(\omega) = f(t,\omega)$. It is asserted that T is a Hilbert-Schmidt operator in $L^2(\mathcal{T})$, and hence compact. The kernel function $\mathbf{C}$ is conjugate-symmetric and semi-positive. As a consequence, T has a sequence of positive eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq \cdots > 0$, and correspondingly, the eigenfunctions $\phi_1, \cdots, \phi_n, \cdots, T\phi_k = \lambda_k\phi_k$, orthogonal with each other. If $\text{span}\{\phi_1, \cdots, \phi_n, \cdots\} \neq L^2(\mathcal{T})$, that is, $R(T) \neq L^2(\mathcal{T})$, a supplementary orthonormal system (corresponding to the zero eigenvalue) may be added to form a complete basis system, called a KL basis, still denoted by $\{\phi_k\}$ with the property $T\phi_k = \lambda_k\phi_k$, where λ_k now may be zero. There holds

$$\mathbf{C}(s,t) = \sum_{k=1}^{\infty} \lambda_k\phi_k(s)\bar{\phi}_k(t). \tag{16}$$

When $\mathbf{C}$ is continuous in $\mathcal{T} \times \mathcal{T}$, all the ϕ_k are continuous, and the above convergence is uniform and absolute. The originally given $f(\cdot,\omega)$, a.s. belonging to $L^2(\mathcal{T})$, has the so-called KL decomposition: for a.e. $t \in \mathcal{T}$, in the probability square-mean sense

$$f(t,\omega) - \mu(t) \overset{L^2(\Omega)}{=} \sum_{k=1}^{\infty} \sqrt{\lambda_k}\phi_k(t)\xi_k(\omega) \overset{L^2(\Omega)}{=} \lim_{n\to\infty} S_n(t,\omega), \tag{17}$$

and, furthermore, in the Bochner norm sense

$$f(t,\omega)-\mu(t)\overset{\mathcal{N}}{=}\sum_{k=1}^{\infty}\sqrt{\lambda_k}\phi_k(t)\xi_k(\omega)\overset{\mathcal{N}}{=}\lim_{n\to\infty}S_n(t,\omega), \tag{18}$$

where $S_n(t,\omega)\triangleq\sum_{k=1}^{n}\sqrt{\lambda_k}\phi_k(t)\xi_k(\omega), S_0=0$, and for the non-zero λ_k,

$$\xi_k(\omega)=\frac{1}{\sqrt{\lambda_k}}\langle f_\omega-\mu,\phi_k\rangle_{L^2(\mathcal{T})}.$$

The random variables are uncorrelated, zero mean, and of unit variance. If the process is Gaussian, then $\xi_k\sim N(0,1)$ iid.

Since $\mathbf{C}(s,t)$ is conjugate-symmetric and semi-positive, by the Moor–Aronszajn Theorem, it uniquely determines a RKHS. We will show that the RKHS under the Moor–Aronszajn Theorem is identical with the following defined Hilbert space $H_{\mathbf{C}}$. Let $\alpha_k=1$ if $\lambda_k>0$; and $\alpha_k=0$ if $\lambda_k=0$. Define

$$H_{\mathbf{C}}=\left\{F\in L^2(\mathcal{T})\,|\,\|F\|^2_{H_{\mathrm{C}}}=\sum_{k=1}^{\infty}\alpha_k\frac{|\langle F,\phi_k\rangle|^2}{\lambda_k}<\infty\right\},$$

whose inner product is defined as

$$\langle F,G\rangle_{H_{\mathrm{C}}}=\sum_{k=1}^{\infty}\alpha_k\frac{\langle F,\phi_k\rangle\overline{\langle G,\phi_k\rangle}}{\lambda_k},$$

where the role of the α_k is that when λ_k is zero, the corresponding terms in the above two series vanish.

We collect in the following theorem the fundamentals of the KL expansion.

Theorem 2.

(i) $H_{\boldsymbol{C}}\subset L^2(\mathcal{T})$.

(ii) $\mathrm{var}[f(t,\cdot)-S_n(t,\cdot)]=\boldsymbol{C}(t,t)-\sum_{k=1}^{n}\lambda_k\phi_k^2(t)$, $\|f-S_n\|_{\mathcal{N}}=\sum_{k=n+1}\lambda_k$, $\|\mathrm{var} f\|_{L^2(\partial\boldsymbol{D})}=\sum_{k=1}^{\infty}\lambda_k$, *and* $\|S_n\|^2_{\mathcal{N}}=\mathbb{E}\sum_{k=1}^{n}|\langle f_\omega-\mu,\phi_k\rangle|^2=\sum_{k=1}^{n}\langle T\phi_k,\phi_k\rangle=\sum_{k=1}^{n}\lambda_k$.

(iii) $H_{\boldsymbol{C}}$ is the RKHS with reproducing kernel $\boldsymbol{C}(s,t)$, and, in particular, $\overline{\mathrm{span}\{\boldsymbol{C}(s,\cdot)\}}=H_{\boldsymbol{C}}$, where the bar stands for the closure under the $H_{\boldsymbol{C}}$ norm.

(iv) The KL basis has the optimality property: For any orthonormal basis $\{\psi_k\}$ of $L^2(\partial\boldsymbol{D})$ and any n, there holds $\sum_{k=1}^{n}\langle T\psi_k,\psi_k\rangle\le\sum_{k=1}^{n}\langle T\phi_k,\phi_k\rangle$.

(v) If there are only finitely many λ_ks being non-zero, then $f_\omega\in H_{\boldsymbol{C}}$ for a.s. $\omega\in\Omega$.

(vi) $R(L)=H_{\boldsymbol{C}}$ in the set-theoretic sense, and moreover, $H_K=H_{\boldsymbol{C}}$, where H_K is the RKHS over $R(L)$ defined with the $\mathcal{H}$-H_K formulation (as in Qian [3]). In particular, when taking $L^2(\Omega)$ as $L^2(\partial\boldsymbol{D})$ and f the parameterized Szegö kernel, we have, as isometric spaces, $R(L)=H_K=H_{\boldsymbol{C}}=H^2(\boldsymbol{B}_1)$, latter being the classical Hardy space in the disc.

(vii) Set $\boldsymbol{C}^\theta(s,t)=\sum_{k=1}^{\infty}\lambda_k^\theta\phi_k(s)\bar{\phi}_k(t), 0<\theta<\infty$, and

$$H^\theta_{\boldsymbol{C}}=\left\{F\in L^2(\partial\boldsymbol{D})\;:\;\|F\|^2_{H^\theta_C}=\sum_{k=1}^{\infty}\alpha_k\frac{|\langle F,\phi_k\rangle|^2}{\lambda_k^\theta}<\infty\right\}.$$

In particular, $\boldsymbol{C}^1=\boldsymbol{C}$ and $H^1_{\boldsymbol{C}}=H_{\boldsymbol{C}}$. Then, $H^\theta_{\boldsymbol{C}}$ is the RKHS with reproducing kernel $\boldsymbol{C}^\theta$, and $TH^\theta_{\boldsymbol{C}}=H^{\theta+2}_{\boldsymbol{C}}$ giving rise to an isometric isomorphism between $H^\theta_{\boldsymbol{C}}$ and $H^{\theta+2}_{\boldsymbol{C}}, 0<\theta<\infty$.

(viii) The identity mappings $R(T)=H^2_{\boldsymbol{C}}\to H_{\boldsymbol{C}}\to L^2(\partial\boldsymbol{D})$ are bounded imbeddings. $H^2_{\boldsymbol{C}}=H_{\boldsymbol{C}}=L^2(\partial\boldsymbol{D})$ in the set-theoretic sense and in the norm equivalent sense if and only if there exist only a finite number of non-zero λ_ks. The bounded imbedding and the isometric isomorphism conclusions between $H^2_{\boldsymbol{C}}=H_{\boldsymbol{C}}=L^2(\partial\boldsymbol{D})$ are extendable to all different $H^\theta_{\boldsymbol{C}}, 0<\theta<\infty$.

Lord et al. [28] is referred as basic reference of KL. For more detailed and general formulation of the KL expansion, we refer to the recent papers Kanagawa et al. [31], Steinwart and Scovel [32], and Steinwart [33]. The nesting RKHSs of the power RKHSs of $H_{\mathbf{C}}$ are summarized in the view point of Sobolev spaces and are presented in terms of the $\mathcal{H}$-H_K formulation of Qian [3]. For $H_{\mathbf{C}}^{\theta}$ in the range $0 < \theta < 1$, we refer to Kanagawa et al. [31] and Steinwart [33]. $H_{\mathbf{C}}^{\theta}$ for the range $1 < \theta < \infty$ corresponds to the Sobolev spaces in the classical setting. For the self-containing and the algorithm concerns, we outline the proofs. The proof of (iv) by using the simplex algorithm may be new.

Proof. (i) follows from the definition of $H_{\mathbf{C}}$. The first relation of (ii) follows from uncorrelation of the ξ_ks (see, for instance, Lord et al. [28]). We now deduce the other relations. Due to the orthonormality of ϕ_k, we have $\|S_n\|^2_{\mathcal{N}} = \mathbb{E}\sum_{k=1}^{n} |\langle f_\omega, \phi_k\rangle|^2$. Then,

$$\begin{aligned}\mathbb{E}\sum_{k=1}^{n} |\langle f_\omega - \mu, \phi_k\rangle|^2 &= \mathbb{E}\sum_{k=1}^{n} \langle f_\omega - \mu, \phi_k\rangle\overline{\langle f_\omega - \mu, \phi_k\rangle} \\ &= \sum_{k=1}^{n}\int_{\partial\mathbf{D}}\int_{\partial\mathbf{D}} \mathbb{E}([f(s,\cdot)-\mu]\overline{[f(t,\cdot)-\mu]})\phi_k(t)\bar{\phi}_k(s)dtds \\ &= \sum_{k=1}^{n}\langle T\phi_k, \phi_k\rangle \\ &= \sum_{k=1}^{n}\lambda_k.\end{aligned} \tag{19}$$

Now, we show (iii). By definition, $H_{\mathbf{C}}$ is a Hilbert space. For any fixed $s \in \mathcal{T}$, as a consequence of (16),

$$\langle \mathbf{C}_s, \phi_k\rangle = \lambda_k\phi_k(s).$$

Hence,

$$\sum_{k=1}^{\infty}\alpha_k\frac{|\langle \mathbf{C}_s, \phi_k\rangle|^2}{\lambda_k} = \sum_{k=1}^{\infty}\lambda_k\phi_k^2(s).$$

Since the $L^1(\partial\mathbf{D})$-norm of the last function in s is equal to $\sum_{k=1}^{\infty}\lambda_k < \infty$, we have $\|\mathbf{C}_s\|_{H_{\mathbf{C}}} < \infty$, a.e.. This implies that for a.e. $s \in \partial\mathbf{D}$, $\mathbf{C}_s$ belongs to $H_{\mathbf{C}}$. Let $F \in H_{\mathbf{C}}$ with $F = \sum_{k=1}^{\infty}\alpha_k\langle F, \phi_k\rangle\phi_k$. Then,

$$\begin{aligned}\langle F, \mathbf{C}_s\rangle_{H_{\mathbf{C}}} &= \sum_{k=1}^{\infty}\alpha_k\frac{\langle F, \phi_k\rangle\overline{\langle \mathbf{C}_s, \phi_k\rangle}}{\lambda_k} \\ &= \sum_{k=1}^{\infty}\alpha_k\frac{\langle F, \phi_k\rangle\lambda_k\phi_k(s)}{\lambda_k} \\ &= F(s),\end{aligned}$$

verifying the reproducing property of $\mathbf{C}_s$.

Next, we show (iv). Let n be fixed, and $m \geq n$, and T_m be the integral operator defined through the kernel $\mathbf{C}_m(s,t) = \sum_{j=1}^{m}\lambda_j\phi_j(s)\bar{\phi}_j(t)$, where $\phi_1, \cdots, \phi_m$ are the first m functions of the entire KL basis $\{\phi_j\}_{j=1}^{\infty}$. For any n-orthonormal system $\{\psi_1, \cdots, \psi_n\}$, there holds

$$\begin{aligned}\sum_{k=1}^{n}\langle T_m\psi_k, \psi_k\rangle &= \sum_{k=1}^{n}\int_{\mathcal{T}}\int_{\mathcal{T}}\mathbf{C}_m(s,t)\psi_k(t)\bar{\psi}_k(s)dtds \\ &= \sum_{k=1}^{n}\int_{\mathcal{T}}\int_{\mathcal{T}}\sum_{j=1}^{m}\lambda_j\phi_j(s)\overline{\phi}_j(t)\psi_k(t)\bar{\psi}_k(s)dtds \\ &= \sum_{k=1}^{n}\sum_{j=1}^{m}\lambda_j|\langle\psi_k, \phi_j\rangle|^2.\end{aligned}$$

Denote $c_{kj} = |\langle \psi_k, \phi_j \rangle|^2$. Now, for the fixed $\lambda_j, j = 1, \cdots, m$, we are to solve the global maximization problem for the linear objective function:

$$A = \sum_{k=1}^{n} \sum_{j=1}^{m} \lambda_j c_{kj},$$

under the constraint conditions

$$\begin{cases} \sum\limits_{k=1}^{n} c_{kj} + \alpha_j = 1, \\ \sum\limits_{j=1}^{m} c_{kj} + \beta_k = 1, \\ 1 \geq c_{kj} \geq 0, \alpha_j \geq 0, \beta_k \geq 0, \\ \lambda_1 \geq \cdots \geq \lambda_n \geq \cdots \lambda_m, \\ 1 \leq k \leq n, 1 \leq j \leq m, \end{cases}$$

where the first two constraint conditions are due to the Bessel inequality. This is a typical simplex algorithm problem that attains the greatest possible value at some vertex points of the defined region. Note that the objective function is invariant under permutations on c_{kj}. Through testing the function values at the vertex points, we obtain that the optimal solution is attainable at and only at $c_{kk} = 1, 1 \leq k \leq n; c_{kj} = 0, k \neq j; \alpha_j = 1, n < j \leq m; \alpha_j = 0, 1 \leq j \leq n; \beta_k = 0, 1 \leq k \leq n$. This solution amounts that $\psi_k = \phi_k, 1 \leq k \leq n$. As a consequence of the above simplex algorithm solution, for any general n-orthonormal system $\{\psi_1, \cdots, \psi_n\}$, there hold

$$\sum_{k=1}^{n} \langle T_m \psi_k, \psi_k \rangle \leq \sum_{k=1}^{n} \lambda_k = \sum_{k=1}^{n} \langle T_m \phi_k, \phi_k \rangle.$$

Letting $m \to \infty$, since $\lim_{m\to\infty} T_m \psi_k = T\psi_k, 1 \leq k \leq n$, we obtain the inequality claimed in (iv).

Next, we show (v). For a.s. $\omega \in \Omega$, we have the series expansion (18). Hence,

$$\begin{aligned} \|f_\omega - \mu\|_{H_{\mathbf{C}}}^2 &= \sum_{k=1}^{\infty} \alpha_k \frac{|\langle f_\omega, \phi_k \rangle|^2}{\lambda_k} \\ &= \sum_{k=1}^{\infty} \alpha_k |\xi_k|^2. \end{aligned}$$

If there are finitely many λ_ks being non-zero, then

$$\mathbb{E}_\omega \left(\sum_{k=1}^{\infty} \alpha_k |\xi_k|^2 \right) = \sum_{k=1}^{\infty} \alpha_k < \infty.$$

This implies that for Probability 1, there holds $\|f_\omega - \mu\|_{H_{\mathbf{C}}} < \infty$, and thus, a.s. $f_\omega - \mu \in H_{\mathbf{C}}$.

We now show (vi). Let $g(\omega) \in L^2(\Omega)$. Denote the image of g under L by $G = Lg$. By using the KL expansion of f,

$$G(t) = \int_\Omega g(\omega) \overline{[f(t,\omega) - \mu(t)]} d\mathbb{P} = \sum_{k=1}^{\infty} \sqrt{\lambda_k} \langle g, \xi_k \rangle_{L^2(\Omega)} \phi_k(t).$$

Now, we examine the $H_{\mathbf{C}}$ norm of G. Noting that $\{\xi_k\}$ is an orthonormal system in $L^2(\Omega)$, by invoking the Bessel inequality, we have

$$\|G\|_{H_{\mathbf{C}}} = \sum_{k=1}^{\infty} \alpha_k |\langle g, \xi_k \rangle_{L^2(\Omega)}|^2 \leq \|g\|_{L^2(\Omega)}^2 < \infty.$$

Therefore, $L(g) \in H_{\mathbf{C}}$. By the Riesz–Fisher Theorem and the definition of $H_{\mathbf{C}}$, the mapping $L : \mathcal{H} \to H_{\mathbf{C}}$ is onto. Hence, in the set-theoretic sense $R(L) = H_{\mathbf{C}}$. With the $\mathcal{H}$-H_K formulation of Qian [3] (also see Saitoh and Sawano [4]), $R(L)$ is equipped with an inner product by which it is a RKHS, H_K, for which $\mathbf{C}$ is the reproducing kernel. The inner product used there is induced by that of $L^2(\Omega)$ on the equivalent classes of the form $L^{-1}(G)$ (as set inverse), $G \in H_{\mathbf{C}}$. Now, $H_{\mathbf{C}}$ is also a RKHS with the same reproducing kernel $\mathbf{C}$. The uniqueness part of the Moor–Aronszajn Theorem asserts that the two RKHSs, namely, H_K and $H_{\mathbf{C}}$, have to be the same. The latter stands as a realization of the former in terms of the eigenvalues and eigenfunctions of the integral operator T.

Next, we prove (vii). Obviously, the norm $\|F\|^2_{H^\theta_{\mathbf{C}}}$ is equipped with the inner product

$$\langle F, G\rangle_{H_{\mathbf{C}_j}} = \sum_{k=1}^{\infty} \alpha_k \frac{\langle F, \phi_k\rangle \overline{\langle G, \phi_k\rangle}}{\lambda_k^\theta}.$$

Under this inner product, the reproducing kernel property can be verified: For $F = \sum_{k=1}^{\infty} \langle F, \phi_k\rangle \phi_k \in H^\theta_{\mathbf{C}}$,

$$\langle F, \left(\mathbf{C}^\theta\right)_s\rangle_{H_{\mathbf{C}^\theta}} = \sum_{k=1}^{\infty} \alpha_k \frac{\langle F, \phi_k\rangle \overline{\left\langle \left(\mathbf{C}^\theta\right)_s, \phi_k\right\rangle}}{\lambda^\theta_k} = \sum_{k=1}^{\infty} \alpha_k \frac{\langle F, \phi_k\rangle \lambda_k^\theta \phi_k(s)}{\lambda_k^\theta} = F(s).$$

To verify the last statement of (viii), since $TF = \sum_{k=1}^{\infty} \lambda_k \langle F, \phi_k\rangle \phi_k$, we have

$$\|TF\|^2_{H^{\theta+2}_{\mathbf{C}}} = \sum_{k=1}^{\infty} \alpha_k \frac{|\lambda_k \langle F, \phi_k\rangle|^2}{\lambda_k^{\theta+2}} = \sum_{k=1}^{\infty} \alpha_k \frac{|\langle F, \phi_k\rangle|^2}{\lambda_k^\theta} = \|F\|^2_{H^\theta_{\mathbf{C}}}.$$

Finally, we show (viii), and first show $R(T) \subset H_{\mathbf{C}}$. Letting $F = \sum_{k=1}^{\infty} \langle F, \phi_k\rangle \phi_k \in L^2(\mathcal{T})$, then

$$TF(s) = \int_{\mathcal{T}} \mathbf{C}(s,t)F(t)dt = \sum_{k=1}^{\infty} \alpha_k \lambda_k \langle F, \phi_k\rangle \phi_k(s) = \sum_{k=1}^{\infty} \tilde{c}_k \phi_k(s), \ \tilde{c}_k = \alpha_k \lambda_k \langle F, \phi_k\rangle.$$

Those coefficients satisfy the condition

$$\sum_{k=1}^{\infty} \frac{|\tilde{c}_k|^2}{\lambda_k^2} = \sum_{k=1}^{\infty} \alpha_k |\langle F, \phi_k\rangle|^2 < \infty.$$

Since $\lambda_k \to 0$, the above condition implies

$$\sum_{k=1}^{\infty} \frac{|\tilde{c}_k|^2}{\lambda_k} < \infty,$$

and hence, $TF \in H_{\mathbf{C}}$. By invoking the Riesz–Fisher Theorem, $R(T)$ has its natural inner product in

$$H^2_{\mathbf{C}} \triangleq \left\{ F \in L^2(\partial \mathbf{D}) \ : \ \sum_{k=1}^{\infty} \alpha_k \frac{|\langle F, \phi_k\rangle|^2}{\lambda_k^2} < \infty \right\}.$$

Equipped with the above induced inner product below, we may identify $R(T)$ with $H^2_{\mathbf{C}}$. Since $\lambda_k \downarrow 0$, being affiliated with the $H^2_{\mathbf{C}}$-norm, the identical mapping from $R(T)$ to $H_{\mathbf{C}}$ is a bounded imbedding: In fact,

$$\|F\|_{H_{\mathbf{C}}} \le \lambda_1 \|F\|_{H^2_{\mathbf{C}}}.$$

In the chain of the imbeddings induced by the monotonous property $H_{\mathbf{C}}^{\delta} \uparrow$ for $\delta \downarrow 0$ in the set-theoretic sense, and in particular, $H_{\mathbf{C}}^{k+\theta} \subset R(T) = H_{\mathbf{C}}^{2} \subset H_{\mathbf{C}}^{1} = R(L) \subset H_{\mathbf{C}}^{\theta} \subset L^2(\partial \mathbf{D}), 0 < \theta < 1, k = 1, 2, \cdots$, any of the two spaces, therefore, all the spaces are identical with $L^2(\partial \mathbf{D})$ in the set-theoretic sense if and only if there exist only finitely many non-zero λ_ks, for otherwise, $\sum_{k=1}^{\infty} \alpha_k \lambda_k < \infty$ and $0 \neq \lambda_k \to 0$ would imply $\sum_{k=1}^{\infty} \sqrt{\lambda_k} \phi_k \in L^2(\partial \mathbf{D}) \backslash H_{\mathbf{C}}$, and $\sum_{k=1}^{\infty} \lambda_k \phi_k \in H_{\mathbf{C}} \backslash R(T)$. Thus, (viii) is proved. The proof of the theorem is complete. □

Remark 3. We note that, in the proof of (iv) of Theorem 2, not only in the particular case $\psi_k = \phi_k, k = 1, \cdots, n$, but also this special simplex algorithm problem also attains its greatest maximal value when $\mathrm{span}\{\psi_1, \cdots, \psi_n\} = \mathrm{span}\{\phi_1, \cdots, \phi_n\}$. In fact, in such cases, $c_{kj} = |\langle \psi_k, \phi_j \rangle|^2 = 0, k \leq n < j \leq m$, and the Bessel inequality becomes the Placherel identity, $\sum_{k=1}^{n} c_{kj} = \sum_{j=1}^{n} c_{kj} = 1$, that implies, following the deductions in the proof,

$$\sum_{k=1}^{n} \langle T_m \psi_k, \psi_k \rangle = \sum_{k=1}^{n} \sum_{j=1}^{m} \lambda_j |\langle \psi_k, \phi_j \rangle|^2 = \sum_{j=1}^{n} \lambda_j \sum_{k=1}^{n} |\langle \psi_k, \phi_j \rangle|^2 = \sum_{j=1}^{n} \lambda_j .$$

Remark 4. The spaces defined in (vii) are called power spaces of $H_{\mathbf{C}}$. The range corresponding to $0 < \theta \leq 1$ of the family $H_{\mathbf{C}}^{\theta}$ has recently been studied in the random function literature [31, 33]. This range of $H_{\mathbf{C}}^{\theta}$ can be regarded as non-smoothness extensions, as opposite to the traditional smooth extensions the usual Sobolev-type spaces. In the occasion that the basic random function $f(t, \omega)$ as standard Brownian motion on $[0, 1]$ being a Gaussian process with the kernel $k(x, y) = \min(x, y)$ for $x, y \in [0, 1]$, the space RKHS $H_{\mathbf{C}} = R(L)$ coincides with the first-order Sobolev space. The study of the non-smooth $H_{\mathbf{C}}^{\theta}$ is intimately related to the zero–one law of Dirscoll [34], being generalized by Lukić and Beder [35], asserting that by Probability 1 or alternatively 0, $f(t, \omega)$ belongs to $H_{\mathbf{C}}$. Based on the non-smooth power RKHS concept and the related studies (see the above cited and references therein), one shows that the initial random function $f(t, \omega)$ interested must belong to a power RKHS $H_{\mathbf{C}}^{\theta}, 0 < \theta < 1$, so to allow implementation of the RKHS methods.

3 | THE SPOAFD METHODS IN COMPARISON WITH KL

3.1 | The algorithms of SAFD, SPOAFD, and SnB

We will give the computational details of SPOAFD in which SAFD corresponds to the cases where SPOAFD is restricted to the Hardy $H^2(\mathbf{B}_1)$ or the Hardy $H^2(\mathbf{C}^+)$ spaces. We note that only in those two cases rational orthogonal (or TM) systems are as orthogonalization of the parameterized Szegö kernels. Let $\{E_q\}$ be a BVC dictionary. We are to find inductively $q_1, \cdots, q_k, \cdots$, such that with the notation of Section 1,

$$q_k = \arg\max \left\{ \mathbb{E}_{\omega} \left| \left\langle f_{\omega} - \mu, E_k^q \right\rangle \right|^2 | \, q \in \mathbf{D} \right\} . \tag{20}$$

As in the proof of (ii) of Theorem 2, the quantity $\mathbb{E}_{\omega} \left| \left\langle f_{\omega} - \mu, E_k^q \right\rangle \right|^2$ may be identically reduced, for each k, to

$$\mathbb{E}_{\omega} |\langle f_{\omega} - \mu, E_k^q \rangle|^2 = \int_{\partial \mathbf{D}} \int_{\partial \mathbf{D}} \mathbf{C}(s, t) E_k^q(t) \bar{E}_k^q(s) dt ds. \tag{21}$$

For the notation E_k^q, see (7) and (13) and the explanations in the corresponding texts. With the above expression in terms of the covariance function $\mathbf{C}$, one can actually work out, with personal computer, approximations of the optimal parameters needed by the SPOAFD expansion. This computation does not require information of the eigenvalues and eigenfunctions of the integral operator defined by the covariance kernel. In contrast, the eigenpairs information is crucial in order to carry on the KL expansion. For a general kernel operator finding the information of its eigenpairs is by no means easy. Practically, one can only get, by using linear algebra based on sampling, numerical approximations of the eigenvalues and eigenfunctions. With SPOAFD, under a sequence of optimally selected parameters according to (21) and (20), the relation (14) holds. For a given n, with SnB, the objective function (21) is replaced by (22), seeking for an n-tuple of parameters $\mathbf{q} = (q_1, \cdots, q_n)$ that maximizes

$$\sum_{k=1}^{n} \int_{\partial \mathbf{D}} \int_{\partial \mathbf{D}} \mathbf{C}(s,t) E_k^{\mathbf{p}}(t) \bar{E}_k^{\mathbf{p}}(s) dt ds, \tag{22}$$

over all $\mathbf{p} = (p_1, \cdots, p_n) \in \mathbf{D}^n$. To actually find a practical solution to (22) one can use the cyclic algorithm given by [26], or its improvement [36]. The main estimation proved in Wang and Qian [25], or Qian et al. [37], or more generally in Qian [27], reduces the optimization problem to one in a compact disc and enables to use an algorithm for finding global maxima of a Lipschitz continuous functions on a compact set (a compact Lipschitz optimizer). Such an algorithm prevents itself from falling into a local minimum that is not global and yet is practical. One such algorithm by using a new compact Lipschitz optimizer is on its way ([37]).

There is a particular type of random fields whose SPOAFD computation may be simpler. If $f(t,\omega)$ has the form $F(t,X)$ where X is a random variable having probability density function $p(u), u \in U \subset (-\infty, \infty)$, then (21) may be computed as

$$\mathbb{E}_\omega \left| \left\langle f_\omega - \mu, E_n^q \right\rangle \right|^2 = \int_U \int_{\partial \mathbf{D}} \left| F(t,u) E_n^q(t) \right|^2 p(u) dt du.$$

See Yang [5] for concrete examples. In such situation with respect to the objective function (22) of SnB, there exists a similar formula.

3.2 | Optimality of KL over SPOAFD and SnB

The assertion (iv) of Theorem 2 is valid for any orthonomal basis or system $\{\psi_k\}$, and especially, for the complex orthonormal system $\{E_k\}$ obtained from optimally selected $\mathbf{q} = (q_1, \cdots, q_n, \cdots)$ under the SPOAFD or SnB maximal selection principle. There hold

$$\begin{aligned} \mathbb{E} \sum_{k=1}^{n} |\langle f_\omega - \mu, E_k \rangle|^2 &= \sum_{k=1}^{n} \langle TE_k, E_k \rangle \\ &\leq \sum_{k=1}^{n} \langle T\phi_k, \phi_k \rangle \\ &= \sum_{k=1}^{n} \lambda_k. \end{aligned} \tag{23}$$

We specially note that the functions E_ks are in general not eigenfunctions of the operator T. In spite of the optimality proved in (iv), experiments in Section 4, as well as the analysis in the proof of (iv) of Theorem 2, all exhibit that when n is large the efficiency of SnB is almost the same as KL. It is noted that the efficiency of SPOAFD and SAFD is very close to that of SnB with, however, much less computational complexity than the latter.

Remark 5. We also take the opportunity to note that in the POAFD algorithm $|\langle f, E_k \rangle|$ are not necessarily in the descending order. Consider e_1, e_2 being two dictionary elements and f is in the span of e_1, e_2. Assume that $\|f_{e_1}\| > |\langle f, e_1 \rangle| > |\langle f, e_2 \rangle|$, where f_{e_1} is denoted as the projection of f into the subspace perpendicular to e_1. Then, $|\langle f, E_2^{e_2} \rangle| = |\langle f_{e_1}, E_2^{e_2} \rangle| = \|f_{e_1}\| > |\langle f, e_1 \rangle| = |\langle f, E_1 \rangle|$, as claimed.

3.3 | Convergence rates

In this section, we adopt a more general formulation whose MSP is called weak stochastic MSP (WSPOMSP) as given in (24). The corresponding algorithm is accordingly phrased as WSPOAFD.

For $\rho \in (0,1]$ and each k, WSPOMSP involves determination of a $q_k \in \mathbf{D}$ such that

$$\mathbb{E}_\omega \left| \left\langle f_\omega - \mu, E_k^{q_k} \right\rangle \right|^2 \geq \rho \sup \left\{ \mathbb{E}_\omega \left| \left\langle f_\omega - \mu, E_k^{q} \right\rangle \right|^2 | q \in \mathbf{D} \right\}. \tag{24}$$

If, in particular, $\rho = 1$, then the WSPOMSP reduces to SPOMSP in (13). In literature (see, for instance, Temlyakov [7]), there exist more general setting allowing different $\rho_k \in (0, 1]$ for different k. We, however, keep it simple and adopt a uniform $\rho \in (0, 1]$ for all k. Denote

$$M(\omega) = \inf \left\{ \sum_{l=1}^{\infty} |c_l(\omega)| \mid f_\omega - \mu = \sum_{l=1}^{\infty} c_l(\omega) E_{q_l}, \forall l, q_l \in \mathbf{D} \right\} \tag{25}$$

and

$$M_0 = \left(\int_\Omega |M(\omega)|^2 d\mathbb{P} \right)^{\frac{1}{2}}.$$

The following is an immediate consequence of the convergence rate of the AFD-type algorithms. See, for instance, Qian [10].

Theorem 6. *Denote g_n the n-standard remainder of the WSPOAFD algorithm,*

$$g_n(x, \omega) = f(x, \omega) - \mu(x) - \sum_{k=1}^{n-1} \langle (f_\omega - \mu(x))_k, E_{q_k} \rangle E_k(x).$$

There holds the estimation

$$\|g_n\|_{\mathcal{N}} \leq \frac{M_0}{\rho\sqrt{n}}.$$

Proof. By invoking the deterministic case result, Theorem 3.3 of Qian [10], there holds, for a.s. $\omega \in \Omega$,

$$\|(g_\omega)_n\|_{L^2(\mathbf{D})} \leq \frac{M(\omega)}{\rho\sqrt{n}}.$$

By taking the square-norm of the probability space to both sides, we obtain the claimed estimation.

For the deterministic case, there holds $M(\omega) \equiv M_0$, and the estimation reduces to the one for WPOAFD. For $\rho = 1$, the estimation gives rise to that of SPOAFD. □

Since the one by one parameters selection model in SPOAFD is surely less optimal than SnB, we obtain that the convergence rate for SnB is at least the same as that for SPOAFD given by Theorem 6. Note that if the used dictionary is itself an orthonormal system, then the above used convergence rate in the deterministic case quoted from Qian [10] coincides with that for $p = 1$ in DeVore and Temlyakov [11].

Theorem 7. *Let n be a fixed positive integer. For a given $f \in L^2(\Omega, L^2(\partial\mathbf{D}))$, there holds*

$$\left\| f - \mu - \sum_{l=1}^{n} \sqrt{\lambda_k} \phi_l \xi_l \right\|_{\mathcal{N}} \leq \frac{\| \sum_{l=n+1}^{\infty} |\langle f_\omega - \mu, \phi_l \rangle| \|_{L^2(\Omega)}}{\sqrt{n}}. \tag{26}$$

Proof. Due to uniqueness of expansion of f_ω in the basis $\{\phi_k\}$, the infimum in (25) reduces to

$$M(\omega) = \sum_{l=n+1}^{\infty} |\langle f_\omega - \mu, \phi_l \rangle|. \tag{27}$$

As a consequence of the optimality property of the KL basis, the n-partial sum expanded by the first n eigenfunctions of $f \in L^2(\Omega, L^2(\partial\mathbf{D}))$ is identical with its SnB with respect to the dictionary $\{\phi_l\}_{l=1}^{\infty}$. There holds, owing to (27) and Theorem 6,

$$\left\| f - \mu - \sum_{k=1}^{n} \sqrt{\lambda_k}\phi_k\xi_k \right\|_{\mathcal{N}} \leq \frac{\left\| \sum_{k=n+1}^{\infty} |\langle f_\omega - \mu, \phi_k \rangle| \right\|_{L^2(\Omega)}}{\sqrt{n}}.$$

The proof is complete. □

We note that the greedy-type method given in Theorem 6 seems to be rather rough. For general KL expansions, we have the energy of the tail

$$\left(\sum_{k=n+1}^{\infty} \lambda_k \right)^{\frac{1}{2}} < \infty.$$

In contrast, the right-hand side of (26) is

$$\frac{\left\| \sum_{k=n+1}^{\infty} \sqrt{\lambda_k}|\xi_k| \right\|_{L^2(\Omega)}}{\sqrt{n}},$$

provided that the last quantity is a finite number.

On the other hand, for the particular Brownian bridge case, for instance, the convergence rate can be precisely estimated as

$$\mathbb{E}\left[\|B - S_n\|^2_{L^2[0,1]} \right] = \sum_{j=n+1}^{\infty} \frac{1}{\pi^2 j^2} \sim \frac{1}{\pi^2 n}$$

(page 206 of Lord et al. [28]), showing that the convergence rate can indeed be as bad as $O\left(\frac{1}{\sqrt{n}}\right)$. As a common sense in Fourier analysis, convergence rate of an expansion is determined by smoothness of the function to be expanded. Almost surely a sample path of Brownian motion is continuous but no where differentiable. There is no wonder that such non-smoothness does not correspond to fast convergence.

The nesting structure of the RKHSs studied in (vii) and (viii) of Theorem 2 offers a sequence of finer and finer RKHSs to fill in the gap between $H_{\mathbf{C}}$ and $L^2(\partial\mathbf{D})$. Referring to the zero–one law of Driscoll, for a given $f(\cdot,\omega) \in L^2(\partial\mathbf{D})\backslash H_{\mathbf{C}}$ with a fixed sample path ω, one may find $\theta \in (0,1)$ such that $f(\cdot,\omega) \in H^\theta_{\mathbf{C}}$. In accordance with the convergence rate results obtained in Theorems 6 and 7, one has discretion to perform an AFD-type approximation in a selected space $H^\theta_{\mathbf{C}}$.

3.4 | Flexibility of dictionary selection for implementing SPOAFD

For a given random signal $f \in L^2(\Omega, L^2(\partial\mathbf{D}))$, both its KL and its SPOAFD decompositions are adaptive. SPOAFD, however, possesses extra adaptivity because the dictionary in use can again be selected according to the concrete task. As an example, in Yang [5], we numerically solve the Dirichelet problem of random data:

$$\begin{cases} \Delta u(x,\omega) = 0, \ \forall x \in \mathbf{D} \subseteq \mathbb{R}^{n+1}, \ \text{a.s.}\, \omega \in \Omega, \\ u(x,\omega) = f(x,\omega), \ \text{for a.e.}\, x \in \partial\mathbf{D}, \ \text{a.s.}\, \omega \in \Omega, \end{cases} \tag{28}$$

where $f \in L^2(\Omega, L^2(\partial\mathbf{D}))$, $\mathbf{D} = B_1$. According to the related Hardy space theory the solution u will belong to $L^2(\Omega, h^2(\mathbf{D}))$. The most convenient dictionary that we use in this case is the parameterized Poisson kernels (for the unit ball) $P_x(y')$ defined as

$$P_x(y') \triangleq P(x,y') \triangleq c_n \frac{1-r^2}{|x-y'|^n}, \ x = rx' \in B_1, x', y' \in \partial B_1, \tag{29}$$

In the algorithm, the random data $f(x,\omega)$ is first sufficiently approximated in the $\mathcal{N}$ norm $\mathcal{N} = L^2(\Omega, L^2(\partial\mathbf{D}))$ by SPOAFD series on the boundary using the parameterized Poisson kernels:

$$f_\omega(x') - \mu(x') \overset{\mathcal{N}}{=} \sum_{k=1}^{\infty} \langle f_\omega - \mu, E_k \rangle E_k(x') = \sum_{k=1}^{\infty} c_k(\omega)\tilde{P}_{x_k}(x'), \ x' \in B_1, \tag{30}$$

where $(E_1, \cdots, E_k, \cdots)$ are, consecutively, the GS orthogonalization of the multiple kernels $(\tilde{P}_{x_1}, \cdots, \tilde{P}_{x_k}, \cdots), x_k = r_k x'_k, k = 1, \cdots, n, \cdots$, and $c_k(\omega)$s are the coefficients in the span of $\{\tilde{P}_{x_k}\}$. Then, based on the semigroup property of the Poisson kernel, we can "lift up" each term of the spherical expansion and then add up to get the solution to the Dirichlet problem with the random data, that is, in the self-explanatory notation,

$$u_{f_\omega - \mu}(x) = \sum_{k=1}^{\infty} c_k(\omega) \tilde{P}_{x_k}(x), \quad x \in B_1. \tag{31}$$

The convergence speed of the n-partial sums of (31) is the same as that for (30), being of the rate $O\left(\frac{1}{\sqrt{n}}\right)$ as proved in Section 3. By using other dictionaries such as the heat kernel or any suitable convolution type kernel, or the eigensystem of the Mocer's kernel, although efficient approximation may be obtained at the boundary, there will be no convenience as lifting up the Poisson kernels to directly obtain the solution of the problem. Some other similar examples are also given in Yang [5] and Qu [29]. We finally note that, not like the KL expansion, SAFD, SPOAFD, and the related SnB are also available in unbounded domains, as long as the random function interested is in the corresponding Bochner-type space.

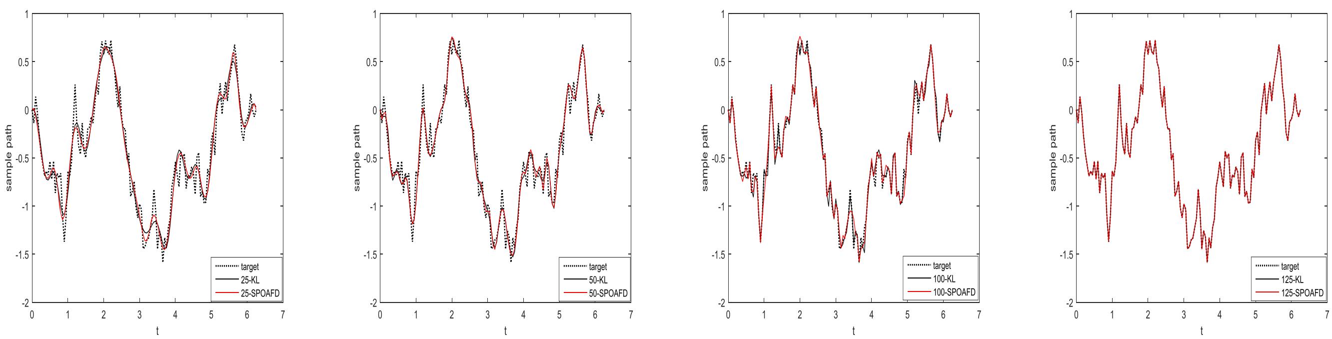
4 | EXPERIMENTS

In this section, we approximate the Brownian bridge in $[0, 2\pi]$. In the following experiments, the graphs of the targeted Brownian bridge are made by using the algorithm on page 195 of Lord et al. [28]. The KL expansions are according to the formula (5.44) on page 206 of Lord et al. [28]. The AFD-type methods are based on the covariance of the Brownian bridge, that is,

$$\mathbf{C}(s, t) = \min(s, t) - \frac{st}{2\pi}.$$

Example 1. (SPOAFD based on the Poisson kernel dictionary) The experiment is based on 126 sampling points in $[0, 2\pi]$ with the uniform spacing $\Delta t \approx 0.05$. As shown in Figure 1, SPOAFD by using the Poisson kernel dictionary has almost the same effect as that of the KL expansion. At the 125th partial sums, both SPOAFD and KL expansions approximately recover the target function. In the local details, SPOAFD seems to have visually better results. The relative errors of the two methods are given in Table 1.

Example 2. (SAFD on the Szegö kernel dictionary for the complex Hardy space on the disc space) We approximate the Brownian bridge by using the KL and the SAFD expansions based on 4096 and 1024 sampling points in $[0, 2\pi]$



SPOAFD: 25 partial sum SPOAFD: 50 partial sum SPOAFD: 100 partial sum SPOAFD: 125 partial sum

FIGURE 1 Sample path I (126 points) of Brownian bridge. [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 1 Relative error.

n partial sum	$n = 25$	$n = 50$	$n = 100$	$n = 125$
KL	0.0331	0.0140	0.0021	6.1397×10^{-31}
SPOAFD	0.0298	0.0113	0.0026	1.0984×10^{-7}

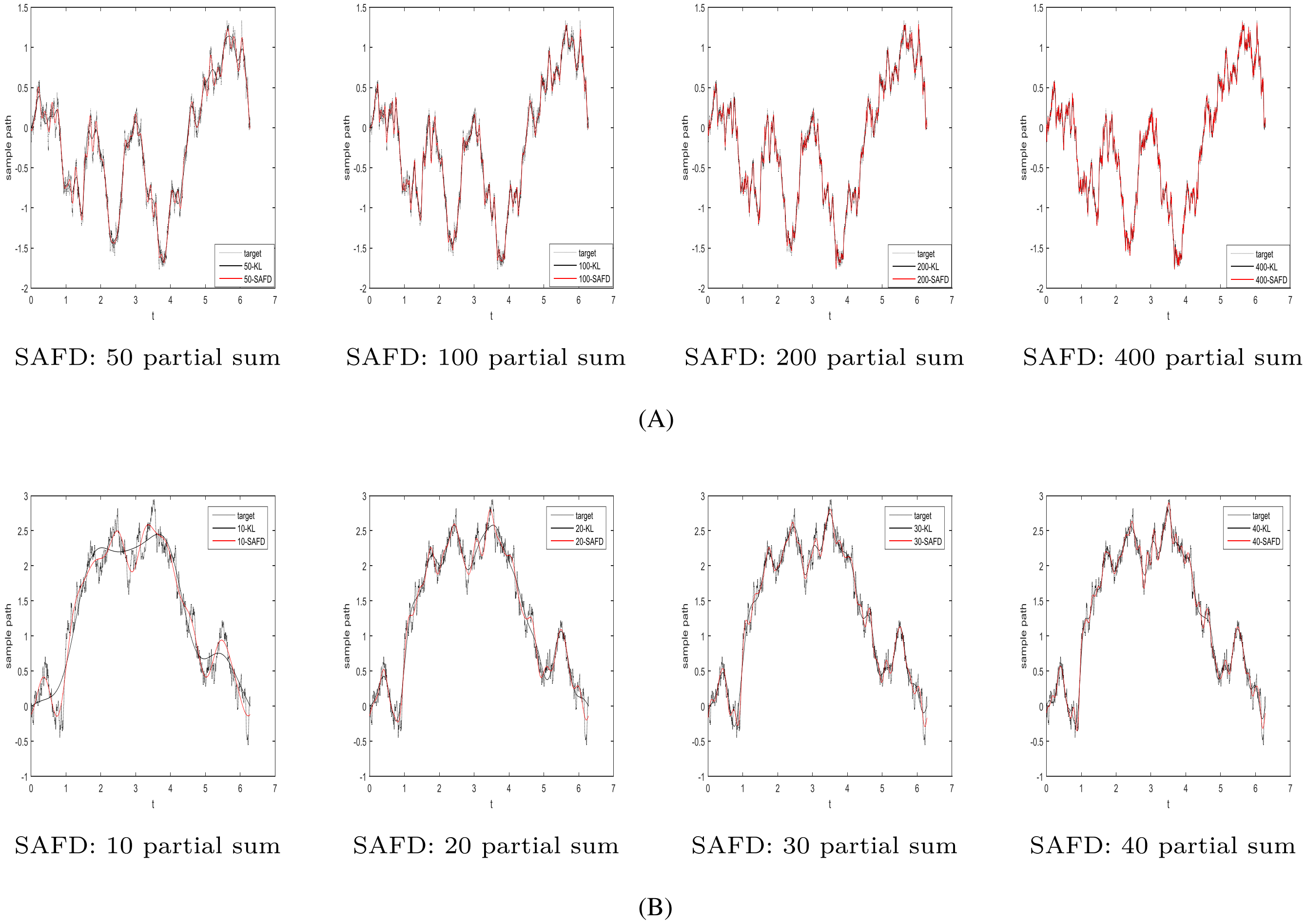


FIGURE 2 Sample path IIa (2^{12} points) of Brownian bridge. [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 2a Relative error.

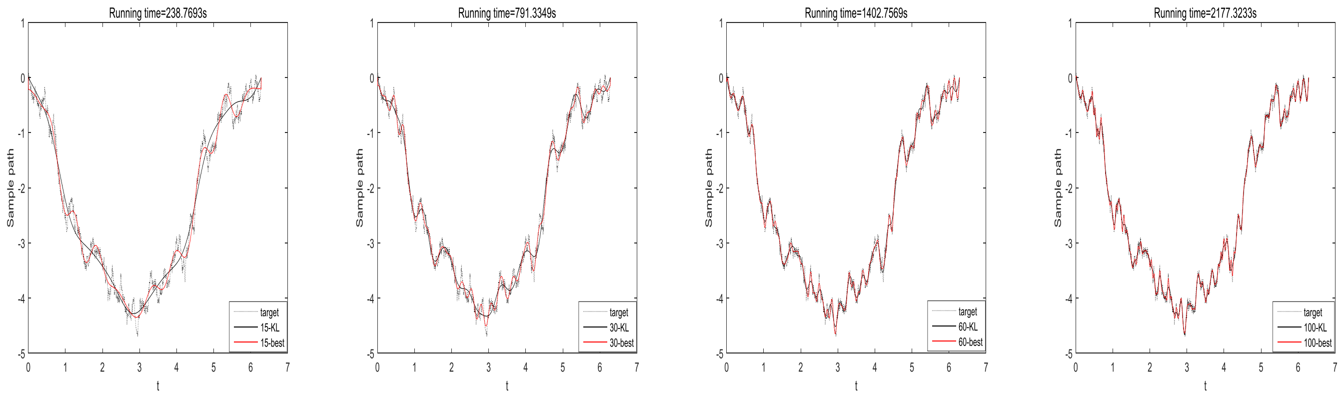
n partial sum	$n = 50$	$n = 100$	$n = 200$	$n = 400$
KL	0.0237	0.0118	0.0055	0.0026
SAFD	0.0119	0.0055	0.0026	0.0012

TABLE 2b Relative error.

n partial sum	$n = 10$	$n = 20$	$n = 30$	$n = 40$
KL	0.0245	0.0103	0.0074	0.0059
SAFD	0.0120	0.0061	0.0046	0.0035

with the uniform spacing $\Delta t \approx 0.002$ and $\Delta t \approx 0.006$, respectively. The results are shown in Figure 2A,B. SAFD has the convenience of using the TM system, that is, in the continuous formulation, orthonormal. Discretely, however, the orthonormal properties are with errors. Hence, SAFD requires more sampling points than SPOAFD. The relative errors are given in Tables 2 and 2b.

Example 3. (SnB on the Szegö kernels dictionary) The Brownian bridge is generated by using 2048 sampling points in $[0, 2\pi]$ with the uniform spacing $\Delta t \approx 0.003$. In this example, we approximate the sample path with the KL expansion and the SnB method. As shown in Figure 3 and Table 1, with all the 15, 30, 60, 100 partial sum approximations, the SnB method outperforms the KL method in the details. The relative errors are given in Table 3.



15-best 30-best 60-best 100-best

FIGURE 3 Sample path III (2^{11} points) of Brownian bridge. [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 3 Relative error.

n partial sum	$n = 15$	$n = 30$	$n = 60$	$n = 100$
KL	0.0068	0.0031	0.0015	8.2008×10^{-4}
SnB	0.0031	0.0015	6.9197×10^{-4}	3.8540×10^{-4}

5 | CONCLUSIONS

In the article, we propose several AFD-type methods, including SAFD, SPOAFD, and SnB, to expand random functions (time series, random processes, and random fields). They enjoy the same optimal convergence rate as that for KL expansion. Compared with KL, the proposed AFD methods require much less computation for they only use the covariance but not eigenpair information. There usually exists a large pool of dictionaries available for implementing an AFD-type method in the question. A suitably chosen dictionary may remarkably reduce the computation and lift up efficiency of the approximation. The proposed AFD-type expansions are well applicable also to infinite time and unbounded space domains, as long as the second moment is integrable there.

ACKNOWLEDGEMENTS

The authors wish to sincerely thank Professor Radu Victor Balan whose relevant questions inspired this study.

CONFLICT OF INTEREST STATEMENT

This work does not have any conflicts of interest.

ORCID

Tao Qian https://orcid.org/0000-0002-8780-9958

REFERENCES

1. T. Qian, *A sparse representation of random signals*, Math. Methods Appl. Sci. **45** (2022), no. 8, 4210–4230.
2. W. Qu, C. K. Chui, G. T. Deng, and T. Qian, *Sparse representation of approximation to identity*, Anal. Appl. **20** (2022), no. 04, 815–837.
3. T. Qian, *Reproducing kernel sparse representations in relation to operator equations*, Complex. Anal. Oper. Theory **14** (2020), no. 2, 1–15.
4. S. Saitoh and Y. Sawano, *Theory of reproducing kernels and applications*, Developments in Mathematics, Vol. **44**, Springer, Singapore, 2016.
5. F. Yang, M. Chen, J. Chen, Y. Zhang, and T. Qian, *Sparse representations of solutions to a class of random boundary value problems*. http://arxiv.org/abs/2112.04132
6. T. Qian and Y. B. Wang, *Adaptive Fourier series—a variation of greedy algorithm*, Adv. Comput. Math. **34** (2011), no. 3, 279–293.
7. V. N. Temlyakov, *Greedy approximation*, Cambridge University Press, Cambridge, 2011.
8. T. Qian, *Boundary derivative of phses of inner and outer functions and applications*, Math. Methods Appl. Sci. **32** (2009), 253–263.
9. T. Qian, Q.-H. Chen, and L.-Q. Li, *Analytic unit quadrature signals with non-linear phase*, Phys. D: Nonlinear Phenom. **303** (2005), no. 1–2, 80–87.

10. T. Qian, *Two-dimensional adaptive Fourier decomposition*, Math. Methods Appl. Sci. **39** (2016), no. 10, 2431–2448.
11. R. A. DeVore and V. N. Temlyakov, *Some remarks on greedy algorithms*, Adv. Comput. Math. **5** (1996), 173–187.
12. T. Qian, *Intrinsic mono-component decomposition of functions: an advance of Fourier theory*, Math. Methods Appl. Sci. **33** (2010), 880–891.
13. M. Nahon. (2000). *Phase evaluation and segmentation*, Ph.D. thesis, Yale University.
14. R. Coifman and S. Steinerberger, *Nonlinear phase unwinding of functions*, J. Fourier Anal. Appl. **23** (2017), no. 4, 778–809.
15. R. Coifman and J. Peyriére, *Phase unwinding, or invariant subspace decompositions of Hardy spaces*, J. Fourier Anal. Appl. **25** (2019), 684–695.
16. D. Alpay, F. Colombo, T. Qian, and I. Sabadini, *Adaptive orthonormal systems for matrix-valued functions*, Proc. Am. Math. Soc. **145** (2017), no. 5, 2089–2106.
17. D. Alpay, F. Colombo, T. Qian, and I. Sabadini, *Adaptive decomposition: the case of the Drury-Arveson space*, J. Fourier Anal. Appl. **23** (2017), no. 6, 1426–1444.
18. T. Qian, *Adaptive Fourier decomposition*, Science Press, Beijing, 2015. (in Chinese).
19. W. Qu and P. Dang, *Reproducing kernel approximation in weighted Bergman spaces: algorithm and applications*, Math. Methods Appl. Sci. **42** (2019), no. 12, 4292–4304.
20. Q. H. Chen, T. Qian, and L. H. Tan, A theory on non-constant frequency decompositions and applications, *Advancements in complex analysis*, Springer, Cham, 2020, pp. 1–37.
21. T. Qian, *A novel Fourier theory on non-linear phase and applications*, Adv. Math. (China) **47**, no. 3, 321–347.
22. T. Qian, W. Sproessig, and J. X. Wang, *Adaptive Fourier decomposition of functions in quaternionic Hardy spaces*, Math. Methods Appl. Sci. **35** (2012), no. 1, 43–64.
23. J. X. Wang and T. Qian, *Orthogonalization in Clifford Hilbert modules and applications*, 2021. arXiv preprint arXiv:2103.09416.
24. W. Qu and P. Dang, *Rational approximation in a class of weighted Hardy spaces*, Compl. Anal. Oper. Theory **13** (2019), no. 4, 1827–1852.
25. Y. B. Wang and T. Qian, *Pseudo-hyperbolic distance and n-best rational approximation in H^2 space*, Math. Methods Appl. Sci. **44** (2021), no. 11, 8497–8504.
26. T. Qian, *Cyclic AFD algorithm for best rational*, Math. Methods Appl. Sci. **37** (2014), no. 6, 846–859.
27. T. Qian, *n-best kernel approximation in reproducing kernel Hilbert spaces*. http://arxiv.org/abs/2201.07228
28. G. J. Lord, C. E. Powell, and T. Shardlow, *An introduction to computational stochastic PDEs*, Cambridge University Press, Cambridge, 2014.
29. W. Qu, T. Qian, I. T. Leongc, and P. T. Li, *The sparse representation of approximation to identity related with fractional heat equations*. arXiv:2207.10078v1 [math.NA].
30. W. Qu, T. Qian, and G. T. Deng, *A stochastic sparse representation: n-best approximation to random signals and computation*, Appl. Comput. Harmon. Anal. **55** (2021), 185–198.
31. M. Kanagawa, P. Hennig, D. Sejdinovic, and B. K. Sriperumbudur, *Gaussian processes and kernel methods: a review on connections and equivalence*, 2018. arXiv preprint:1807.02582.
32. I. Steinwart and C. Scovel, *Mercer's theorem on general domains: on the interaction between measures, kernels, and RKHSs*, Constr. Approx. **35** (2012), no. 3, 363–417.
33. I. Steinwart, *Convergence types and rates in generic Karhunen-Loéve expansions with applications to sample path properties*, Potential Anal. **51** (2019), 361–395. DOI 10.1007/s11118-018-9715-5
34. M. F. Dirscoll, *The reproducing kernel Hilbert space structure of the sample paths of a Gaussian process*, Z. Wahrscheinlichkeitstheorie und verwandte Gebiete **26** (1973), no. 4, 309–316.
35. N. Lukić and J. H. Beder, *Stochastic processes with sample paths in reproducing Hilbert spaces*, Trans. Am. Math. Soc. **353** (2001), no. 10, 729–748.
36. T. Qian, J. Z. Wang, and W. X. Mai, *An enhancement algorithm for cyclic adaptive Fourier decomposition*, Appl. Comput. Harmon. Anal. **47** (2019), no. 2, 516–525.
37. T. Qian, L. Dai, L. Zhang, and Z. Chen, *Granular sieving algorithm for selecting best n parameters*, Math. Methods Appl. Sci. **45** (2022), no. 12, 7495–7509.

How to cite this article: T. Qian, Y. Zhang, W. Liu, W. Qu, *Adaptive Fourier decomposition-type sparse representations versus the Karhunen–Loève expansion for decomposing stochastic processes*, Math. Meth. Appl. Sci. **46** (2023), 14007–14025. DOI 10.1002/mma.9301