Applied Mathematics and Computation 362 (2019) 124538

Contents lists available at ScienceDirect

Applied Mathematics and Computation

journal homepage: www.elsevier.com/locate/amc

Dimensionwise multivariate orthogonal polynomials in general probability spaces☆

Sharif Rahman

Applied Mathematical and Computational Sciences, The University of Iowa, Iowa City, Iowa 52242, USA

ARTICLE INFO

Keywords:
Uncertainty quantification
Multivariate orthogonal polynomials
ANOVA
Polynomial dimensional decomposition
Non-product-type probability measures

ABSTRACT

This paper puts forward a new generalized polynomial dimensional decomposition (PDD), referred to as GPDD, encompassing hierarchically organized, measure-consistent multivariate orthogonal polynomials in dependent random variables. In contrast to the existing PDD, which is valid strictly for independent random variables, no tensor-product structure or product-type probability measure is imposed or necessary. Fundamental mathematical properties of GPDD are examined by creating dimensionwise decomposition of polynomial spaces, deriving statistical properties of random orthogonal polynomials, demonstrating completeness of orthogonal polynomials for requisite assumptions, and upholding mean-square convergence to the correct limit. The GPDD approximation proposed should be effective in solving high-dimensional stochastic problems subject to dependent variables with a large class of non-product-type probability measures.

1. Introduction

Polynomial dimensional decomposition (PDD) represents a dimensionwise Fourier series expansion in random orthogonal polynomials [1,2]. Methods premised on PDD are often used to solve high-dimensional uncertainty quantification or stochastic problems in many fields of engineering [3–6]. The decomposition, introduced as the polynomial refinement of the analysis-of-variance (ANOVA) dimensional decomposition (ADD) [7–12], mitigates the curse of dimensionality to an appreciable magnitude by constructing a sequence of low-dimensional approximations of the input–output mapping. However, the existing PDD mandates that all input random variables be statistically independent, endowed with product-type probability measures. This enables forming multivariate orthogonal polynomials from a tensor product of univariate orthogonal polynomials. In practice, though, there are considerable correlation or dependence among input variables, negating the application of PDD and most stochastic methods available today. A common approach to tackle dependent variables is by invoking measure transformations [13], where dependent variables are mapped to independent variables and then use existing stochastic methods developed precisely for independent variables. Depending on the problem at hand, this may require stochastic analysis of a highly nonlinear function, likely vitiating approximation quality of stochastic solutions. Therefore, advancements above and beyond tensor-product structure and product-type probability measures, able to deal with dependent variables head-on, are most desirable.

This paper proposes a new generalized PDD, referred to as GPDD, to account for dependent, non-product-type probability measures of input random variables. A number of fundamental mathematical properties of GPDD are studied. Readers

☆ Grant sponsor: U.S. National Science Foundation; Grant No. CMMI-1462385.
E-mail address: sharif-rahman@uiowa.edu

https://doi.org/10.1016/j.amc.2019.06.052

interested in computational algorithm and applications should consult the companion paper [14]. The paper is structured as follows. Section 2 clarifies requisite assumptions on the input probability measure. Section 3 briefly reviews the generalized ADD for dependent variables, providing a vital link to the development of GPDD. A concise description of multivariate orthogonal polynomials consistent with a general, non-product-type probability measure, including derivation of their second-moment properties, is presented in Section 4. Appropriate polynomial spaces and their dimensionwise decompositions are explained. The validity of orthogonal basis and completeness of multivariate orthogonal polynomials have been demonstrated. Section 5 officially unveils GPDD, followed by a discussion on its convergence, approximation quality, and optimality. The relevance of GPDD for infinitely many input variables is also established. Section 6 presents a numerical example and discusses the potential effectiveness of GPDD. Finally, Section 7 draws conclusions.

2. Random input and its probability measure

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability triple consisting of Ω as a sample space, $\mathcal{F}$ as a σ-algebra on Ω, and $\mathbb{P}: \mathcal{F} \to [0,1]$ as a probability measure. For $N \in \mathbb{N} := \{1, 2, \ldots\}$, denote by $\mathbf{X} := (X_1, \ldots, X_N)^T : (\Omega, \mathcal{F}) \to (\mathbb{A}^N, \mathcal{B}^N)$ a measurable function, that is, $\mathbf{X}^{-1}(B) \in \mathcal{F}$ for every Borel set $B \in \mathcal{B}^N$. Here, $\mathbb{A}^N \subseteq \mathbb{R}^N$ represents a bounded or unbounded subdomain of $\mathbb{R}^N$. This function, referred to as $\mathbb{A}^N$-valued input random vector, describes all uncertain parameters of a stochastic problem and induces the probability measure $\mathbb{P}$ on the measurable space $(\Omega, \mathcal{F})$. In general, $\mathbf{X}$ comprises N dependent random variables, where N defines the dimension of the stochastic problem.

Let $F_{\mathbf{X}}(\mathbf{x}) := \mathbb{P}(\cap_{i=1}^N \{X_i \le x_i\})$ represent the joint cumulative distribution function of $\mathbf{X}$, which admits the joint probability density function $f_{\mathbf{X}}(\mathbf{x}) := \partial^N F_{\mathbf{X}}(\mathbf{x}) / \partial x_1 \cdots \partial x_N$. Associated with the abstract probability triple $(\Omega, \mathcal{F}, \mathbb{P})$, there exists the image probability triple $(\mathbb{A}^N, \mathcal{B}^N, f_{\mathbf{X}} d\mathbf{x})$, where $\mathbb{A}^N$ is the image of Ω from the mapping $\mathbf{X}: \Omega \to \mathbb{A}^N$. Statements and objects appropriate in one triple have obvious counterparts in the other triple. Both probability triples will be considered in this paper.

Assumption 1. The following list describes a set of requisite assumptions by GPDD on the input probability measure:

(1) The input random vector $\mathbf{X}$ is endowed with an absolutely continuous joint distribution function $F_{\mathbf{X}}(\mathbf{x})$. Its joint probability density function $f_{\mathbf{X}}(\mathbf{x})$ is continuous and has a bounded or unbounded support $\mathbb{A}^N \subseteq \mathbb{R}^N$.
(2) All absolute moments of $\mathbf{X}$ exist. That is, for all $\mathbf{j} := (j_1, \ldots, j_N) \in \mathbb{N}_0^N$, $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$,

$$\mathbb{E}\left[|\mathbf{X}^{\mathbf{j}}|\right] := \int_{\Omega} |\mathbf{X}(\omega)^{\mathbf{j}}| d\mathbb{P}(\omega) = \int_{\mathbb{A}^N} |\mathbf{x}^{\mathbf{j}}| f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} < \infty,$$

where $\mathbf{X}^{\mathbf{j}} = X_1^{j_1} \cdots X_N^{j_N}$ and $\mathbb{E}$ is the mathematical expectation operator.
(3) The density function $f_{\mathbf{X}}(\mathbf{x})$ fulfills, at least, one of the two following properties:
   (a) there exists a compact subset $\mathbb{A}^N \subset \mathbb{R}^N$ such that $\mathbb{P}(\mathbf{X} \in \mathbb{A}^N) = 1$;
   (b) there exists a real number $a > 0$ such that

$$\int_{\mathbb{A}^N} \exp\left(a\|\mathbf{x}\|\right) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} < \infty,$$

   where $\|\cdot\| : \mathbb{A}^N \to \mathbb{R}_0^+ := [0, +\infty)$ is an arbitrary norm.
(4) The density function $f_{\mathbf{X}}(\mathbf{x})$ has a grid-closed support.

The first three items of Assumption 1 are identical to those required by a recently developed generalized polynomial chaos expansion (PCE) or GPCE [15]. Therefore, for added explanations, readers should check prior work. Item (4) of Assumption 1 ensures that for any point $\mathbf{x} \in \mathrm{supp}(f_{\mathbf{X}})$, one can move in each coordinate direction and find another point $\mathbf{x}' \in \mathrm{supp}(f_{\mathbf{X}})$. To explain this further, consider a density function $f_{X_1X_2}(x_1, x_2)$ of a bivariate input random vector $(X_1, X_2)^T$. The support of $f_{X_1X_2}(x_1, x_2)$ is referred to as grid-closed if there is a point (x_1, x_2) in $\mathrm{supp}(f_{X_1X_2})$, there will always be another point in $\mathrm{supp}(f_{X_1X_2})$ with only one of the two variables being identical. Conversely, consider a density function defined on a straight line $x_1 = x_2$. Then the support is not grid-closed, because there does not exist another point with only one variable being identical. Therefore, Item (4) is needed to avoid such pathological density functions. For a technical definition of grid closure, consult Appendix B in Hooker's paper [16]. Assumption 1 is used throughout the paper.

3. A generalized ANOVA dimensional decomposition

Defined on the same probability triple $(\Omega, \mathcal{F}, \mathbb{P})$, let $y(\mathbf{X}) := y(X_1, \ldots, X_N)$ represent a real-valued output random variable of interest. A major objective of stochastic analysis, given the probability measure of random input $\mathbf{X}$, is to calculate the probabilistic characteristics of $y(\mathbf{X})$. Frequently, y is assumed to belong to a reasonably large function class, such as the Hilbert space

$$L^2(\Omega, \mathcal{F}, \mathbb{P}) := \left\{ y : \Omega \to \mathbb{R} : \int_{\Omega} y^2(\mathbf{X}(\omega)) d\mathbb{P}(\omega) < \infty \right\},$$

endowed with an inner product and associated norm, defined by

$$(y(\mathbf{X}), z(\mathbf{X}))_{L^2(\Omega,\mathcal{F},\mathbb{P})} := \int_\Omega y(\mathbf{X}(\omega)) z(\mathbf{X}(\omega)) d\mathbb{P}(\omega)$$

and

$$\|y(\mathbf{X})\|_{L^2(\Omega,\mathcal{F},\mathbb{P})} := \sqrt{(y(\mathbf{X}), y(\mathbf{X}))_{L^2(\Omega,\mathcal{F},\mathbb{P})}},$$

respectively. Similarly, for the image probability triple $(\mathbb{A}^N, \mathcal{B}^N, f_{\mathbf{X}} d\mathbf{x})$, there is an equivalent Hilbert space

$$L^2(\mathbb{A}^N, \mathcal{B}^N, f_{\mathbf{X}} d\mathbf{x}) := \left\{ y : \mathbb{A}^N \to \mathbb{R} : \int_{\mathbb{A}^N} y^2(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} < \infty \right\},$$

with the corresponding inner product and norm, defined by

$$(y(\mathbf{x}), z(\mathbf{x}))_{f_{\mathbf{X}} d\mathbf{x}} := \int_{\mathbb{A}^N} y(\mathbf{x}) z(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}$$

and

$$\|y(\mathbf{x})\|_{f_{\mathbf{X}} d\mathbf{x}} := \sqrt{(y(\mathbf{x}), y(\mathbf{x}))_{f_{\mathbf{X}} d\mathbf{x}}},$$

respectively. If $y(\mathbf{x}) \in L^2(\mathbb{A}^N, \mathcal{B}^N, f_{\mathbf{X}} d\mathbf{x})$, then it follows that $y(\mathbf{X}(\omega)) \in L^2(\Omega, \mathcal{F}, \mathbb{P})$.

3.1. Generalized ADD

For a stochastic problem involving $N \in \mathbb{N}$ variables, let $\{1, \ldots, N\}$ be an index set. Denote by $u \subseteq \{1, \ldots, N\}$ a subset of the index set, including the empty set $\emptyset$. The subset u has cardinality $0 \le |u| \le N$. Let $\mathbf{X}_u := (X_{i_1}, \ldots, X_{i_{|u|}})^T$, $1 \le i_1 < \cdots < i_{|u|} \le N$, be a subvector of $\mathbf{X}$, which is defined on the abstract probability triple $(\Omega^u, \mathcal{F}^u, \mathbb{P}^u)$, comprising the sample space Ω^u, σ-algebra $\mathcal{F}^u$ on Ω^u, and probability measure $\mathbb{P}^u$. For a complementary subset of u, defined by $-u := \{1, \ldots, N\} \backslash u$, there is a complementary subvector $\mathbf{X}_{-u} := \mathbf{X}_{\{1,\ldots,N\}\backslash u}$. The image probability triple of $\mathbf{X}_u$ is denoted by $(\mathbb{A}^u, \mathcal{B}^u, f_{\mathbf{X}_u} d\mathbf{x}_u)$. Here, $\mathbb{A}^u \subseteq \mathbb{R}^{|u|}$ is the image sample space, $\mathcal{B}^u$ is the Borel σ-algebra on $\mathbb{A}^u$, and $f_{\mathbf{X}_u}(\mathbf{x}_u) := \int_{\mathbb{A}^{-u}} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}_{-u}$ is the marginal probability density function of $\mathbf{X}_u$ with support $\mathbb{A}^u$.

From Item (4) of Assumption 1, fulfilled by common probability distributions, there exists, for a square-integrable function $y(\mathbf{X}) \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ of input variables $\mathbf{X}$, a unique, finite, hierarchical[1] expansion [16]

$$y(\mathbf{X}) = \sum_{u \subseteq \{1,\ldots,N\}} y_u(\mathbf{X}_u), \tag{1}$$

referred to as the generalized ADD [17], in terms of the component functions y_u, $u \subseteq \{1, \ldots, N\}$, of input variables with increasing dimensions. Here, $y_u(\mathbf{X}_u) = y_u(X_{i_1}, \ldots, X_{i_{|u|}})$ represents a $|u|$-variate component function of y, expressing a constant or a $|u|$-variate interactive effect of $\mathbf{X}_u := (X_{i_1}, \ldots, X_{i_{|u|}})$ on y when $|u| = 0$ or $|u| > 0$. Analogous to the classical ADD [7–12], there are 2^N component functions in the summation of (1), each of which depends on the group of variables indexed by a specific subset of $\{1, \ldots, N\}$, including the empty set $\emptyset$.

3.2. Component functions of generalized ADD

A simple way to link the component functions of ADD, be it classical or generalized, to the function y involves exploiting annihilating conditions. However, the original annihilating conditions [10–12], applicable to the classical ADD, are too strong for dependent random variables and hence not appropriate for the generalized ADD. Therefore, the conditions must be weakened to the degree possible under Item (4) of Assumption 1. Indeed, there exist such weak annihilating conditions, which mandate all non-constant component functions y_u of the generalized ADD to integrate to *zero* with respect to the marginal density function $f_{\mathbf{X}_u}(\mathbf{x}_u)$ of $\mathbf{X}_u$ in each coordinate direction of u, that is [16],

$$\int_{\mathbb{A}^{\{i\}}} y_u(\mathbf{x}_u) f_{\mathbf{X}_u}(\mathbf{x}_u) dx_i = 0 \text{ for } i \in u \neq \emptyset. \tag{2}$$

Compared with the original annihilating conditions, (2) represents a milder version, but it still produces two remarkable properties of the generalized ADD [17]:

(1) The means of all non-constant generalized ADD component functions y_u, $\emptyset \neq u \subseteq \{1, \ldots, N\}$, vanish, that is,

$$\mathbb{E}[y_u(\mathbf{X}_u)] = 0. \tag{3}$$

(2) Two different generalized ADD component functions $y_{u,G}$ and $y_{v,G}$, where $v \subset u$, are mutually orthogonal, that is,

$$\mathbb{E}[y_u(\mathbf{X}_u) y_v(\mathbf{X}_v)] = 0. \tag{4}$$

[1] The adjective *hierarchical* is used here in the context of dimensionwise hierarchy of input variables.

Applying the weak annihilating conditions (2) and the second-moment properties from (3) and (4), the master formulae for all component functions y_u, $u \subseteq \{1,\ldots,N\}$, of the generalized ADD are [17]

$$y_\emptyset = \int_{\mathbb{A}^N} y(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}, \tag{5a}$$

$$\begin{aligned} y_u(\mathbf{X}_u) = & \int_{\mathbb{A}^{-u}} y(\mathbf{X}_u, \mathbf{x}_{-u}) f_{\mathbf{X}_{-u}}(\mathbf{x}_{-u}) d\mathbf{x}_{-u} - \sum_{v \subset u} y_v(\mathbf{X}_v) \\ & - \sum_{\substack{\emptyset \ne v \subseteq \{1,\cdots,N\} \\ v \cap u \ne \emptyset, v \not\subseteq u}} \int_{\mathbb{A}^{v \cap -u}} y_v(\mathbf{X}_{v \cap u}, \mathbf{x}_{v \cap -u}) f_{\mathbf{X}_{v \cap -u}}(\mathbf{x}_{v \cap -u}) d\mathbf{x}_{v \cap -u}. \end{aligned} \tag{5b}$$

Although the original formulae were reported using $\mathbb{A}^N = \mathbb{R}^N$ [17], the generalization for the case of $\mathbb{A}^N \subseteq \mathbb{R}^N$ is straightforward. Here, $(\mathbf{X}_u, \mathbf{x}_{-u})$ stands for an N-dimensional vector whose ith component is X_i if $i \in u$ and x_i if $i \notin u$. When $u = \emptyset$, both summations in (5b) go away, producing the expression of the constant function $y_\emptyset$ in (5a). When $u = \{1,\ldots,N\}$, the integration in the first line of (5b) is performed on the empty set and the summation in the second line of (5b) disappears, forming (1) and hence determining the last function $y_{\{1,\ldots,N\}}$. In fact, all component functions of y can be generated from (5b).

From (5a) and (5b), two important observations stand out. First, the constant component function of ADD, whether classical or generalized, is the same as the expected value of $y(\mathbf{X})$. Second, in contrast to the classical ADD, the component functions of the generalized ADD, satisfying (5b), are coupled and must be solved simultaneously. In this case, for a given $\emptyset \ne u \subseteq \{1,\ldots,N\}$, the component function y_u depends not only on the component functions y_v, where $v \subset u$, but also on the component functions y_v, where $v \cap u \ne \emptyset$, $v \not\subseteq u$. For additional details of the generalized ADD, readers are directed to a prior work [17].

The generalized ADD discussed earlier can also be attained by splitting the Hilbert space

$$L^2(\mathbb{A}^N, \mathcal{B}^N, f_{\mathbf{X}} d\mathbf{x}) = \mathbf{1} \oplus \bigcup_{\emptyset \ne u \subseteq \{1,\ldots,N\}} \mathcal{W}_u \tag{6}$$

into a collection of generalized ADD subspaces

$$\mathcal{W}_u := \left\{ y_u \in L^2(\mathbb{A}^u, \mathcal{B}^u, f_{\mathbf{X}_u} d\mathbf{x}_u) : \int_{\mathbb{A}^{\{i\}}} y_u(\mathbf{x}_u) f_{\mathbf{X}_u}(\mathbf{x}_u) dx_i = 0 \text{ for } i \in u \ne \emptyset \right\}, \tag{7}$$

subsuming $|u|$-variate component functions of y. However, $\mathcal{W}_u$, $\emptyset \ne u \subseteq \{1,\ldots,N\}$, are all infinite-dimensional subspaces, indicating a need for their finite-dimensional discretizations. Indeed, by using measure-consistent multivariate orthogonal polynomial basis, to be formally introduced next, a component function $y_u(\mathbf{X}_u) \in \mathcal{W}_u$ can be expanded in terms of these basis functions. The end result is a polynomial variant of the generalized ADD, namely GPDD, which is the principal motivation for this work.

4. Measure-consistent multivariate orthogonal polynomials

Let $\mathbf{X}$ be an input random vector with a general probability measure $f_{\mathbf{X}}(\mathbf{x})d\mathbf{x}$ on $\mathbb{A}^N$, obeying Assumption 1. It is elementary to show that the same assumptions also hold for any subvector $\mathbf{X}_u := (X_{i_1},\ldots,X_{i_{|u|}})^T$ with the marginal probability measure $f_{\mathbf{X}_u}(\mathbf{x}_u)d\mathbf{x}_u$ on $\mathbb{A}^u$. When $\emptyset \ne u \subseteq \{1,\ldots,N\}$, denote by $\mathbf{j}_u := (j_{i_1},\ldots,j_{i_{|u|}}) \in \mathbb{N}_0^{|u|}$ a $|u|$-dimensional multi-index with total degree $|\mathbf{j}_u| := j_{i_1} + \cdots + j_{i_{|u|}}$, where $j_{i_p} \in \mathbb{N}_0$, $p = 1,\ldots,|u|$, is the pth component of $\mathbf{j}_u$. The same symbol $|\cdot|$ represents the degree of a multi-index and the cardinality of a set.

4.1. Orthogonal polynomials

Let

$$\Pi^u := \mathbb{R}[\mathbf{x}_u] = \mathbb{R}[x_{i_1},\ldots,x_{i_{|u|}}]$$

be the space of all real-valued polynomials in $\mathbf{x}_u$. Define, for any polynomial pair P_u, $Q_u \in \Pi^u$, $\emptyset \ne u \subseteq \{1,\ldots,N\}$, an inner product

$$(P_u, Q_u)_{f_{\mathbf{X}_u} d\mathbf{x}_u} := \int_{\mathbb{A}^u} P_u(\mathbf{x}_u) Q_u(\mathbf{x}_u) f_{\mathbf{X}_u}(\mathbf{x}_u) d\mathbf{x}_u = \mathbb{E}[P_u(\mathbf{X}_u) Q_u(\mathbf{X}_u)] \tag{8}$$

on Π^u in relation to the probability measure $f_{\mathbf{X}_u}(\mathbf{x}_u)d\mathbf{x}_u$ and the concomitant norm

$$\|P_u\|_{f_{\mathbf{X}_u} d\mathbf{x}_u} := \sqrt{(P_u, P_u)_{f_{\mathbf{X}_u} d\mathbf{x}_u}} = \left(\int_{\mathbb{A}^u} P_u^2(\mathbf{x}_u) f_{\mathbf{X}_u}(\mathbf{x}_u) d\mathbf{x}_u \right)^{1/2} = \sqrt{\mathbb{E}\left[P_u^2(\mathbf{X}_u)\right]}.$$

If $(P_u, Q_u)_{f_{\mathbf{X}_u} d\mathbf{x}_u} = 0$, then the polynomials $P_u \in \Pi^u$ and $Q_u \in \Pi^u$ are said to be mutually orthogonal in relation to the measure $f_{\mathbf{X}_u}(\mathbf{x}_u)d\mathbf{x}_u$. In addition, if

$$(P_u, Q_u)_{f_{\mathbf{X}_u} d\mathbf{x}_u} = 0 \ \forall Q_u \in \Pi^u \text{ with } \deg Q_u < \deg P_u, \tag{9}$$

then a polynomial $P_u \in \Pi^u$ is referred to as an orthogonal polynomial in relation to the probability measure $f_{\mathbf{X}_u}(\mathbf{x}_u)d\mathbf{x}_u$ [18].

In recognition of Assumption 1, all moments of $\mathbf{X}_u$ exist, meaning that the inner product in (8) is well-defined and positive-definite. Therefore, the norm $\|P_u\|_{f_{\mathbf{X}_u} d\mathbf{x}_u}$ is also positive for all $0 \neq P_u \in \Pi^u$. Then, an infinite set of multivariate orthogonal polynomials, denoted by $\{P_{u,\mathbf{j}_u}(\mathbf{x}_u) : \mathbf{j}_u \in \mathbb{N}_0^{|u|}\}$, $P_{u,\mathbf{0}} = 1$, $P_{u,\mathbf{j}_u} \neq 0$, exists. This set of multivariate orthogonal polynomials is consistent with the probability measure $f_{\mathbf{X}_u}(\mathbf{x}_u)d\mathbf{x}_u$, conforming, for any $\mathbf{j}_u, \mathbf{k}_u \in \mathbb{N}_0^{|u|}$, to

$$\left(P_{u,\mathbf{j}_u}, P_{u,\mathbf{k}_u}\right)_{f_{\mathbf{X}_u} d\mathbf{x}_u} = 0 \text{ whenever } |\mathbf{j}_u| \neq |\mathbf{k}_u|. \tag{10}$$

In (10), $P_{u,\mathbf{j}_u} \in \Pi^u$ has a total degree $|\mathbf{j}_u|$ and is an orthogonal polynomial satisfying (9). Note that $P_{u,\mathbf{j}_u}$ is orthogonal to all polynomials of different degrees, but it may not be so with respect to other orthogonal polynomials of the same degree.

For each non-negative integer $l \in \mathbb{N}_0$, consider the set of total degrees $\{\mathbf{j}_u \in \mathbb{N}_0^{|u|} : |\mathbf{j}_u| = l\}$, where its elements have been organized as $\mathbf{j}_u^{(1)}, \ldots, \mathbf{j}_u^{(K_{u,l})}$. Any monomial order can be selected. The cardinality of the aforementioned set is

$$K_{u,l} = \#\left\{\mathbf{j}_u \in \mathbb{N}_0^{|u|} : |\mathbf{j}_u| = l\right\} = \binom{|u| + l - 1}{l}.$$

Define two $K_{u,l}$-dimensional column vectors

$$\mathbf{x}_{u,l} = (\mathbf{x}_u^{\mathbf{j}_u^{(1)}}, \ldots, \mathbf{x}_u^{\mathbf{j}_u^{(K_{u,l})}})^T$$

and

$$\mathbf{P}_{u,l}(\mathbf{x}_u) := (P_{u,\mathbf{j}_u^{(1)}}(\mathbf{x}_u), \ldots, P_{u,\mathbf{j}_u^{(K_{u,l})}}(\mathbf{x}_u))^T,$$

comprising as elements the monomials $\mathbf{x}_u^{\mathbf{j}_u}$ for $|\mathbf{j}_u| = l$ and the polynomial sequence $\{P_{u,\mathbf{j}_u}(\mathbf{x}_u)\}_{|\mathbf{j}_u|=l}$, respectively. The elements of $\mathbf{x}_{u,l}$ and $\mathbf{P}_{u,l}(\mathbf{x}_u)$ are ordered in the same manner. Inspired by Dunkl and Xu [18], a formal definition of multivariate orthogonal polynomials follows.

Definition 2. Consistent with a given probability measure $f_{\mathbf{X}_u}(\mathbf{x}_u)d\mathbf{x}_u$ of $\mathbf{X}_u$, denote by $\{P_{u,\mathbf{j}_u}(\mathbf{x}_u) : |\mathbf{j}_u| = l, \mathbf{j}_u \in \mathbb{N}_0^{|u|}\}$, $P_{u,\mathbf{j}_u}(\mathbf{x}_u) \in \Pi_l^{|u|}$, a set of polynomials in $\mathbf{x}_u$ of degree $l \in \mathbb{N}_0$. Arranged according to a monomial order of choice, let $\mathbf{P}_{u,l}(\mathbf{x}_u)$ be the corresponding $K_{u,l}$-dimensional column vector of polynomials. The polynomial set or vector is called orthogonal with respect to the inner product $(\cdot,\cdot)_{f_{\mathbf{X}_u} d\mathbf{x}_u}$ or the probability measure $f_{\mathbf{X}_u}(\mathbf{x}_u)d\mathbf{x}_u$, if, for $l, r \in \mathbb{N}_0$,

$$\left(\mathbf{x}_{u,r}, \mathbf{P}_{u,l}^T(\mathbf{x}_u)\right)_{f_{\mathbf{X}_u} d\mathbf{x}_u} := \int_{\mathbb{A}^u} \mathbf{x}_{u,r}\mathbf{P}_{u,l}^T(\mathbf{x}_u) f_{\mathbf{X}_u}(\mathbf{x}_u)d\mathbf{x}_u =: \mathbb{E}\left[\mathbf{X}_{u,r}\mathbf{P}_{u,l}^T(\mathbf{X}_u)\right] = \mathbf{0}, \quad l > r, \tag{11}$$

where

$$\mathbf{S}_{u,l} := \left(\mathbf{x}_{u,l}, \mathbf{P}_{u,l}^T(\mathbf{x}_u)\right)_{f_{\mathbf{X}_u} d\mathbf{x}_u} := \int_{\mathbb{A}^u} \mathbf{x}_{u,l}\mathbf{P}_{u,l}^T(\mathbf{x}_u) f_{\mathbf{X}_u}(\mathbf{x}_u)d\mathbf{x}_u =: \mathbb{E}\left[\mathbf{X}_{u,l}\mathbf{P}_{u,l}^T(\mathbf{X}_u)\right] \tag{12}$$

is a $K_{u,l} \times K_{u,l}$ invertible matrix.

A concise expression of the aforementioned polynomial vector, employing the vector notation, reads

$$\mathbf{P}_{u,r}(\mathbf{x}_u) = \mathbf{H}_{u,r,r}\mathbf{x}_{u,r} + \mathbf{H}_{u,r,r-1}\mathbf{x}_{u,r-1} + \cdots + \mathbf{H}_{u,r,0}\mathbf{x}_{u,0}, \quad r \in \mathbb{N}_0.$$

Here, $\mathbf{H}_{u,r,r-k}$, $k = 0, 1, \ldots, r$, represent various $K_{u,r} \times K_{u,r-k}$ coefficient matrices. Subsequently, using the properties in (11) and (12), the inner product

$$(\mathbf{P}_{u,r}(\mathbf{x}_u), \mathbf{P}_{u,l}^T(\mathbf{x}_u))_{f_{\mathbf{X}_u} d\mathbf{x}_u} = \begin{cases} \mathbf{0}, & l > r, \\ \mathbf{H}_{u,l,l}\mathbf{S}_{u,l} & l = r. \end{cases}$$

Clearly, Definition 2 matches with the conventional statement on orthogonal polynomials satisfying (9). A notable example of classical multivariate orthogonal polynomials is multivariate Hermite polynomials, obtained consistent with the Gaussian probability measure [19,20]. For further details of multivariate orthogonal polynomials in relation to other measures, readers should review the works of Appell and de Fériet [21], Erdelyi [19], Krall and Sheffer [22], and Dunkl and Xu [18]. A few non-Gaussian measures, described by Gegenbauer density function on a unit ball or Dirichlet density function on a simplex, and their associated multivariate orthogonal polynomials obtained from Rodrigues-type formulae are discussed in the companion paper [14].

For arbitrary probability measures, the associated multivariate orthogonal polynomials must be determined by numerical methods, such as the well-known Gram-Schmidt orthogonalization process [23]. However, the process is known to be ill-conditioned, suggesting a need for more stable methods. Another matter in generating multivariate polynomials stems from the existence of different monomial orders, leading to different sequences of orthogonal polynomials. No one choice is necessarily better than the other. This is contrary to the case of univariate orthogonal polynomials, where the Gram-Schmidt orthogonalization process produces a single sequence of such polynomials. Finally, it should be underscored that for a general non-product-type density function, there is no tensor product structure, thereby building multivariate orthogonal polynomials from univariate orthogonal polynomials impossible.

After the multivariate orthogonal polynomials have been generated either analytically or numerically, they can be standardized, as follows.

Definition 3. A standardized multivariate orthogonal polynomial $\Psi_{u,\mathbf{j}_u}(\mathbf{x}_u)$, $\emptyset \neq u \subseteq \{1,\ldots,N\}$, $\mathbf{j}_u \in \mathbb{N}_0^{|u|}$, of degree $|\mathbf{j}_u| = j_{i_1} + \ldots + j_{i_{|u|}}$ that is consistent with the probability measure $f_{\mathbf{X}_u}(\mathbf{x}_u)d\mathbf{x}_u$ is defined as

$$\Psi_{u,\mathbf{j}_u}(\mathbf{x}_u) := \frac{P_{u,\mathbf{j}_u}(\mathbf{x}_u)}{\|P_{u,\mathbf{j}_u}\|_{f_{\mathbf{X}_u}d\mathbf{x}_u}} = \frac{P_{u,\mathbf{j}_u}(\mathbf{x}_u)}{\sqrt{\mathbb{E}[P^2_{u,\mathbf{j}_u}(\mathbf{X}_u)]}}. \tag{13}$$

The standardized version is not absolutely necessary. However, it yields a comparatively concise statement of GPDD, to be presented in Section 5.

4.2. Dimensionwise splitting of polynomial spaces

The original idea of dimensional decomposition of the polynomial spaces for independent random variables also applies for dependent random variables. However, there is one important difference: while for the independent variables the splitting leads to a complete orthogonal decomposition, only a partial orthogonal decomposition can be achieved for dependent random variables. This is due to the fundamental properties of multivariate orthogonal polynomials consistent with non-product-type probability measures.

Consider the polynomial space Π^u defined in Section 4.1. For any $\emptyset \neq u \subseteq \{1,\ldots,N\}$, restrict the component j_{i_p} associated with the i_pth variable, where $i_p \in u \subseteq \{1,\ldots,N\}$, $p = 1,\ldots,|u|$, and $|u| > 0$, to accept only positive integer values. Therefore, $\mathbf{j}_u := (j_{i_1},\ldots,j_{i_{|u|}}) \in \mathbb{N}^{|u|}$ – the multi-index of $P_{u,\mathbf{j}_u}(\mathbf{x}_u)$ – possesses a degree $|\mathbf{j}_u| = j_{i_1} + \cdots + j_{i|u|}$, which can vary from $|u|$ to ∞ as $j_{i_1} \neq \cdots \neq j_{i_{|u|}} \neq 0$.

For $\emptyset \neq u \subseteq \{1,\ldots,N\}$ and $\mathbf{j}_u \in \mathbb{N}^{|u|}$, a monomial in $\mathbf{x}_u := (x_{i_1},\ldots,x_{i_{|u}})$ is the product

$$\mathbf{x}_u^{\mathbf{j}_u} = x_{i_1}^{j_{i_1}} \ldots x_{i_{|u|}}^{j_{i_{|u|}}}$$

with total degree $|\mathbf{j}_u|$. The expression obtained from an arbitrary linear combination of $\mathbf{x}_u^{\mathbf{j}_u}$, where $|\mathbf{j}_u| = l$, $|u| \le l < \infty$, is a homogeneous polynomial in $\mathbf{x}_u$ of degree l. This leads to the definition of

$$\mathcal{Q}_l^u := \operatorname{span}\{\mathbf{x}_u^{\mathbf{j}_u} : |\mathbf{j}_u| = l,\ \mathbf{j}_u \in \mathbb{N}^{|u|}\},\ \ |u| \le l < \infty,$$

as the space of homogeneous polynomials in $\mathbf{x}_u$ of degree l, followed by the definition of

$$\Theta_m^u := \operatorname{span}\{\mathbf{x}_u^{\mathbf{j}_u} : |u| \le |\mathbf{j}_u| \le m,\ \mathbf{j}_u \in \mathbb{N}^{|u|}\},\ \ |u| \le m < \infty,$$

as the space of polynomials in $\mathbf{x}_u$ of degree at least $|u|$ and at most m. In both spaces, the individual degree of each variable is non-zero. It is easy to figure out the dimensions of these two vector spaces as

$$\dim \mathcal{Q}_l^u = \#\left\{\mathbf{j}_u \in \mathbb{N}^{|u|} : |\mathbf{j}_u| = l\right\} = \binom{l-1}{|u|-1}$$

and

$$\dim \Theta_m^u = \sum_{l=|u|}^{m} \dim \mathcal{Q}_l^u = \sum_{l=|u|}^{m} \binom{l-1}{|u|-1} = \binom{m}{|u|},$$

respectively.

Given the definitions of the vector spaces $\mathcal{Q}_l^u$ and Θ_m^u, the space of orthogonal polynomials can be defined as follows. At the beginning, set $\mathcal{Z}_{|u|}^u := \Theta_{|u|}^u$. Then, for each $|u|+1 \le l < \infty$, denote by $\mathcal{Z}_l^u \subset \Theta_l^u$ the vector space of orthogonal polynomials of degree exactly l that are orthogonal to all polynomials in Θ_{l-1}^u. Formally, the definition reads

$$\mathcal{Z}_l^u := \{P_u \in \Theta_l^u : (P_u, Q_u)_{f_{\mathbf{X}_u}d\mathbf{x}_u} = 0\ \forall\, Q_u \in \Theta_{l-1}^u\},\ \ |u|+1 \le l < \infty.$$

Assuming that the support of $f_{\mathbf{X}_u}(\mathbf{x}_u)$ has non-empty interior, the dimension of $\mathcal{Z}_l^u$ is

$$M_{u,l} := \dim \mathcal{Z}_l^u = \dim \mathcal{Q}_l^u = \binom{l-1}{|u|-1}.$$

There are many choices for the basis of $\mathcal{Z}_l^u$. A natural selection is the set

$$\{P_{u,\mathbf{j}_u}(\mathbf{x}_u) : |\mathbf{j}_u| = l, \mathbf{j}_u \in \mathbb{N}^{|u|}\} \subset \mathcal{Z}_l^u,$$

comprising $M_{u,l}$ number of basis functions. It will be proved in Section 4.3 that, indeed, such a selection forms a basis of $\mathcal{Z}_l^u$. Each basis function $P_{u,\mathbf{j}_u}(\mathbf{x}_u)$ is a multivariate orthogonal polynomial of degree $|\mathbf{j}_u|$, as described previously. Obviously,

$$\mathcal{Z}_l^u = \text{span}\{P_{u,\mathbf{j}_u}(\mathbf{x}_u) : |\mathbf{j}_u| = l, \mathbf{j}_u \in \mathbb{N}^{|u|}\}, \ \ |u| \le l < \infty.$$

From the definition, any two distinct polynomial subspaces $\mathcal{Z}_l^u$ and $\mathcal{Z}_{l'}^u$, where $l \neq l'$, are mutually orthogonal. As a result, Θ_m^u admits an orthogonal decomposition

$$\begin{aligned}\Theta_m^u &= \bigoplus_{l=|u|}^{m} \mathcal{Z}_l^u = \bigoplus_{l=|u|}^{m} \text{span}\{P_{u,\mathbf{j}_u}(\mathbf{x}_u) : |\mathbf{j}_u| = l, \mathbf{j}_u \in \mathbb{N}^{|u|}\} \\ &= \text{span}\{P_{u,\mathbf{j}_u}(\mathbf{x}_u) : |u| \le |\mathbf{j}_u| \le m, \mathbf{j}_u \in \mathbb{N}^{|u|}\},\end{aligned}$$

in which the symbol $\oplus$ denotes the orthogonal sum of vector spaces. Furthermore, this facilitates a dimensionwise splitting of

$$\begin{aligned}\Pi^u &= \mathbf{1} \oplus \bigcup_{\emptyset \neq v \subseteq u} \bigoplus_{l=|v|}^{\infty} \mathcal{Z}_l^v = \mathbf{1} \oplus \bigcup_{\emptyset \neq v \subseteq u} \bigoplus_{l=|v|}^{\infty} \text{span}\{P_{v,\mathbf{j}_v}(\mathbf{x}_v) : |\mathbf{j}_v| = l, \mathbf{j}_v \in \mathbb{N}^{|v|}\} \\ &= \mathbf{1} \oplus \bigcup_{\emptyset \neq v \subseteq u} \text{span}\{P_{v,\mathbf{j}_v}(\mathbf{x}_v) : \mathbf{j}_v \in \mathbb{N}^{|v|}\},\end{aligned} \tag{14}$$

where the constant subspace, defined as $\mathbf{1} := \text{span}\{1\}$, has to be supplemented because $\mathcal{Z}_l^v$ does not contain constant functions.

Let

$$\Pi^N := \mathbb{R}[\mathbf{x}] = \mathbb{R}[x_1, \dots, x_N]$$

be the space of all real polynomials in $\mathbf{x}$. Then, setting $u = \{1, \dots, N\}$ in (14) and interchanging v for u produces yet another dimensionwise splitting of

$$\begin{aligned}\Pi^N &= \mathbf{1} \oplus \bigcup_{\emptyset \neq u \subseteq \{1,\dots,N\}} \bigoplus_{l=|u|}^{\infty} \mathcal{Z}_l^u \\ &= \mathbf{1} \oplus \bigcup_{\emptyset \neq u \subseteq \{1,\dots,N\}} \bigoplus_{l=|u|}^{\infty} \text{span}\{P_{u,\mathbf{j}_u}(\mathbf{x}_u) : |\mathbf{j}_u| = l, \mathbf{j}_u \in \mathbb{N}^{|u|}\} \\ &= \mathbf{1} \oplus \bigcup_{\emptyset \neq u \subseteq \{1,\dots,N\}} \text{span}\{P_{u,\mathbf{j}_u}(\mathbf{x}_u) : \mathbf{j}_u \in \mathbb{N}^{|u|}\}.\end{aligned} \tag{15}$$

According to (15), the polynomial space Π^N has been decomposed into spaces of hierarchically ordered multivariate orthogonal polynomials in $\mathbf{x}_u$, $\emptyset \neq u \subseteq \{1, \dots, N\}$. Some of these spaces are orthogonal, but not all. However, for a product-type probability measure, the union operator in (15) becomes an orthogonal sum, further reducing to a complete orthogonal decomposition [2]. The complete orthogonal decomposition leads to PDD meant for independent random variables. For dependent variables and GPDD, though, no further reduction is possible for a general non-product-type probability measure.

4.3. Orthogonal polynomials as a basis

While the multivariate orthogonal polynomials under Assumption 1 can be generated for a given probability measure, questions arise whether are they are complete and they form a basis in the Hilbert space of interest. The following two propositions provide answers to these important questions.

Proposition 4. *Given $N \in \mathbb{N}$, let $\mathbf{X} := (X_1, \dots, X_N)^T$ be an N-dimensional random vector with multivariate probability density function $f_{\mathbf{X}}(\mathbf{x})$, which satisfies Assumption 1. For any $\emptyset \neq u \subseteq \{1, \dots, N\}$, denote by $\mathbf{X}_u := (X_{i_1}, \dots, X_{i_{|u|}})^T$, a subvector of $\mathbf{X}$, which has the probability measure $f_{\mathbf{X}_u}(\mathbf{x}_u)d\mathbf{x}_u$. Then the measure-consistent orthogonal polynomial set $\{P_{u,\mathbf{j}_u}(\mathbf{x}_u) : |\mathbf{j}_u| = l, \mathbf{j}_u \in \mathbb{N}^{|u|}\}$, $|u| \le l < \infty$, forms a basis of $\mathcal{Z}_l^u$.*

Proof. In reference to Items (1) and (2) of Assumption 1, there exist orthogonal polynomials with respect to the probability measure $f_{\mathbf{X}_u}(\mathbf{x}_u)d\mathbf{x}_u$. Let $\mathbf{a}_{u,l}^T = (a_{u,l}^{(1)}, \dots, a_{u,l}^{(K_{u,l})})$ be a row vector consisting of real-valued constants $a_{u,l}^{(i)} \in \mathbb{R}, \ i = 1, \dots, K_{u,l}$. First, set the equality condition

$$\mathbf{a}_{u,l}^T \mathbf{P}_{u,l}(\mathbf{x}_u) = 0,$$

where $\mathbf{P}_{u,l}(\mathbf{x}_u)$ is a $K_{u,l}$-dimensional column vector constructed from the elements of $\{P_{u,\mathbf{j}_u}(\mathbf{x}_u) : |\mathbf{j}_u| = l, \mathbf{j}_u \in \mathbb{N}_0^{|u|}\}$. Second, multiply both sides of the equality from the right by $\mathbf{x}_{u,l}^T$, then integrate with respect to the measure $f_{\mathbf{X}_u}(\mathbf{x}_u)d\mathbf{x}_u$ over $\mathbb{A}^u$, and finally take transpose to attain

$$\mathbf{S}_{u,l}\mathbf{a}_{u,l} = \mathbf{0}. \tag{16}$$

Since $\mathbf{S}_{u,l}$, as defined in (12), is a $K_{u,l} \times K_{u,l}$ invertible matrix, (16) gives $\mathbf{a}_{u,l} = \mathbf{0}$. This proves linear independence of the elements of $\mathbf{P}_{u,l}(\mathbf{x}_u)$ or $\{P_{u,\mathbf{j}_u}(\mathbf{x}_u) : |\mathbf{j}_u| = l, \mathbf{j}_u \in \mathbb{N}_0^{|u|}\}$. Obviously, the elements of the subset $\{P_{u,\mathbf{j}_u}(\mathbf{x}_u) : |\mathbf{j}_u| = l, \mathbf{j}_u \in \mathbb{N}^{|u|}\}$, excluding the elements associated with zero components of $j_{i_1}, \ldots, j_{i_{|u|}}$, are also linearly independent. Furthermore, the dimension $M_{u,l}$ of $\mathcal{Z}_l^u$ matches exactly the number of elements of the aforementioned subset. Therefore, the spanning set $\{P_{u,\mathbf{j}_u}(\mathbf{x}_u) : |\mathbf{j}_u| = l, \mathbf{j}_u \in \mathbb{N}^{|u|}\}$ constitutes a basis of $\mathcal{Z}_l^u$. □

Proposition 5. *Given the preamble of Proposition 4, the set of polynomials from the union-sum collection*

$$\mathbf{1} \oplus \bigcup_{\emptyset \neq u \subseteq \{1,\ldots,N\}} \bigoplus_{l=|u|}^{\infty} \operatorname{span}\{P_{u,\mathbf{j}_u}(\mathbf{x}_u) : |\mathbf{j}_u| = l, \mathbf{j}_u \in \mathbb{N}^{|u|}\} \tag{17}$$

is L^2-dense in the Hilbert space $L^2(\mathbb{A}^N, \mathcal{B}^N, f_{\mathbf{X}}d\mathbf{x})$. Furthermore, with the overline representing set closure,

$$L^2(\mathbb{A}^N, \mathcal{B}^N, f_{\mathbf{X}}d\mathbf{x}) = \overline{\mathbf{1} \oplus \bigcup_{\emptyset \neq u \subseteq \{1,\ldots,N\}} \bigoplus_{l=|u|}^{\infty} \mathcal{Z}_l^u}. \tag{18}$$

Proof. As Items (1) and (2) of Assumption 1 are fulfilled, there exist multivariate orthogonal polynomials $P_{u,\mathbf{j}_u}(\mathbf{x}_u)$ with respect to the probability measure $f_{\mathbf{X}_u}(\mathbf{x}_u)d\mathbf{x}_u$. In addition, since Item (3) of Assumption 1 is also satisfied, the polynomial space Π^N is L^2-dense in the Hilbert space $L^2(\mathbb{A}^N, \mathcal{B}^N, f_{\mathbf{X}}d\mathbf{x})$, as explained in Theorem 3.2.18 of Dunkl and Xu [18]. However, according to the dimensionwise decomposition, with Items (3) and (4) of Assumption 1 in mind, Π^N is equal to the set of polynomials from the union-sum collection in (17). Therefore, the union-sum collection in (17) is also L^2-dense in $L^2(\mathbb{A}^N, \mathcal{B}^N, f_{\mathbf{X}}d\mathbf{x})$. When the limit points of (17) are included, the result is (18). □

Propositions 4 and 5 presented here are general versions of those presented in a prior work [2]. So are their proofs, especially the proof of Proposition 4, which require a more general reasoning for dependent random variables.

4.4. Statistical properties of random multivariate polynomials

Instead of real deterministic variables $x_1, \ldots, x_N$, if the input random variables $X_1, \ldots, X_N$ are used as arguments, then the polynomials $P_{u,\mathbf{j}_u}(\mathbf{X}_u)$ and $\Psi_{u,\mathbf{j}_u}(\mathbf{X}_u)$ become random functions. Here, their second-moment properties are derived, to be used in the remainder of this paper.

Lemma 6. *Let $\mathbf{X} := (X_1, \ldots, X_N)^T : (\Omega, \mathcal{F}) \to (\mathbb{A}^N, \mathcal{B}^N)$, $N \in \mathbb{N}$, be an N-dimensional random vector with multivariate probability density function $f_{\mathbf{X}}(\mathbf{x})$, satisfying Assumption 1 and $\mathbf{X}_u := (X_{i_1}, \ldots, X_{i_{|u|}})^T : (\Omega^u, \mathcal{F}^u) \to (\mathbb{A}^u, \mathcal{B}^u)$, $\emptyset \neq u \subseteq \{1, \ldots, N\}$, be a subvector of $\mathbf{X}$. Consistent with the probability measure $f_{\mathbf{X}_u}(\mathbf{x}_u)d\mathbf{x}_u$, let $\{P_{u,\mathbf{j}_u}(\mathbf{x}_u) : \mathbf{j}_u \in \mathbb{N}^N\}$ be an infinite set of multivariate orthogonal polynomials. Then each polynomial of the set satisfies the weak annihilating conditions, that is,*

$$\int_{\mathbb{A}^{\{i\}}} P_{u,\mathbf{j}_u}(\mathbf{x}_u) f_{\mathbf{X}_u}(\mathbf{x}_u) dx_i = 0 \text{ for } i \in u \neq \emptyset,\ \mathbf{j}_u \in \mathbb{N}^{|u|}. \tag{19}$$

Proof. Let $y(\mathbf{x}) \in L^2(\mathbb{A}^N, \mathcal{B}^N, f_{\mathbf{X}}d\mathbf{x})$ be an arbitrary function, where the input probability measure satisfies Assumption 1. Therefore, as explained in Section 2, there exists a unique generalized ADD of $y(\mathbf{x})$ where the component functions $y_u(\mathbf{x}_u)$, $\emptyset \neq u \subseteq \{1, \ldots, N\}$, obey the weak annihilating conditions described in (2). Since y is a square-integrable function, so are the component functions of y, that is, $y_u(\mathbf{x}_u)$ is an element of $\mathcal{W}_u$ defined in (7). Furthermore, from similar considerations given to Propositions 4 and 5, one can prove that the set $\{P_{u,\mathbf{j}_u}(\mathbf{x}_u) : \mathbf{j}_u \in \mathbb{N}^N\}$ is a basis of $\mathcal{W}_u$. Consequently, there exists a Fourier-like series expansion of

$$y_u(\mathbf{x}_u) \sim \sum_{\mathbf{j}_u \in \mathbb{N}^{|u|}} \hat{C}_{u,\mathbf{j}_u} P_{u,\mathbf{j}_u}(\mathbf{x}_u) \tag{20}$$

with $\hat{C}_{u,\mathbf{j}_u}$ denoting the associated expansion coefficients. The coefficients depend on $y_u(\mathbf{x}_u)$, which, in turn, depends on y. Here, the symbol $\sim$ represents equality in a weaker sense, such as equality in mean-square, but not necessarily pointwise nor almost everywhere. From the standard Hilbert space theory, the infinite series on the right hand side of (20) converges in mean-square. Combining (2) and (20) and interchanging the integral and summation operators, which is admissible for the convergent sum, yields

$$\sum_{\mathbf{j}_u \in \mathbb{N}^{|u|}} \hat{C}_{u,\mathbf{j}_u} \int_{\mathbb{A}^{\{i\}}} P_{u,\mathbf{j}_u}(\mathbf{x}_u) f_{\mathbf{X}_u}(\mathbf{x}_u) dx_i = 0 \tag{21}$$

for $i \in u \neq \emptyset$ and $\mathbf{j}_u \in \mathbb{N}^{|u|}$. Since $y(\mathbf{x}) \in L^2(\mathbb{A}^N, \mathcal{B}^N, f_{\mathbf{X}} d\mathbf{x})$ is arbitrary, so are the component functions $y_u(\mathbf{x}_u) \in \mathcal{W}_u$ and the resultant coefficients $\hat{C}_{u,\mathbf{j}_u}$, yet the sum in (21) must vanish. This is only possible if the integral in (21) vanishes, resulting in (19). □

Proposition 7. *Let* $\mathbf{X} := (X_1, \ldots, X_N)^T : (\Omega, \mathcal{F}) \to (\mathbb{A}^N, \mathcal{B}^N)$, $N \in \mathbb{N}$, *be an N-dimensional random vector with multivariate probability density function* $f_{\mathbf{X}}(\mathbf{x})$, *satisfying Assumption 1. For* $\emptyset \neq u \subseteq \{1, \ldots, N\}$, $\emptyset \neq v \subseteq \{1, \ldots, N\}$, $\mathbf{j}_u \in \mathbb{N}^{|u|}$, *and* $\mathbf{k}_v \in \mathbb{N}^{|v|}$, *the second-moment properties of non-standardized orthogonal polynomials are as follows:*

$$\mathbb{E}\left[P_{u,\mathbf{j}_u}(\mathbf{X}_u)\right] = 0. \tag{22}$$

$$\mathbb{E}\left[P_{u,\mathbf{j}_u}(\mathbf{X}_u) P_{v,\mathbf{k}_v}(\mathbf{X}_v)\right] = \begin{cases} 0, & (u \subset v) \text{ or } (v \subset u), \forall\, \mathbf{j}_u, \mathbf{k}_v, \\ 0, & \forall\, u, v, |\mathbf{j}_u| \neq |\mathbf{k}_v|, \\ \displaystyle\int_{\mathbb{A}^u} P^2_{u,\mathbf{j}_u}(\mathbf{x}_u) f_{\mathbf{X}_u}(\mathbf{x}_u) d\mathbf{x}_u, & u = v, \mathbf{j}_u = \mathbf{k}_v, \\ \displaystyle\int_{\mathbb{A}^{u\cup v}} P_{u,\mathbf{j}_u}(\mathbf{x}_u) P_{v,\mathbf{k}_v}(\mathbf{x}_v) f_{\mathbf{X}_{u\cup v}}(\mathbf{x}_{u\cup v}) d\mathbf{x}_{u\cup v}, & \text{otherwise.} \end{cases} \tag{23}$$

Proof. In reference to Definition 2, set $r = 0$ and $1 \leq l < \infty$ for any $\emptyset \neq u \subseteq \{1, \ldots, N\}$. Then $\mathbf{x}_{u,0} = (1)^T = (1)$, so that (11) becomes

$$\left(1, \mathbf{P}^T_{u,l}(\mathbf{x}_u)\right)_{f_{\mathbf{X}_u} d\mathbf{x}_u} := \int_{\mathbb{A}^u} \mathbf{P}^T_{u,l}(\mathbf{x}_u) f_{\mathbf{X}_u}(\mathbf{x}_u) d\mathbf{x}_u = \mathbf{0}$$

for all $1 \leq l < \infty$. Since the vector $\mathbf{P}_{u,l}(\mathbf{X}_u)$ comprises as elements the orthogonal polynomials $P_{u,\mathbf{j}_u}(\mathbf{X}_u)$, $|\mathbf{j}_u| = l$, one obtains (22) for any $\emptyset \neq u \subseteq \{1, \ldots, N\}$ and $\mathbf{j}_u \in \mathbb{N}^{|u|}$.

For the first *zero* result of (23), consider two subsets $\emptyset \neq u, v \subseteq \{1, \ldots, N\}$, where $v \subset u$, $\mathbf{j}_u \in \mathbb{N}^{|u|}$, and $\mathbf{k}_v \in \mathbb{N}^{|v|}$. Obviously, $u = v \cup (u \setminus v)$. Let $i \in (u \setminus v) \subseteq u$. Then

$$\begin{aligned} \mathbb{E}\left[P_{u,\mathbf{j}_u}(\mathbf{X}_u) P_{v,\mathbf{k}_v}(\mathbf{X}_v)\right] &:= \int_{\mathbb{A}^N} P_{u,\mathbf{j}_u}(\mathbf{x}_u) P_{v,\mathbf{k}_v}(\mathbf{x}_v) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathbb{A}^u} P_{u,\mathbf{j}_u}(\mathbf{x}_u) P_{v,\mathbf{k}_v}(\mathbf{x}_v) f_{\mathbf{X}_u}(\mathbf{x}_u) d\mathbf{x}_u \\ &= \int_{\mathbb{A}^v} P_{v,\mathbf{k}_v}(\mathbf{x}_v) \int_{\mathbb{A}^{u\setminus v}} P_{u,\mathbf{j}_u}(\mathbf{x}_u) f_{\mathbf{X}_u}(\mathbf{x}_u) d\mathbf{x}_{u\setminus v} d\mathbf{x}_v \\ &= \int_{\mathbb{A}^v} P_{v,\mathbf{k}_v}(\mathbf{x}_v) \int_{\mathbb{A}^{(u\setminus v)\setminus\{i\}}} \int_{\mathbb{A}^{\{i\}}} P_{u,\mathbf{j}_u}(\mathbf{x}_u) f_{\mathbf{X}_u}(\mathbf{x}_u) dx_i \prod_{\substack{j\in(u\setminus v) \\ j\neq i}} dx_j d\mathbf{x}_v \\ &= 0, \end{aligned}$$

where the equality to *zero* in the last line results from the innermost integral vanishing as per (19) in Lemma 6. Interchanging u and v obtains the complete result.

To derive the second *zero* result of (23), recognize that the polynomials $P_{u,\mathbf{j}_u}(\mathbf{x}_u)$ and $P_{v,\mathbf{k}_v}(\mathbf{x}_v)$ are members of $\mathcal{Z}^u_{|\mathbf{j}_u|}$ and $\mathcal{Z}^v_{|\mathbf{k}_v|}$, respectively. Therefore, they can both be expanded in terms of orthogonal polynomials $P_{\mathbf{j}}(\mathbf{x})$ in $\mathbf{x}$, for instance,

$$P_{u,\mathbf{j}_u}(\mathbf{x}_u) = \sum_{|\mathbf{j}|=|\mathbf{j}_u|} C_{u,\mathbf{j}} P_{\mathbf{j}}(\mathbf{x}), \quad P_{v,\mathbf{k}_v}(\mathbf{x}_v) = \sum_{|\mathbf{k}|=|\mathbf{k}_v|} C_{v,\mathbf{k}} P_{\mathbf{k}}(\mathbf{x}),$$

with $C_{u,\mathbf{j}}$ and $C_{v,\mathbf{k}}$ denoting the associated expansion coefficients. Then, for any u, v, and $|\mathbf{j}_u| \neq |\mathbf{k}_v|$,

$$\begin{aligned} \mathbb{E}\left[P_{u,\mathbf{j}_u}(\mathbf{X}_u) P_{v,\mathbf{k}_v}(\mathbf{X}_v)\right] &:= \int_{\mathbb{A}^N} P_{u,\mathbf{j}_u}(\mathbf{x}_u) P_{v,\mathbf{k}_v}(\mathbf{x}_v) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathbb{A}^N} \sum_{|\mathbf{j}|=|\mathbf{j}_u|} \sum_{|\mathbf{k}|=|\mathbf{k}_v|} C_{u,\mathbf{j}} C_{v,\mathbf{k}} P_{\mathbf{j}}(\mathbf{x}) P_{\mathbf{k}}(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \\ &= \sum_{|\mathbf{j}|=|\mathbf{j}_u|} \sum_{|\mathbf{k}|=|\mathbf{k}_v|} C_{u,\mathbf{j}} C_{v,\mathbf{k}} \int_{\mathbb{A}^N} P_{\mathbf{j}}(\mathbf{x}) P_{\mathbf{k}}(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \\ &= 0, \end{aligned}$$

where the equality to *zero* in the last line stems from the vanishing integral according to (10) with $u = \{1, \ldots, N\}$.

Finally, the two non-zero expressions of (23) are rooted in their respective definitions. For a general non-product-type probability measure, it is not possible to reduce them any more. □

Corollary 8. *For $\emptyset \neq u \subseteq \{1,\ldots,N\}$, $\emptyset \neq v \subseteq \{1,\ldots,N\}$, $\mathbf{j}_u \in \mathbb{N}^{|u|}$, and $\mathbf{k}_v \in \mathbb{N}^{|v|}$, the second-moment properties of standardized orthogonal polynomials are as follows:*

$$\mathbb{E}\left[\Psi_{u,\mathbf{j}_u}(\mathbf{X}_u)\right] = 0.$$

$$\mathbb{E}\left[\Psi_{u,\mathbf{j}_u}(\mathbf{X}_u)\Psi_{v,\mathbf{k}_v}(\mathbf{X}_v)\right] = \begin{cases} 0, & (u \subset v) \text{ or } (v \subset u), \forall\, \mathbf{j}_u, \mathbf{k}_v, \\ 0, & \forall\, u, v, |\mathbf{j}_u| \neq |\mathbf{k}_v|, \\ 1, & u = v, \mathbf{j}_u = \mathbf{k}_v, \\ \displaystyle\int_{\mathbb{A}^{u\cup v}} \Psi_{u,\mathbf{j}_u}(\mathbf{x}_u)\Psi_{v,\mathbf{k}_v}(\mathbf{x}_v) f_{\mathbf{X}_{u\cup v}}(\mathbf{x}_{u\cup v})d\mathbf{x}_{u\cup v}, & \text{otherwise}. \end{cases}$$

Corollary 9. *Let $\mathbf{X} = (X_1,\ldots,X_N)^T$ be a vector of independent, but not necessarily identical, input random variables, which satisfies Assumption 1. For $i = 1,\ldots,N$, designate by $f_{X_i}(x_i)$ the marginal probability density function of X_i and by $\Psi_{\{i\},j_i}(x_i)$ the j_ith-degree univariate orthonormal polynomial in x_i, where the latter is generated consistent with the probability measure $f_{X_i}(x_i)dx_i$. Then the multivariate orthonormal polynomial in $\mathbf{x}_u = (x_{i_1},\ldots,x_{i_{|u|}})$ of degree $|\mathbf{j}_u| = j_{i_1} + \cdots + j_{i_{|u|}}$, $\emptyset \neq u \subseteq \{1,\ldots,N\}$, $\mathbf{j}_u \in \mathbb{N}^{|u|}$, is constructed from the tensor product structure as*

$$\Psi_{u,\mathbf{j}_u}(\mathbf{x}_u) = \prod_{i\in u} \Psi_{\{i\},j_i}(x_i) = \prod_{p=1}^{|u|} \Psi_{\{i_p\},j_{i_p}}(x_{i_p}).$$

Moreover, the second-moment properties of such multivariate orthonormal polynomials, where $\emptyset \neq u, v \subseteq \{1,\ldots,N\}$, $\mathbf{j}_u \in \mathbb{N}^{|u|}$, and $\mathbf{k}_v \in \mathbb{N}^{|v|}$, are as follows:

$$\mathbb{E}\left[\Psi_{u,\mathbf{j}_u}(\mathbf{X}_u)\right] = 0.$$

$$\mathbb{E}\left[\Psi_{u,\mathbf{j}_u}(\mathbf{X}_u)\Psi_{v,\mathbf{k}_v}(\mathbf{X}_v)\right] = \begin{cases} 1, & u = v,\ \mathbf{j}_u = \mathbf{k}_v, \\ 0, & \text{otherwise}. \end{cases}$$

According to Corollary 9, the statistical properties of random orthonormal polynomials for independent random variables, especially the covariances between $\Psi_{u,\mathbf{j}_u}(\mathbf{X}_u)$ and $\Psi_{v,\mathbf{k}_v}(\mathbf{X}_v)$, simplify greatly. These simplified results, albeit applicable to existing PDD [2], are invalid for dependent variables. In contrast, Corollary 8 is a general result, is meant for dependent variables, and will be needed in the next section.

5. Generalized polynomial dimensional decomposition

The GPDD of a random variable $y(\mathbf{X}) \in L^2(\Omega,\mathcal{F},\mathbb{P})$ is an infinite series expansion of $y(\mathbf{X})$ with respect to a complete, hierarchically ordered, orthogonal polynomial basis of $L^2(\Omega,\mathcal{F},\mathbb{P})$. A preliminary, less sophisticated version of GPDD was mentioned in a prior work by the author [17]. Here, a more rigorous treatment of GPDD entailing dimensionwise splitting of polynomial spaces and functional analysis is presented. This version is new, has not been published elsewhere, and is, therefore, formally presented here as Theorem 10.

Theorem 10. *For a stochastic problem, let $\mathbf{X} := (X_1,\ldots,X_N)^T$ be a vector of $N \in \mathbb{N}$ input random variables with known probability measure $f_{\mathbf{X}}(\mathbf{x})d\mathbf{x}$, which satisfies Assumption 1. Suppose, for $\emptyset \neq u \subseteq \{1,\ldots,N\}$ and $\mathbf{X}_u := (X_{i_1},\ldots,X_{i_{|u|}})^T$, that a set of standardized multivariate orthogonal polynomials $\{\Psi_{u,\mathbf{j}_u}(\mathbf{x}_u) : \mathbf{j}_u \in \mathbb{N}^{|u|}\}$ has been obtained, consistent with the probability measure $f_{\mathbf{X}_u}(\mathbf{x})d\mathbf{x}_u$ of $\mathbf{X}_u$. Then*

(1) *there exists a Fourier-like series expansion in multivariate orthogonal polynomials in $\mathbf{X}_u$ for any square-integrable random variable $y(\mathbf{X}) \in L^2(\Omega,\mathcal{F},\mathbb{P})$, designated as the GPDD of*

$$\begin{aligned} y(\mathbf{X}) &\sim y_\emptyset + \sum_{\emptyset\neq u\subseteq\{1,\ldots,N\}} \sum_{l=|u|}^{\infty} \sum_{\substack{\mathbf{j}_u\in\mathbb{N}^{|u|} \\ |\mathbf{j}_u|=l}} C_{u,\mathbf{j}_u}\Psi_{u,\mathbf{j}_u}(\mathbf{X}_u) \\ &= y_\emptyset + \sum_{\emptyset\neq u\subseteq\{1,\ldots,N\}} \sum_{\mathbf{j}_u\in\mathbb{N}^{|u|}} C_{u,\mathbf{j}_u}\Psi_{u,\mathbf{j}_u}(\mathbf{X}_u). \end{aligned} \tag{24}$$

The expansion comprises a real-valued, zero-variate Fourier coefficient

$$y_\emptyset := \mathbb{E}[y(\mathbf{X})] := \int_{\mathbb{A}^N} y(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x})d\mathbf{x} \tag{25}$$

and an infinite number of real-valued, $|u|$-variate Fourier coefficients $C_{u,\mathbf{j}_u}$, which satisfy the infinite-dimensional linear system

$$\sum_{\emptyset\neq v\subseteq\{1,\ldots,N\}} \sum_{\mathbf{k}_v\in\mathbb{N}^{|v|}} C_{v,\mathbf{k}_v} J_{u,\mathbf{j}_u;v,\mathbf{k}_v} = I_{u,\mathbf{j}_u},\ \emptyset \neq u \subseteq \{1,\ldots,N\},\ \mathbf{j}_u \in \mathbb{N}^{|u|}, \tag{26}$$

with the integrals defined as

$$I_{u,\mathbf{j}_u} := \mathbb{E}\left[y(\mathbf{X})\Psi_{u,\mathbf{j}_u}(\mathbf{X}_u)\right] := \int_{\mathbb{A}^N} y(\mathbf{x})\Psi_{u,\mathbf{j}_u}(\mathbf{x}_u) f_{\mathbf{X}}(\mathbf{x})d\mathbf{x}, \tag{27a}$$

$$J_{u,\mathbf{j}_u;v,\mathbf{k}_v} := \mathbb{E}\left[\Psi_{u,\mathbf{j}_u}(\mathbf{X}_v)\Psi_{v,\mathbf{k}_v}(\mathbf{X}_v)\right] := \int_{\mathbb{A}^N} \Psi_{u,\mathbf{j}_u}(\mathbf{x}_u)\Psi_{v,\mathbf{k}_v}(\mathbf{x}_v) f_{\mathbf{X}}(\mathbf{x})d\mathbf{x}; \tag{27b}$$

and

(2) *the GPDD of* $y(\mathbf{X}) \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ *is mean-square convergent to* $y(\mathbf{X})$*, that is, for*

$$y_{S,m}(\mathbf{X}) := y_\emptyset + \sum_{\substack{\emptyset \neq u \subseteq \{1,\ldots,N\} \\ 1 \le |u| \le S}} \sum_{\substack{\mathbf{j}_u \in \mathbb{N}^{|u|} \\ |u| \le |\mathbf{j}_u| \le m}} C_{u,\mathbf{j}_u}\Psi_{u,\mathbf{j}_u}(\mathbf{X}_u),$$

where $1 \le S \le N$ *and* $|u| \le m < \infty$ *are integers,*

$$\lim_{S\to N,\ m\to\infty} \mathbb{E}\left[y_{S,m}^2(\mathbf{X})\right] = \mathbb{E}\left[y^2(\mathbf{X})\right];$$

is convergent in probability, that is, for any $\epsilon > 0$,

$$\lim_{S\to N,\ m\to\infty} \mathbb{P}(|y_{S,m}(\mathbf{X}) - y(\mathbf{X})| > \epsilon) = 0;$$

and is convergent in distribution, that is, for any $\xi \in \mathbb{R}$,

$$\lim_{S\to N,\ m\to\infty} F_{S,m}(\xi) = F(\xi)$$

such that $F_{S,m}(\xi) := \mathbb{P}(y_{S,m}(\mathbf{X}) \le \xi)$ *and* $F(\xi) := \mathbb{P}(y(\mathbf{X}) \le \xi)$ *are continuous distribution functions of* $y_{S,m}(\mathbf{X})$ *and* $y(\mathbf{X})$*, respectively.*

Proof. In reference to Assumption 1, there exists a complete infinite set of multivariate orthogonal polynomials in $\mathbf{x}_u$ that is consistent with the probability measure $f_{\mathbf{X}_u}(\mathbf{x}_u)d\mathbf{x}_u$. Invoking Proposition 5, with the cognizance that standardization is simply scaling, the set of standardized orthogonal polynomials from the union-sum collection

$$\mathbf{1} \oplus \bigcup_{\emptyset \neq u \subseteq \{1,\ldots,N\}} \bigoplus_{l=|u|}^{\infty} \mathrm{span}\{\Psi_{u,\mathbf{j}_u}(\mathbf{x}_u) : |\mathbf{j}_u| = l, \mathbf{j}_u \in \mathbb{N}^{|u|}\} = \Pi^N$$

is also L^2-dense in $L^2(\mathbb{A}^N, \mathcal{B}^N, f_{\mathbf{X}}d\mathbf{x})$. Equivalently, the set of random orthogonal polynomials

$$\mathbf{1} \oplus \bigcup_{\emptyset \neq u \subseteq \{1,\ldots,N\}} \bigoplus_{l=|u|}^{\infty} \mathrm{span}\{\Psi_{u,\mathbf{j}_u}(\mathbf{X}_u) : |\mathbf{j}_u| = l, \mathbf{j}_u \in \mathbb{N}^{|u|}\} \tag{28}$$

is L^2-dense in $L^2(\Omega, \mathcal{F}, \mathbb{P})$ as well. Therefore, the expansion in (24) readily applies to any random variable $y(\mathbf{X}) \in L^2(\Omega, \mathcal{F}, \mathbb{P})$. Merging the two inner sums in the first line of (24) produces the equality in the second line of (24).

From the denseness, every element of $L^2(\Omega, \mathcal{F}, \mathbb{P})$ is a limit point of the set of random orthogonal polynomials in (28). Therefore, the infinite series in (24) converges in mean-square to the correct limit, that is,

$$\mathbb{E}\left[\left\{y_\emptyset + \sum_{\emptyset \neq u \subseteq \{1,\ldots,N\}} \sum_{\mathbf{j}_u \in \mathbb{N}^{|u|}} C_{u,\mathbf{j}_u}\Psi_{u,\mathbf{j}_u}(\mathbf{X}_u)\right\}^2\right] = \mathbb{E}\left[y^2(\mathbf{X})\right],$$

which is similar to the Parseval identity [24] for a multivariate orthonormal system.

As the mean-square convergence is stronger than the convergence in probability or the convergence in distribution, the latter modes of convergence follow naturally.

To determine the expressions of the Fourier coefficients, define by

$$e_{\mathrm{GPDD}} := \mathbb{E}\left[\left\{y(\mathbf{X}) - y_\emptyset - \sum_{\emptyset \neq v \subseteq \{1,\ldots,N\}} \sum_{\mathbf{k}_v \in \mathbb{N}^{|v|}} C_{v,\mathbf{k}_v}\Psi_{v,\mathbf{k}_v}(\mathbf{X}_v)\right\}^2\right] \tag{29}$$

a second moment error of interest. Perform differentiation on both sides of (29) with respect to $y_\emptyset$ and $C_{u,\mathbf{j}_u}$, $\emptyset \neq u \subseteq \{1,\ldots,N\}$, $\mathbf{j}_u \in \mathbb{N}^{|u|}$, yielding

$$\begin{aligned}
\frac{\partial e_{\mathrm{GPDD}}}{\partial y_\emptyset} &= \frac{\partial}{\partial y_\emptyset}\mathbb{E}\left[\left\{y(\mathbf{X}) - y_\emptyset - \sum_{\emptyset \neq v \subseteq \{1,\ldots,N\}} \sum_{\mathbf{k}_v \in \mathbb{N}^{|v|}} C_{v,\mathbf{k}_v}\Psi_{v,\mathbf{k}_v}(\mathbf{X}_v)\right\}^2\right] \\
&= \mathbb{E}\left[\frac{\partial}{\partial y_\emptyset}\left\{y(\mathbf{X}) - y_\emptyset - \sum_{\emptyset \neq v \subseteq \{1,\ldots,N\}} \sum_{\mathbf{k}_v \in \mathbb{N}^{|v|}} C_{v,\mathbf{k}_v}\Psi_{v,\mathbf{k}_v}(\mathbf{X}_v)\right\}^2\right]
\end{aligned}$$

$$
\begin{aligned}
&= 2\mathbb{E}\left[\left\{y_\emptyset + \sum_{\emptyset\neq v\subseteq\{1,\ldots,N\}}\sum_{\mathbf{k}_v\in\mathbb{N}^{|v|}} C_{v,\mathbf{k}_v}\Psi_{v,\mathbf{k}_v}(\mathbf{X}_v) - y(\mathbf{X})\right\}\times 1\right] \\
&= 2\{y_\emptyset - \mathbb{E}[y(\mathbf{X})]\}
\end{aligned} \tag{30}
$$

and

$$
\begin{aligned}
\frac{\partial e_{\text{GPDD}}}{\partial C_{u,\mathbf{j}_u}} &= \frac{\partial}{\partial C_{u,\mathbf{j}_u}}\mathbb{E}\left[\left\{y(\mathbf{X}) - y_\emptyset - \sum_{\emptyset\neq v\subseteq\{1,\ldots,N\}}\sum_{\mathbf{k}_v\in\mathbb{N}^{|v|}} C_{v,\mathbf{k}_v}\Psi_{v,\mathbf{k}_v}(\mathbf{X}_v)\right\}^2\right] \\
&= \mathbb{E}\left[\frac{\partial}{\partial C_{u,\mathbf{j}_u}}\left\{y(\mathbf{X}) - y_\emptyset - \sum_{\emptyset\neq v\subseteq\{1,\ldots,N\}}\sum_{\mathbf{k}_v\in\mathbb{N}^{|v|}} C_{v,\mathbf{k}_v}\Psi_{v,\mathbf{k}_v}(\mathbf{X}_v)\right\}^2\right] \\
&= 2\mathbb{E}\left[\left\{y_\emptyset + \sum_{\emptyset\neq v\subseteq\{1,\ldots,N\}}\sum_{\mathbf{k}_v\in\mathbb{N}^{|v|}} C_{v,\mathbf{k}_v}\Psi_{v,\mathbf{k}_v}(\mathbf{X}_v) - y(\mathbf{X})\right\}\Psi_{u,\mathbf{j}_u}(\mathbf{X}_u)\right] \\
&= 2\left\{\sum_{\emptyset\neq v\subseteq\{1,\ldots,N\}}\sum_{\mathbf{k}_v\in\mathbb{N}^{|v|}} C_{v,\mathbf{k}_v}\mathbb{E}\left[\Psi_{u,\mathbf{j}_u}(\mathbf{X}_u)\Psi_{v,\mathbf{k}_v}(\mathbf{X}_v)\right] - \mathbb{E}\left[y(\mathbf{X})\Psi_{u,\mathbf{j}_u}(\mathbf{X}_u)\right]\right\} \\
&= 2\left\{\sum_{\emptyset\neq v\subseteq\{1,\ldots,N\}}\sum_{\mathbf{k}_v\in\mathbb{N}^{|v|}} C_{v,\mathbf{k}_v}J_{u,\mathbf{j}_u;v,\mathbf{k}_v} - I_{u,\mathbf{j}_u}\right\}.
\end{aligned} \tag{31}
$$

In (30) and (31), the second, third, and fourth lines of (30) and (31) are attained by exchanging the differential and expectation operators; carrying out the differentiation; and exchanging the expectation and summation operators and then applying Corollary 8, respectively. The exchanges are allowable as the infinite sum converges, as demonstrated in the foregoing paragraph. The last line of (31) is reached using Corollary 8 and invoking definitions of the two integrals in (27a) and (27b). Finally, setting $\partial e_{\text{GPDD}}/\partial y_\emptyset = 0$ in (30) and $\partial e_{\text{GPDD}}/\partial C_{u,\mathbf{j}_u} = 0$ in (31) results in (25) and (26), respectively. □

It is important to recognize that the definitions of the integrals $I_{u,\mathbf{j}_u}$ and $J_{u,\mathbf{j}_u;v,\mathbf{k}_v}$ in (27a) and (27b) are not identical to those presented in past work [17]. There, the aforementioned integrals were defined as[2]

$$
\bar{I}_{u,\mathbf{j}_u} := \int_{\mathbb{A}^N} y(\mathbf{x})\Psi_{u,\mathbf{j}_u}(\mathbf{x}_u)f_{\mathbf{X}_u}(\mathbf{x}_u)f_{\mathbf{X}_{-u}}(\mathbf{x}_{-u})d\mathbf{x} \tag{32}
$$

and

$$
\bar{J}_{u,\mathbf{j}_u;v,\mathbf{k}_v} := \int_{\mathbb{A}^N} \Psi_{u,\mathbf{j}_u}(\mathbf{x}_u)\Psi_{v,\mathbf{k}_v}(\mathbf{x}_v)f_{\mathbf{X}_u}(\mathbf{x}_u)f_{\mathbf{X}_{v\cap -u}}(\mathbf{x}_{v\cap -u})d\mathbf{x}. \tag{33}
$$

Clearly, the density functions in (32) and (33) are different than those in (27a) and (27b). As a result, the former integrals cannot be interpreted as expectations as the latter integrals. The difference arises due to distinct perspectives involved in deriving the final expressions of the expansion coefficients. Consequently, the resultant linear systems from this and past works are also different, although both lead to the calculation of the expansion coefficients. Indeed, the linear system (26) is new and has not been published elsewhere. More importantly, the new definitions of the two integrals in (27a) and (27b) and the linear system (26) enable a decoupling procedure for calculating the expansion coefficients efficiently, to be discussed next.

The linear system (26) can be broken down further as many of the integral coefficients, that is, $J_{u,\mathbf{j}_u;v,\mathbf{k}_v}$, vanishes, according to Corollary 8. As a result, the Fourier coefficients possess interaction only for a specific degree – a consequence of employing orthogonal polynomial basis. Reshuffling the coefficients according to the degree $1 \le l < \infty$, the GPDD in (24) can also be expressed by

$$
y(\mathbf{X}) \sim y_\emptyset + \sum_{l\in\mathbb{N}}\sum_{\substack{\emptyset\neq u\subseteq\{1,\ldots,N\}\\ 1\le|u|\le\min(N,l)}}\sum_{\substack{\mathbf{j}_u\in\mathbb{N}^{|u|}\\ |\mathbf{j}_u|=l}} C_{u,\mathbf{j}_u}\Psi_{u,\mathbf{j}_u}(\mathbf{X}_u). \tag{34}
$$

For each degree l, there are

$$
Q_{N,l} = \sum_{s=1}^{\min(N,l)}\binom{N}{s}\binom{l-1}{s-1} < \infty \tag{35}
$$

number of Fourier coefficients $C_{u,\mathbf{j}_u}$, $1 \le |u| \le \min(N,l)$, $|\mathbf{j}_u| = l$, which satisfy the $Q_{N,l}\times Q_{N,l}$ linear system

$$
\sum_{\substack{\emptyset\neq v\subseteq\{1,\ldots,N\}\\ 1\le|v|\le\min(N,l)}}\sum_{\substack{\mathbf{k}_v\in\mathbb{N}^{|v|}\\ |\mathbf{k}_v|=|\mathbf{j}_u|}} C_{v,\mathbf{k}_v}J_{u,\mathbf{j}_u;v,\mathbf{k}_v} = I_{u,\mathbf{j}_u},\ \ 1\le|u|\le\min(N,l),\ \ |\mathbf{j}_u| = l. \tag{36}
$$

[2] Strictly speaking, the integrals in [17] were defined using the support of the density function of $\mathbf{X}$ as $\mathbb{R}^N$. The extension to a support $\mathbb{A}^N \subseteq \mathbb{R}^N$ should follow readily.

The coefficient matrix in the matrix form of (36) comprises expectations of the product of two orthogonal polynomials. In other words, the coefficient matrix is a Gram matrix, which is positive-definite and hence invertible. Indeed, (34)–(36) facilitate a systematic and computationally efficient procedure for determining the Fourier coefficients of GPDD. Read the companion paper [14] for an algorithm to calculate the coefficients.

Corollary 11. *Given the introductory statements of Theorem 10, assume further that* $\mathbf{X} := (X_1, \ldots, X_N)^T$ *has independent, but not necessarily identical, components with known marginal probability density functions* $f_{X_i}(x_i)$, $i = 1, \ldots, N$. *Suppose, for* $i = 1, \ldots, N$, *that a set of univariate orthonormal polynomials* $\{\Psi_{\{i\},j_i}(x_i) : j_i = 1, \ldots, \infty\}$ *has been obtained, consistent with the probability measure* $f_{X_i}(x_i)dx_i$. *Then the proposed GPDD shrinks to the existing PDD, resulting in*

$$y(\mathbf{X}) \sim y_\emptyset + \sum_{\emptyset \neq u \subseteq \{1,\ldots,N\}} \sum_{\mathbf{j}_u \in \mathbb{N}^{|u|}} C_{u,\mathbf{j}_u} \prod_{p=1}^{|u|} \Psi_{\{i_p\},j_{i_p}}(X_{i_p}), \tag{37}$$

consisting of the Fourier coefficients

$$y_\emptyset = \mathbb{E}[y(\mathbf{X})] = \int_{\mathbb{A}^N} y(\mathbf{x}) \prod_{i=1}^{N} f_{X_i}(x_i)dx_i$$

and

$$C_{u,\mathbf{j}_u} = \mathbb{E}\left[y(\mathbf{X}) \prod_{p=1}^{|u|} \Psi_{\{i_p\},j_{i_p}}(X_{i_p}) \right] := \int_{\mathbb{A}^N} y(\mathbf{x}) \prod_{p=1}^{|u|} \Psi_{\{i_p\},j_{i_p}}(x_{i_p}) \prod_{i=1}^{N} f_{X_i}(x_i)dx_i.$$

Proof. For independent input variables, one has $f_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^{N} f_{X_i}(x_i)$, leading to

$$\Psi_{u,\mathbf{j}_u}(\mathbf{X}_u) = \prod_{p=1}^{|u|} \Psi_{\{i_p\},j_{i_p}}(X_{i_p}),$$

as per Corollary 9. Also, the integral $J_{u,\mathbf{j}_u;v,\mathbf{k}_v}$ vanishes whenever $u \neq v$. Subsequently, the result of Corollary 11 follows readily. □

It is important to emphasize that the infinite series in (24) or (37) is not guaranteed to converge almost surely to $y(\mathbf{X})$. In addition, higher-order moments of PDD or GPDD do not necessarily converge to their correct limits. They are rooted in fundamental limitations of Fourier or Fourier-like series.

5.1. Connection to the generalized ADD

It is important to point out the relation between GPDD and the generalized ADD discussed in Section 3. For instance, contrasting (6) and (18) produces the closure of an orthogonal decomposition of

$$\mathcal{W}_u = \overline{\bigoplus_{l=|u|}^{\infty} \mathcal{Z}_l^u}$$

with respect to the polynomial spaces $\mathcal{Z}_l^u$, $|u| \le l < \infty$. This leads to the convergent expansion

$$y_u(\mathbf{X}_u) \sim \sum_{\mathbf{j}_u \in \mathbb{N}^{|u|}} C_{u,\mathbf{j}_u} \Psi_{u,\mathbf{j}_u}(\mathbf{X}_u). \tag{38}$$

Indeed, the connection between GPDD and the generalized ADD is clearly palpable, where the former can be viewed as a polynomial variant of the latter. According to (24) or (38), the product $C_{u,\mathbf{j}_u} \Psi_{u,\mathbf{j}_u}(\mathbf{X}_u)$ characterizes a $|u|$-variate, $|\mathbf{j}_u|$th-order GPDD component function of $y(\mathbf{X})$, signifying the $|\mathbf{j}_u|$th-order polynomial refinement of the $|u|$-variate component function $y_u(\mathbf{X}_u)$ of the generalized ADD.

Furthermore, given the statistical properties of multivariate orthogonal polynomials in Proposition 7 or Corollary 8, the second-moment properties of $y_u(\mathbf{X}_u)$, that is, (3) and (4), are automatically satisfied when $y_u(\mathbf{X}_u)$ is expanded compliant with (38). Therefore, GPDD is bequeathed with all desirable properties of the generalized ADD, thus fulfilling a fundamental requirement for any adaptation of the latter.

5.2. Truncated GPDD

For computational tractability, it is necessary that the infinite series of GPDD be truncated. However, the truncation can be performed in several ways. A natural yet forthright approach adopted here involves (1) retaining all polynomials in at most $0 \le S \le N$ variables, thereby maintaining the degrees of interaction among input variables less than or equal to S, and (2) upholding polynomial expansion orders or degrees (total) less than or equal to $S \le m < \infty$. As a result, an S-variate, mth-order GPDD approximation of $y(\mathbf{X})$, expressed by[3]

$$\begin{aligned} y_{S,m}(\mathbf{X}) &= y_\emptyset + \sum_{s=1}^{S} \sum_{l=s}^{m} \sum_{\substack{\emptyset \ne u \subseteq \{1,\ldots,N\} \\ |u|=s}} \sum_{\substack{\mathbf{j}_u \in \mathbb{N}^{|u|} \\ |\mathbf{j}_u|=l}} C_{u,\mathbf{j}_u} \Psi_{u,\mathbf{j}_u}(\mathbf{X}_u) \\ &= y_\emptyset + \sum_{\substack{\emptyset \ne u \subseteq \{1,\ldots,N\} \\ 1 \le |u| \le S}} \sum_{\substack{\mathbf{j}_u \in \mathbb{N}^{|u|} \\ |u| \le |\mathbf{j}_u| \le m}} C_{u,\mathbf{j}_u} \Psi_{u,\mathbf{j}_u}(\mathbf{X}_u), \end{aligned} \tag{39}$$

is envisioned, which consists of

$$L_{S,m} = 1 + \sum_{s=1}^{S} \binom{N}{s} \binom{m}{s} \tag{40}$$

number of Fourier coefficients, including $y_\emptyset$. The truncated GPDD proposed requires a few clarifications. First, the truncation with respect to the polynomial expansion order is connected to the total degree index set. There are other types of truncations, for instance, the ones entailing the tensor product and hyperbolic cross index sets. It is common to use either of these two index sets, although the former choice is susceptible to the curse of dimensionality, rendering it impractical for large-scale problems. The hyperbolic cross index set, popular with applied mathematicians, is mostly used in the context of PCE. Anisotropic versions of all these choices and others may be used for truncating GPDD. In this work, however, only the total degree index set is used for the GPDD approximation. Second, the right side of (39) encompasses sums of orthogonal polynomials in at most S variables. Hence, the designation "S-variate" adopted for the GPDD approximation should be understood from the perspective of including at most S-degree interaction of input variables. Finally, by varying $S \le N$ and $S \le m < \infty$, (39) furnishes a hierarchical and convergent sequence of GPDD approximations.

A strong incentive to exploit the ADD- and GPDD-derived approximations is computational efficiency. In a real-world setting, the function $y(\mathbf{X})$ often has a hidden dimensional structure, as described by (1), where the low-dimensional component functions $y_u(\mathbf{X})$, $|u| \ll N$, are dominant over their high-dimensional counterparts. In which case, $y(\mathbf{X})$ can be accurately approximated by sums of lower-dimensional component functions, but still retaining effects of all random variables $\mathbf{X}$ in a high-dimensional stochastic problem. For example, an S-variate, mth-order GPDD approximation $y_{S,m}(\mathbf{X})$ contains sums of at most S-dimensional, mth-order component functions $C_{u,\mathbf{j}_u} \Psi_{u,\mathbf{j}_u}(\mathbf{X}_u)$, preserving at most S degree of interactions among input variables. The approximation does not exclude effects of any input random variables and can pick up arbitrarily high-order effects of random variables commensurate with the chosen value of m. Indeed, if $S \ll N$, as anticipated in practical applications, then many high-dimensional functions from real-world applications can be effectively represented by GPDD approximations. From (40), the computational complexity of a truncated GPDD is polynomial, as opposed to exponential. Therefore, GPDD reduces the curse of dimensionality to an extent possible.

Given an S-variate, mth-order GPDD approximation $y_{S,m}(\mathbf{X})$ of $y(\mathbf{X})$, what can be said about its approximation quality? The following two propositions address this question.

Proposition 12. *Let*

$$\Pi_{S,m}^{N} := \mathbf{1} \oplus \bigcup_{\substack{\emptyset \ne u \subseteq \{1,\ldots,N\} \\ 1 \le |u| \le S}} \bigoplus_{\substack{\mathbf{j}_u \in \mathbb{N}^{|u|} \\ |u| \le |\mathbf{j}_u| \le m}} \mathrm{span}\{\Psi_{u,\mathbf{j}_u}(\mathbf{X}_u) : \mathbf{j}_u \in \mathbb{N}^{|u|}\} \subseteq L^2(\Omega, \mathcal{F}, \mathbb{P}) \tag{41}$$

be a subspace consisting of all polynomials in $\mathbf{X}$ *with at most* $1 \le S \le N$ *degree of interaction and at most* $S \le m < \infty$ *order, including constants. For any* $y(\mathbf{X}) \in L^2(\Omega, \mathcal{F}, \mathbb{P})$*, denote by* $y_{S,m}(\mathbf{X})$ *and* $y_{N,m}(\mathbf{X})$ *its S-variate, mth-order and N-variate, mth-order GPDD approximations, respectively. Then the truncation error* $y(\mathbf{X}) - y_{N,m}(\mathbf{X})$ *is orthogonal to the subspace* $\Pi_{N,m}^{N} \subseteq L^2(\Omega, \mathcal{F}, \mathbb{P})$*. Furthermore,* $\mathbb{E}[\{y(\mathbf{X}) - y_{S,m}(\mathbf{X})\}^2] \to 0$ *as* $S \to N$ *and* $m \to \infty$.

Proof. Set $S = N$ in (41) to define the subspace $\Pi_{N,m}^{N} \subseteq L^2(\Omega, \mathcal{F}, \mathbb{P})$. Denote by

$$\bar{y}_{N,m}(\mathbf{X}) := \bar{y}_\emptyset + \sum_{\emptyset \ne v \subseteq \{1,\ldots,N\}} \sum_{\substack{\mathbf{k}_v \in \mathbb{N}^{|v|} \\ |v| \le |\mathbf{k}_v| \le m}} \bar{C}_{v,\mathbf{k}_v} \Psi_{v,\mathbf{k}_v}(\mathbf{X}_v),$$

an arbitrary element of $\Pi_{N,m}^{N}$ involving arbitrary expansion coefficients $\bar{y}_\emptyset$ and $\bar{C}_{v,\mathbf{k}_v}$. Then the expectation

$$\begin{aligned} &\mathbb{E}[\{y(\mathbf{X}) - y_{N,m}(\mathbf{X})\} \bar{y}_{N,m}(\mathbf{X})] \\ &= \mathbb{E}\left[\left\{ \sum_{\emptyset \ne u \subseteq \{1,\ldots,N\}} \sum_{\substack{\mathbf{j}_u \in \mathbb{N}^{|u|} \\ m+1 \le |\mathbf{j}_u| < \infty}} C_{u,\mathbf{j}_u} \Psi_{u,\mathbf{j}_u}(\mathbf{X}_u) \right\} \left\{ \bar{y}_\emptyset + \sum_{\emptyset \ne v \subseteq \{1,\ldots,N\}} \sum_{\substack{\mathbf{k}_v \in \mathbb{N}^{|v|} \\ |v| \le |\mathbf{k}_v| \le m}} \bar{C}_{v,\mathbf{k}_v} \Psi_{v,\mathbf{k}_v}(\mathbf{X}_v) \right\} \right] \\ &= 0, \end{aligned}$$

[3] The terms *degree* and *order* associated with GPDD or orthogonal polynomials are used equivalently in the paper.

where the equality to *zero* in the last line arises from Corollary 8. Therefore, the first part of the proposition is proved. For the last part, bring up the Pythagoras theorem, producing

$$\mathbb{E}[\{y(\mathbf{X})-y_{N,m}(\mathbf{X})\}^2]+\mathbb{E}[y_{N,m}^2(\mathbf{X})]=\mathbb{E}[y^2(\mathbf{X})].$$

Then

$$\begin{aligned}\lim_{S\to N, m\to\infty}\mathbb{E}\left[\{y(\mathbf{X})-y_{S,m}(\mathbf{X})\}^2\right] &= \lim_{m\to\infty}\mathbb{E}\left[\{y(\mathbf{X})-y_{N,m}(\mathbf{X})\}^2\right]\\ &= \lim_{m\to\infty}\left(\mathbb{E}\left[y^2(\mathbf{X})\right]-\mathbb{E}\left[y_{N,m}^2(\mathbf{X})\right]\right)\\ &= \mathbb{E}\left[y^2(\mathbf{X})\right]-\lim_{S\to N, m\to\infty}\mathbb{E}\left[y_{S,m}^2(\mathbf{X})\right]\\ &= 0,\end{aligned}$$

where the second line uses the result of the Pythagoras theorem; and the equality to *zero* in the last line stems from Theorem 10, which says that $\mathbb{E}[y_{S,m}^2(\mathbf{X})]\to\mathbb{E}[y^2(\mathbf{X})]$ as $S\to N$ and $m\to\infty$. □

Proposition 12 and its proof confirm mean-square convergence also described in Theorem 10.

Proposition 13. *Let $\Pi_{S,m}^N$ be as defined in Proposition 12. Then the S-variate, mth-order GPDD approximation $y_{S,m}(\mathbf{X})$ is the best approximation of $y(\mathbf{X})\in L^2(\Omega,\mathcal{F},\mathbb{P})$ in the sense of L^2, that is,*

$$\mathbb{E}\left[\{y(\mathbf{X})-y_{S,m}(\mathbf{X})\}^2\right]=\inf_{\bar{y}_{S,m}\in\Pi_{S,m}^N}\mathbb{E}\left[\{y(\mathbf{X})-\bar{y}_{S,m}(\mathbf{X})\}^2\right].$$

Proof. Let

$$\bar{y}_{S,m}(\mathbf{X}):=\bar{y}_\emptyset+\sum_{\substack{\emptyset\ne u\subseteq\{1,\ldots,N\}\\ 1\le|u|\le S}}\sum_{\substack{\mathbf{j}_u\in\mathbb{N}^{|u|}\\ |u|\le|\mathbf{j}_u|\le m}}\bar{C}_{v,\mathbf{k}_v}\Psi_{v,\mathbf{k}_v}(\mathbf{X}_v),$$

with arbitrary expansion coefficients $\bar{y}_\emptyset$ and $\bar{C}_{u,\mathbf{j}_u}$, be any element of the subspace $\Pi_{S,m}^N\subseteq L^2(\Omega,\mathcal{F},\mathbb{P})$ defined in (41). To minimize $\mathbb{E}[\{y(\mathbf{X})-\bar{y}_{S,m}(\mathbf{X})\}^2]$, the derivatives with respect to the coefficients must be *zero*, that is,

$$\frac{\partial}{\partial\bar{y}_\emptyset}\mathbb{E}\left[\{y(\mathbf{X})-\bar{y}_{S,m}(\mathbf{X})\}^2\right]=\frac{\partial}{\partial\bar{C}_{u,\mathbf{j}_u}}\mathbb{E}\left[\{y(\mathbf{X})-\bar{y}_{S,m}(\mathbf{X})\}^2\right]=0.$$

From the proof of Theorem 10, for instance, (30) and (31) and the following text, the derivatives are *zero* only when $\bar{y}_\emptyset=y_\emptyset$ and $\bar{C}_{u,\mathbf{j}_u}=C_{u,\mathbf{j}_u}$, where $y_\emptyset$ and $C_{u,\mathbf{j}_u}$ are the Fourier coefficients of GPDD in (25) and (26), respectively. □

5.3. Infinite number of variables

Consider a stochastic problem involving a countable sequence of input random variables, each of which satisfies Assumption 1 and is defined on a sufficiently rich probability triple. The presence of such problems is ubiquitous in information theory and stochastic process [25]. Proposition 14 explains GPDD's relevance even when there are infinitely many random variables, provided that certain assumptions hold.

Proposition 14. *For a stochastic problem, let $\{X_i\}_{i\in\mathbb{N}}$ be a countable sequence of input random variables, satisfying Assumption 1. Denote by $(\Omega,\mathcal{F}_\infty,\mathbb{P})$ the probability triple of $\{X_i\}_{i\in\mathbb{N}}$, where $\mathcal{F}_\infty:=\sigma(\{X_i\}_{i\in\mathbb{N}})$ is the associated σ-algebra engendered. Let $y(\{X_i\}_{i\in\mathbb{N}})\in L^2(\Omega,\mathcal{F}_\infty,\mathbb{P})$ be an output random variable of interest, where $y:\mathbb{A}^{\mathbb{N}}\to\mathbb{R}$. Then the GPDD of $y(\{X_i\}_{i\in\mathbb{N}})$ converges to $y(\{X_i\}_{i\in\mathbb{N}})$ in mean-square, in probability, and in distribution.*

Proposition 14 is similar to the case of infinitely many input variables studied in the context of PDD [2] or GPCE [15]. The proof is omitted here, as it can be easily obtained from a straightforward extension of prior works.

5.4. Comparison with generalized polynomial chaos expansion

While the paper focuses on a dimensionwise Fourier-like series in orthogonal polynomials, a comparison with competing expansions entailing orthogonal polynomials without dimensional hierarchy should be intriguing. One such expansion is GPCE, developed recently for tackling dependent input random variables [15]. It is derived from a degree-wise splitting of the polynomial spaces, so that any square-integrable output random variable $y(\mathbf{X})$ can be expanded as [15]

$$y(\mathbf{X})\sim\sum_{\mathbf{j}\in\mathbb{N}_0^N}C_{\mathbf{j}}\Psi_{\mathbf{j}}(\mathbf{X}), \tag{42}$$

where $\{\Psi_{\mathbf{j}}(\mathbf{X}):\mathbf{j}\in\mathbb{N}_0^N\}$ is an infinite set of measure-consistent multivariate orthonormal polynomials in $\mathbf{X}$ and $C_{\mathbf{j}}\in\mathbb{R}$, $\mathbf{j}\in\mathbb{N}_0^N$, are the Fourier coefficients of GPCE. Like GPDD, the GPCE of $y(\mathbf{X})\in L^2(\Omega,\mathcal{F},\mathbb{P})$ under Assumption 1 also converges to $y(\mathbf{X})$ in mean-square, in probability, and in distribution [15]. When truncated according to the total degree index set, the pth-order GPCE approximation of $y(\mathbf{X})$, where $0 \le p < \infty$, reads

$$y_p(\mathbf{X}) = \sum_{\substack{\mathbf{j} \in \mathbb{N}_0^N \\ 0 \le |\mathbf{j}| \le p}} C_{\mathbf{j}} \Psi_{\mathbf{j}}(\mathbf{X}). \tag{43}$$

The use of total degree index set for the GPCE approximation is consistent with the order truncation of GPDD in (39). Therefore, a comparison between GPDD and GPCE approximations, as conducted in this work, is fair. Clearly, the two infinite series from GPDD and GPCE, defined by (24) and (42), respectively, are the same or equivalent with respect to their identical second-moment properties. However, GPDD and GPCE when truncated are not. Indeed, three notable observations jump out. First, the terms of the GPCE approximation in (43) are arranged strictly accordingly to the order of orthogonal polynomials. On the other hand, the terms of the GPDD approximation in (39) are organized in connection with both the degree of interaction among random variables and the order of orthogonal polynomials. Consequently, the approximation quality and computational efficiency of their truncated series may vary appreciably.

Second, for stochastic problems involving highly nonlinear functions but encompassing weak interactions among input variables – a premise supported by real-world applications – the GPDD approximation is anticipated to be vastly more computationally efficient than the GPCE approximation. This is largely due to GPDD's dimensionwise hierarchical structure that is not present in GPCE. To better understand this point, a numerical example discussing error analysis of both approximations is illustrated in the following section.

Third, GPDD works with low-dimensional marginal distributions, avoiding the need for the joint distribution of a high-dimensional random input. In consequence, a GPDD approximation entails only low-dimensional multivariate polynomials as compared with high-dimensional multivariate polynomials involved in a GPCE approximation. Therefore, GPDD's effort is generally lower than GPCE's just to form the polynomial basis. In a practical setting, though, it is highly likely that the joint distribution is unknown, but only the marginals are available. In such a case, a GPDD approximation can still be constructed by generating measure-consistent multivariate polynomials in subsets of input variables.

6. A numerical example

The mathematical example presented in this section is simple and meant for highlighting potential computational advantage of GPDD over GPCE approximations. For more real-life examples, readers are referred to the companion paper [14].

Consider a polynomial

$$y(\mathbf{X}) = 10\left(X_1^6 + X_2^6 + X_3^6\right) + \frac{1}{c_1}(X_1X_2 + X_1X_3 + X_2X_3) + \frac{1}{c_2}X_1^2X_2^2X_3^2$$

in three real-valued, dependent random variables (X_1, X_2, X_3), which follow the Dirichlet probability density function

$$f_{X_1X_2X_3}(x_1, x_2, x_3) = \begin{cases} \dfrac{\Gamma\left(\sum_{i=1}^{4} \kappa_i + 2\right)}{\prod_{i=1}^{4} \Gamma\left(\kappa_i + \frac{1}{2}\right)} \left(\prod_{i=1}^{3} x_i^{\kappa_i - \frac{1}{2}}\right)(1 - x_1 - x_2 - x_3)^{\kappa_4 - \frac{1}{2}}, & \mathbf{x} = (x_1, x_2, x_3)^T \in \mathbb{T}^3, \\ 0, & \text{otherwise}, \end{cases}$$

on the standard tetrahedron $\mathbb{T}^3 := \{(x_1, x_2, x_3) : 0 \le x_1, x_2, x_3; x_1 + x_2 + x_3 \le 1\}$, where $\kappa_1 = \kappa_2 = \kappa_3 = \kappa_4 = 1$. Here, c_1 and c_2 are two real-valued constants. Two distinct sets of their values were chosen: (1) $c_1 = 10$, $c_2 = 1000$; and (2) $c_1 = 1$, $c_2 = 1$, reflecting weak and strong interactions, respectively, among input variables. The objective of this example is to evaluate the approximation quality of GPDD approximations in terms of the second-moment statistics of $y(\mathbf{X})$ and contrast the GPDD results with those obtained from the GPCE approximations. In both approximations, the total degree index set was employed.

Under Assumption 1, bases comprising multivariate orthogonal polynomials consistent with the Dirichlet probability density function exist. One such basis, obtained using a Rodrigues-type formula [18] and subsequent scaling, leads to the standardized version $\{\Psi_{u,\mathbf{j}_u}(\mathbf{x}_u)\}$, as described in (13). More explicitly, Table 1 presents first-, second-, and third-order (-degree) orthogonal polynomials in $\mathbf{x}_u$, $1 \le |u| \le 3$, obtained for the Dirichlet density function.

Define two relative errors

$$e_{S,m} := \frac{|\mathrm{var}[y(\mathbf{X})] - \mathrm{var}[y_{S,m}(\mathbf{X})]|}{\mathrm{var}[y(\mathbf{X})]} \quad \text{and} \quad e_p := \frac{|\mathrm{var}[y(\mathbf{X})] - \mathrm{var}[y_p(\mathbf{X})]|}{\mathrm{var}[y(\mathbf{X})]}$$

in the variances, committed by the S-variate, mth-order GPDD approximation $y_{S,m}(\mathbf{X})$ and the pth-order GPCE approximation $y_p(\mathbf{X})$, respectively, of $y(\mathbf{X})$. Here, the exact variance $\mathrm{var}[y(\mathbf{X})]$ and the GPDD variance $\mathrm{var}[y_{S,m}(\mathbf{X})]$, given S and m, were determined analytically from their definitions, which is possible as (1) y and $y_{S,m}$ are both polynomials and (2) expectations of monomials $\{\mathbf{X}^{\mathbf{j}}, 0 \le |\mathbf{j}| < \infty\}$ for $\mathbf{X}$ following a Dirichlet distribution are known analytically. In contrast, the GPCE variance, given p, was calculated using the analytical formula from a prior work [15]. Therefore, all errors were calculated exactly.

Table 1
A few orthogonal polynomials consistent with the Dirichlet density function.[a]

$$\Psi_{\{i\}1} = \sqrt{\frac{7}{3}} - 4\sqrt{\frac{7}{3}}x_i,$$
$$\Psi_{\{i\}2} = \frac{224x_i^2}{\sqrt{55}} - 28\sqrt{\frac{5}{11}}x_i + 3\sqrt{\frac{5}{11}},$$
$$\Psi_{\{i\}3} = -128\sqrt{\frac{15}{13}}x_i^3 + 672\sqrt{\frac{3}{65}}x_i^2 - 112\sqrt{\frac{5}{39}}x_i + 7\sqrt{\frac{5}{39}},$$
$$\Psi_{\{i_1,i_2\}11} = 11\sqrt{\frac{42}{19}}x_{i_1}^2 + 2\sqrt{798}x_{i_2}x_{i_1} - 14\sqrt{\frac{42}{19}}x_{i_1} + 11\sqrt{\frac{42}{19}}x_{i_2}^2 - 14\sqrt{\frac{42}{19}}x_{i_2} + 3\sqrt{\frac{42}{19}},$$
$$\begin{aligned}\Psi_{\{i_1,i_2\}12} = {} & -6\sqrt{\frac{1001}{37}}x_{i_1}^3 - 614\sqrt{\frac{77}{481}}x_{i_2}x_{i_1}^2 + 174\sqrt{\frac{77}{481}}x_{i_1}^2 - 2\sqrt{37037}x_{i_2}^2x_{i_1} + 788\sqrt{\frac{77}{481}}x_{i_2}x_{i_1} \\ & -114\sqrt{\frac{77}{481}}x_{i_1} - 18\sqrt{\frac{1001}{37}}x_{i_2}^3 + 30\sqrt{\frac{1001}{37}}x_{i_2}^2 - 174\sqrt{\frac{77}{481}}x_{i_2} + 18\sqrt{\frac{77}{481}},\end{aligned}$$
$$\begin{aligned}\Psi_{\{i_1,i_2\}21} = {} & -18\sqrt{\frac{1001}{37}}x_{i_1}^3 - 2\sqrt{37037}x_{i_2}x_{i_1}^2 + 30\sqrt{\frac{1001}{37}}x_{i_1}^2 - 614\sqrt{\frac{77}{481}}x_{i_2}^2x_{i_1} + 788\sqrt{\frac{77}{481}}x_{i_2}x_{i_1} \\ & -174\sqrt{\frac{77}{481}}x_{i_1} - 6\sqrt{\frac{1001}{37}}x_{i_2}^3 + 174\sqrt{\frac{77}{481}}x_{i_2}^2 - 114\sqrt{\frac{77}{481}}x_{i_2} + 18\sqrt{\frac{77}{481}},\end{aligned}$$
$$\begin{aligned}\Psi_{\{1,2,3\}111} = {} & -12\sqrt{55}x_1^3 - 50\sqrt{55}x_2x_1^2 - 50\sqrt{55}x_3x_1^2 + 138\sqrt{\frac{11}{5}}x_1^2 - 50\sqrt{55}x_2^2x_1 - 50\sqrt{55}x_3^2x_1 \\ & +346\sqrt{\frac{11}{5}}x_2x_1 - 128\sqrt{55}x_2x_3x_1 + 346\sqrt{\frac{11}{5}}x_3x_1 - 96\sqrt{\frac{11}{5}}x_1 - 12\sqrt{55}x_2^3 \\ & -12\sqrt{55}x_3^3 + 138\sqrt{\frac{11}{5}}x_2^2 - 50\sqrt{55}x_2x_3^2 + 138\sqrt{\frac{11}{5}}x_3^2 - 96\sqrt{\frac{11}{5}}x_2 - 50\sqrt{55}x_2^2x_3 \\ & +346\sqrt{\frac{11}{5}}x_2x_3 - 96\sqrt{\frac{11}{5}}x_3 + 18\sqrt{\frac{11}{5}}.\end{aligned}$$

[a]Here, $i = 1, 2, 3;\ i_1, i_2 = 1, 2, 3,\ i_2 > i_1$.

Table 2
Relative errors in variances by GPDD and GPCE approximations for the fist case ($c_1 = 10,\ c_2 = 1000$).

	Univariate GPDD		Bivariate GPDD		GPCE	
m or p	$e_{1,m}$	$L_{1,m}$	$e_{2,m}$	$L_{2,m}$	e_p	L_p
1	0.856363	4			0.856363	4
2	0.219054	7	0.219038	10	0.219038	10
3	0.038876	10	0.038860	19	0.038860	20
4	1.62697×10^{-3}	13	1.61074×10^{-3}	31	1.61074×10^{-3}	35
5	3.31864×10^{-5}	16	1.69589×10^{-5}	46	1.69589×10^{-5}	56

Table 2 presents the errors $e_{S,m}$ and e_p for the first case of $c_1 = 10,\ c_2 = 1000$, entailing weak interactions among input variables. The results are displayed for various combinations of the truncations parameters of GPDD and GPCE: $S = 1, 2$, $m = 1, 2, 3, 4, 5$, and $p = 1, 2, 3, 4, 5$. The two truncations with respect to the degree of interaction $S = 1$ and $S = 2$ represent the univariate GPDD and bivariate GPDD approximations, respectively. According to Table 2, the GPDD approximation errors drop with respect to S and m as expected. With the exception of $m = 5$, the errors from the univariate and bivariate GPDD approximations are nearly identical. Moreover, the errors from the two GPDD approximations are the same or very close to the errors from the respective GPCE approximations. This is because the chosen function y, albeit it is highly nonlinear with respect to $\mathbf{X}$, is endowed with little interactions among input variables. From the comparisons of computational efforts, measured in terms of the numbers of expansion coefficients also listed in Table 2, both GPDD approximations are more efficient than the GPCE approximations for the same expansion order. For instance, the univariate, fifth-order GPDD approximation ($S = 1,\ m = 5$) achieves a relative error of 3.31864×10^{-5} employing only 16 expansion coefficients. In contrast, to match the same-order error, the fifth-order GPCE approximation ($p = 5$) is needed, committing a relative error of 1.69589×10^{-5} at the cost of 56 expansion coefficients. Therefore, the univariate GPDD approximation is substantially more economical than the GPCE approximation for a similar accuracy. The bivariate, fifth-order GPDD approximation produces practically the same result of the fifth-order GPCE approximation, but still upholding some computational advantage over the latter. However, the gain in efficiency from the bivariate GPDD approximation is much less than that from the univariate GPDD approximation. This is expected due to the added computational expense to include, in addition to the main effects, all two-variable interaction effects, in the bivariate approximation. Nonetheless, when the main effects of the input variables on y are dominant over their interactive effects, as is the first case in this example, the GPDD approximation is expected to be more effective than the GPCE approximation.

Similar results are presented in Table 3 for the second case of $c_1 = 1,\ c_2 = 1$. As, in this case, the interactive effects are stronger than before, the quality of the best univariate GPDD approximation in Table 3 has noticeably deteriorated when

Table 3
Relative errors in variances by GPDD and GPCE approximations for the second case ($c_1 = 1,\ c_2 = 1$).

	Univariate GPDD		Bivariate GPDD		GPCE	
m or p	$e_{1,m}$	$L_{1,m}$	$e_{2,m}$	$L_{2,m}$	e_p	L_p
1	0.738999	4			0.738999	4
2	0.215212	7	0.213612	10	0.213612	10
3	0.039505	10	0.037905	19	0.037905	20
4	3.16727×10^{-3}	13	1.56798×10^{-3}	31	1.56796×10^{-3}	35
5	1.61589×10^{-3}	16	1.65961×10^{-5}	46	1.65714×10^{-5}	56

compared with that in Table 2. More precisely, the error committed by the univariate, fifth-order GPDD approximation in the second case is two-order magnitude larger than the error in the first case. However, the error stemming from the bivariate, fifth-order GPDD approximation in the second case still compares favorably to that in the first case and the corresponding GPCE error in the second case. If the three-variable interaction is any more stronger, then the bivariate GPDD approximations will suffer the same fate of the univariate GPDD approximations. Given these results, the computational efficacy of GPDD over GPCE clearly depends on the existence of a hidden yet desirable dimensional hierarchy of input variables, where the interaction among input variables attenuates or vanish altogether. Fortunately, such dimensional hierarchy is commonly found in many real-life engineering problems [3–6].

For a more general discussion on the computational efforts by the two aforementioned approximations, consider the respective numbers of Fourier coefficients involved: (1) $L_{S,m}$ in (40) for the S-variate, mth-order GPDD approximation; and (2) $L_p = (N+p)!/(N!p!)$ for the pth-order GPCE approximations [15]. In other words, $L_{S,m}$ grows S-degree polynomially with respect to N, whereas L_p scales p-degree polynomially with N. For stochastic problems entailing highly nonlinear functions but containing mostly low-variate interactive effects of input variables, p is expected to be much larger than S. Consequently, the GPDD approximation should offer a hefty computational benefit over the GPDD approximation for the same expansion order. As an example, consider a stochastic problem involving 20 input random variables ($N = 20$) and the following truncation parameters of GPDD and GPCE: $S = 1, 2,\ m = 4$, and $p = 4$. In this case, the univariate, fourth-order GPDD approximation ($S = 1,\ m = 4$), the bivariate, fourth-order GPDD approximation ($S = 2,\ m = 4$), and the fourth-order GPCE approximation ($p = 4$) require 81, 1221, and 10,626 Fourier coefficients, respectively. Clearly, the growth of the number of Fourier coefficients in GPCE is much sharper than that in GPDD. This is chiefly because a GPCE approximation is determined by one truncation parameter p, which dictates the largest polynomial expansion order retained, but not the degree of interaction on its own. On the other hand, two truncation parameters S and m are engaged in a GPDD approximation, offering additional freedom in preserving the largest degree of interaction and largest polynomial expansion order independently. As a result, the number of expansion coefficients in GPCE rises exponentially, rendering the GPCE approximation significantly more expensive than the GPDD approximation.

Finally, comparisons between GPDD and GPCE approximations stemming from other kinds of truncations are desirable. For instance, using the hyperbolic cross index set, GPCE is able to draw in some interactive effects of input variables, making it perhaps more competitive with GPDD. They are subjects of future work.

7. Concluding remarks

A new generalized PDD, referred to as GPDD, of a square-integrable output random variable, encompassing hierarchically structured, multivariate orthogonal polynomials in dependent input random variables, was developed. A dimensional decomposition of polynomial spaces into respective subspaces, each spanned by a set of measure-consistent orthogonal polynomials, was established. The outcome is a polynomial surrogate of the generalized ADD, ultimately leading to GPDD of a general L^2-function without enforcing any tensor-product structure. Under requisite assumptions, the set of measure-consistent orthogonal polynomials constitutes a complete basis of each subspace, producing a union-sum collection of such sets of basis functions to span the space of all polynomials. Moreover, the aforementioned collection is L^2-dense in a Hilbert space of square-integrable functions, resulting in mean-square convergence of GPDD to the correct limit, including for the instance of infinitely many random variables. New results determining the statistical properties of random orthogonal polynomials were derived. The optimality and approximation quality of a truncated GPDD were verified or explained. For independent random variables, the proposed PDD reduces to the existing PDD, justifying the appellation GPDD introduced in this work. By exploiting the hierarchical structure of a function, if it exists, the GPDD approximation is expected to efficiently solve high-dimensional stochastic problems in the presence of dependent random variables with non-product-type probability measures.

References

[1] S. Rahman, A polynomial dimensional decomposition for stochastic computing, Int. J. Numer. Methods Eng. 76 (2008) 2091–2116.
[2] S. Rahman, Mathematical properties of polynomial dimensional decomposition, SIAM/ASA J. Uncert. Quant. 6 (2018) 816–844.
[3] A. Chakraborty, S. Rahman, Stochastic multiscale models for fracture analysis of functionally graded materials, Eng. Fract. Mech. 75 (2008) 2062–2086.
[4] S. Rahman, V. Yadav, Orthogonal polynomial expansions for solving random eigenvalue problems, Int. J. Uncert. Quant. 1 (2) (2011) 163–187.

[5] K. Tang, P.M. Congedo, R. Abgrall, Adaptive surrogate modeling by ANOVA and sparse polynomial dimensional decomposition for global sensitivity analysis in fluid simulation, J. Comput. Phys. 314 (1) (2016) 557–589.
[6] X. Ren, V. Yadav, S. Rahman, Reliability-based design optimization by adaptive-sparse polynomial dimensional decomposition, Struct. Multidiscipl. Optim. 53 (3) (2016) 425–452.
[7] M. Griebel, M. Holtz, Dimension-wise integration of high-dimensional functions with applications to finance, J. Complex. 26 (5) (2010) 455–489.
[8] W. Hoeffding, A class of statistics with asymptotically normal distribution, Annal. Math. Stat. 19 (3) (1948) pp.293–325. http://www.jstor.org/stable/2235637.
[9] F.Y. Kuo, I.H. Sloan, G.W. Wasilkowski, H. Wozniakowski, On decompositions of multivariate functions, Math. Comput. 79 (2011) 953–966.
[10] H. Rabitz, O. Alis, General foundations of high dimensional model representations, J. Math. Chem. 25 (1999) 197–233, doi:10.1023/A:1019188517934.
[11] S. Rahman, Approximation errors in truncated dimensional decompositions, Math. Comput. 83 (290) (2014) 2799–2819.
[12] I.M. Sobol, Sensitivity estimates for nonlinear mathematical models, Math. Model. Comput. Exp. 1 (1993) 407–414.
[13] M. Rosenblatt, Remarks on a multivariate transformation, Ann. Math. Statist. 23 (1952) 470–472.
[14] S. Rahman, Uncertainty quantification under dependent random variables by a generalized polynomial dimensional decomposition, Comput. Methods Appl. Mech. Eng. 344 (2019) 910–937.
[15] S. Rahman, A polynomial chaos expansion in dependent random variables, J. Appl. Math. Appl. 4 (2018) 1–26.
[16] G. Hooker, Generalized functional ANOVA diagnostics for high-dimensiobal functions of dependent variables, J. Comput. Graph. Stat. 16 (3) (2007) 709–732.
[17] S. Rahman, A generalized ANOVA dimensional decomposition for dependent probability measures, SIAM/ASA J. Uncert. Quant. 2 (2014) 670–697.
[18] C.F. Dunkl, Y. Xu, Orthogonal Polynomials of Several Variables, Encyclopedia of Mathematics and its Applications, 155, second, Cambridge University Press, 2001.
[19] A. Erdélyi, Higher Transcendental Functions volume II, Encyclopedia of Mathematics and its Applications, 155, McGraw-Hill, 1953.
[20] B. Holmquist, The d-variate vector hermite polynomial of order k, Linear Algebra Appl. 237/238 (1996) 155–190.
[21] P. Appell, J.K. de Fériet, Fonctions hyprgéométriques et hypersphériques, polynomes d'hermite, Gauthier-Villars, Paris, 1926.
[22] H.L. Krall, I.M. Sheffer, Orthogonal polynomials in two variables, Ann. Mat. Pura Appl 76 (4) (1967) 325–376.
[23] G.H. Golub, C.F. van Loan, Matrix Computations, third ed., The John Hopkins University Press, 1996.
[24] R. Courant, D. Hilbert, Methods of Mathematical Physics, I, Interscience Publishers Inc., 1966.
[25] M. Griebel, F.Y. Kuo, I.H. Sloan, The ANOVA decomposition of a non-smooth function of infinitely many variables can have every term smooth, Math. Comput. 86 (306) (2017) 1855–1876.