Second Order Moments of Multivariate Hermite Polynomials in Correlated Random Variables

Laura Lyman[1] [0000−0002−3590−3651] and
Gianluca Iaccarino[2]

[1] Stanford University, Stanford CA 94305, USA
lymanla@stanford.edu
[2] Stanford University, Stanford CA 94305, USA
jops@stanford.edu

Abstract. Polynomial chaos methods can be used to estimate solutions of partial differential equations under uncertainty described by random variables. The stochastic solution is represented by a polynomial expansion, whose deterministic coefficient functions are recovered through Galerkin projections. In the presence of multiple uncertainties, the projection step introduces products (second order moments) of the basis polynomials. When the input random variables are correlated Gaussians, calculating the products of the corresponding multivariate basis polynomials is not straightforward and can become computationally expensive. We present a new expression for the products by introducing multiset notation for the polynomial indexing, which allows for simple and efficient evaluation of the second-order moments of correlated multivariate Hermite polynomials.

Keywords: Polynomial chaos · Multivariate Hermite polynomials · Stochastic Galerkin methods

1 Introduction

Uncertainty quantification (UQ) is crucial for developing confidence in predictions resulting from mathematical models of physical phenomena such as those described by partial differential equations (PDEs) [2,12]. A common strategy in computational science is to represent sources of uncertainty by random variables, which causes the solution to the original differential equation(s) to become a function of stochastic parameters [2,5,12]. A polynomial chaos expansion (PCE) expresses the solution as an infinite series of square-integrable orthogonal polynomials of independent random variables [21]. A truncated version of this expansion, as first suggested by Ghanem and Spanos [7], can be used as a solution approximation; further, for the case in which the orthogonal polynomials are Hermite and the random variables centered Gaussians, this truncation is proved to converge in mean-square by Xiu et al. [22] via an application of a theorem by Cameron and Martin [1,5,9].

The most common UQ strategies involve Monte Carlo (MC) sampling, which suffers from a slow convergence rate proportional to the inverse square root of the number of samples [2,14]. If each sample evaluation is expensive — as is

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77980-1_53

often true for the solutions of PDEs — this slow convergence rate can make obtaining tens of thousands of samples computationally infeasible [2,12]. PCE approximations can offer significant computational advantages over Monte Carlo methods in such instances, although there are some exceptions [15].

In [15], Rahman generalizes the classical PCE to account for arbitrary but dependent multi-dimensional Gaussian parameters, proving convergence in mean-square, probability, and distribution. Prior to this work, the multivariate PCE was constrained by the assumption that its input random variables were independent. By introducing correlation into this more general representation, the multi-dimensional analog of Hermite polynomials — referred to as *multivariate Hermite polynomials* — become only *weakly* orthogonal rather than orthogonal [15,22]. Namely, for two multivariate Hermite polynomials $H_{\boldsymbol{\alpha}}$ and $H_{\boldsymbol{\beta}}$ with multi-indices $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, which we define formally in Sect. 1.2, weak orthogonality guarantees that

$$\mathbb{E}(H_{\boldsymbol{\alpha}}(\boldsymbol{\xi})H_{\boldsymbol{\beta}}(\boldsymbol{\xi})) = 0 \quad \text{if } |\boldsymbol{\alpha}| \neq |\boldsymbol{\beta}|, \qquad \boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \Sigma) \tag{1}$$

where $\Sigma \in \mathbb{R}^{n\times n}$ is a real, symmetric positive definite (SPD) covariance matrix. However, $\mathbb{E}(H_{\boldsymbol{\alpha}}(\boldsymbol{\xi})H_{\boldsymbol{\beta}}(\boldsymbol{\xi}))$ can be (and often is) nonzero for *distinct* $\boldsymbol{\alpha}, \boldsymbol{\beta}$ satisfying $|\boldsymbol{\alpha}| = |\boldsymbol{\beta}|$. The quantities $\mathbb{E}(H_{\boldsymbol{\alpha}}(\boldsymbol{\xi})H_{\boldsymbol{\beta}}(\boldsymbol{\xi}))$ for various $\boldsymbol{\alpha}, \boldsymbol{\beta}$ are called the *double products* or *second moments* of the multivariate Hermite polynomials. To demonstrate why these double products are important, we will illustrate how a multi-dimensional PCE can be applied in a general setting.

1.1 Application Case and Motivation

For convenience, let $\mathcal{D} \subset \mathbb{R}^n \times \mathbb{R}_{t\geq 0}$ be a compact subset of a spatial and time domain $\mathbb{R}^n_{\boldsymbol{x}} \times \mathbb{R}_{t\geq 0}$ with initial time $t_0 \geq 0$. Let $u : \mathcal{D} \to \mathbb{R}$ be continuous and differentiable in both its spatial and temporal derivatives; further, let $u \in \mathcal{L}^2(\mathcal{D})$. This u represents the solution a differential equation

$$\mathcal{F}(u, \boldsymbol{x}, t) = 0. \tag{2}$$

Here $\mathcal{F}$ is a general differential operator, often a mix of linear and nonlinear terms. Let $\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \Sigma)$ be an n-dimensional random variable with known SPD covariance matrix $\Sigma = \mathbb{E}(\boldsymbol{\xi}\boldsymbol{\xi}^T)$. Then $\boldsymbol{\xi}$ has the joint probability density function [13,15,22]

$$\phi : \mathbb{R}^n \to \mathbb{R}_{\geq 0} \qquad \phi(\boldsymbol{x}; \Sigma) = \frac{1}{(2\pi)^{n/2}|\det(\Sigma)|^{1/2}} \exp\left(-\frac{1}{2}\boldsymbol{x}^T\Sigma^{-1}\boldsymbol{x}\right). \tag{3}$$

We assume uncertainty is present in the initial condition $u(\,\cdot\,, t_0)$ and represent it by setting

$$u(\boldsymbol{x}, t_0; \boldsymbol{\xi}) : \mathbb{R}^n \to \mathbb{R} \qquad u(\boldsymbol{x}, t_0; \boldsymbol{\xi}) = f(\boldsymbol{x}, \boldsymbol{\xi})$$

where f is a known function of $\boldsymbol{x}$ and $\boldsymbol{\xi}$. As statistics of $\boldsymbol{\xi}$, we require that both $u(\boldsymbol{x}, t; \boldsymbol{\xi})$ and $f(\boldsymbol{x}, \boldsymbol{\xi})$ have existing second moments.

The multivariate polynomial chaos expansion *separates* the deterministic and random components of u by writing $u(\boldsymbol{x}, t; \boldsymbol{\xi}) = \sum_{\boldsymbol{\alpha}\in\mathbb{N}^n_0} u_{\boldsymbol{\alpha}}(\boldsymbol{x}, t)H_{\boldsymbol{\alpha}}(\boldsymbol{\xi})$ where

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77980-1_53

$\boldsymbol{\alpha} \in \mathbb{N}_0^n$ is a multi-index with ℓ_1 norm $|\boldsymbol{\alpha}|$ (see Sect. 1.2) and $H_{\boldsymbol{\alpha}}$ is a *multivariate Hermite polynomial* (Def. 3). The $u_{\boldsymbol{\alpha}} : \mathcal{D} \to \mathbb{R}$ output deterministic coefficients. The $H_{\boldsymbol{\alpha}}$ are weakly orthogonal (Eq. 1) with respect to the measure $d\boldsymbol{\xi}$ induced by $\boldsymbol{\xi}$. When truncating to some M value, we can define

$$u^{(M)}(\boldsymbol{x}, t; \boldsymbol{\xi}) := \sum_{\substack{\boldsymbol{\alpha} \in \mathbb{N}_0^n \\ 0 \leq |\boldsymbol{\alpha}| \leq M}} u_{\boldsymbol{\alpha}}^{(M)}(\boldsymbol{x}, t) H_{\boldsymbol{\alpha}}(\boldsymbol{\xi}). \tag{4}$$

As proved in [15], $u^{(M)}(\boldsymbol{x}, t; \boldsymbol{\xi})$ converges to $u(\boldsymbol{x}, t; \boldsymbol{\xi})$ in mean-square, probability, and distribution. The procedure is to substitute $u^{(M)}(\boldsymbol{x}, t; \boldsymbol{\xi})$ into Eq. 2, multiply through by an arbitrary $H_{\boldsymbol{\beta}}$, and integrate with respect to the $d\boldsymbol{\xi}$ measure, repeating for each $H_{\boldsymbol{\beta}}$; this is effectively projecting onto the polynomial basis. From there, orthogonality conditions can be used to eliminate terms. For instance, if Eq. 2 represents the inviscid Burgers' equation,

$$\frac{\partial u}{\partial t} + u\left[\boldsymbol{a}^T \nabla_{\boldsymbol{x}} u\right] = 0 \qquad u(\boldsymbol{x}, 0; \boldsymbol{\xi}) = f(\boldsymbol{\xi}, \boldsymbol{x}) \qquad \boldsymbol{a} \in \mathbb{R}^n \text{ fixed}, \tag{5}$$

then this projection process and weak orthogonality (Eq. 1) gives

$$\sum_{\substack{\boldsymbol{i} \in \mathbb{N}_0^n \\ |\boldsymbol{i}| = |\boldsymbol{k}|}} \frac{\partial u_{\boldsymbol{i}}^{(M)}}{\partial t} \langle H_{\boldsymbol{i}}, H_{\boldsymbol{k}} \rangle + \sum_{\substack{\boldsymbol{i} \in \mathbb{N}_0^n \\ 0 \leq |\boldsymbol{i}| \leq M}} \sum_{\substack{\boldsymbol{j} \in \mathbb{N}_0^n \\ 0 \leq |\boldsymbol{j}| \leq M}} u_{\boldsymbol{i}}^{(M)} (\boldsymbol{a}^T \nabla_{\boldsymbol{x}} u_{\boldsymbol{j}}^{(M)}) \langle H_{\boldsymbol{i}}, H_{\boldsymbol{j}}, H_{\boldsymbol{k}} \rangle = 0 \tag{6}$$

for *every* $\boldsymbol{k} \in \mathbb{N}_0^n, 1 \leq |\boldsymbol{k}| \leq M$. Here $\langle H_{\boldsymbol{i}}, H_{\boldsymbol{k}} \rangle := \mathbb{E}_{\boldsymbol{\xi}}(H_{\boldsymbol{i}}(\boldsymbol{\xi}) H_{\boldsymbol{k}}(\boldsymbol{\xi}))$ denotes the *double product* and $\langle H_{\boldsymbol{i}}, H_{\boldsymbol{j}}, H_{\boldsymbol{k}} \rangle := \mathbb{E}_{\boldsymbol{\xi}}(H_{\boldsymbol{i}}(\boldsymbol{\xi}) H_{\boldsymbol{j}}(\boldsymbol{\xi}) H_{\boldsymbol{k}}(\boldsymbol{\xi}))$ denotes the *triple product*.

Eq. 6 is a system of *deterministic* PDEs to solve, where the number of PDEs is equal to the number of $\boldsymbol{k} \in \mathbb{N}_0^n$ such that $1 \leq |\boldsymbol{k}| \leq M$ for the selected M bound. In particular, a computer can solve such a system via standard numeric techniques if the coefficients $\langle H_{\boldsymbol{i}}, H_{\boldsymbol{k}} \rangle$ and $\langle H_{\boldsymbol{i}}, H_{\boldsymbol{j}}, H_{\boldsymbol{k}} \rangle$ are known. Moreover, *any linear term* of a general differential operator $\mathcal{F}(u, \boldsymbol{x}, t)$ of Eq. 2 will generate double product coefficients; this is not specific to inviscid Burgers' equation.

When the centered $\boldsymbol{\xi}$ is one dimensional, the double and triple products are given by simple expressions [19]. However, when $\boldsymbol{\xi}$ is n-dimensional and correlated, the expressions (first proved in [15]) become cumbersome. *Our contribution provides a new formula (Thm. 2) for the double product of multivariate Hermite polynomials of a centered Gaussian with generic covariance that is both simpler and more computationally efficient to implement than its previous formulation.*

This paper is organized as follows. In Sect. 1.2, we establish the necessary preliminaries for proving our contribution (Thm. 2). Sect. 2 outlines both the previous formula (Thm. 1) and our contribution for the double product and provides an instructive example (Ex. 1) in which each formula is applied. In Sect. 3, we compare and discuss computational complexity. Appendices A (Sect. 5) and B (Sect. 6) prove Thm. 2 and report some technical lemmas applied throughout the document.

1.2 Definitions and Notation

The following notation will be utilized throughout this paper:

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77980-1_53

- $\mathbb{N}$:= the natural numbers $= \{1, 2, 3, \ldots\}$,
- $\mathbb{N}_0 := \mathbb{N} \cup \{0\} = \{0, 1, 2, \ldots\}$,
- $[n] := \{1, \ldots, n\}$ for any $n \in \mathbb{N}$.

Definition 1 (Multi-index). *An n-dimensional multi-index $\boldsymbol{\alpha}$ is an n-tuple $(\alpha_1, \ldots, \alpha_n) \in \mathbb{N}_0^n$ of non-negative integers. For any $n \in \mathbb{N}$, a multi-index $\boldsymbol{\alpha}$ over $[n]$ is an n-tuple such that every $\alpha_i \in \mathbb{N}_0$ satisfies $\alpha_i \leq n$. Each α_i is referred to as the ith element of the multi-index $\boldsymbol{\alpha}$.*

Unless otherwise specified, a multi-index $\boldsymbol{\alpha}$ is assumed to have the notation given in Def. 1. For multi-indices $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{N}_0^n$, one defines [15,17]

1. *componentwise sum and difference*: $\boldsymbol{\alpha} \pm \boldsymbol{\beta} = (\alpha_1 \pm \beta_1, \ldots, \alpha_n \pm \beta_n)$,
2. *absolute value*: $|\boldsymbol{\alpha}| := \|\boldsymbol{\alpha}\|_1 = \alpha_1 + \cdots + \alpha_n$, which we call the *order* of $\boldsymbol{\alpha}$,
3. *factorial*: $\boldsymbol{\alpha}! = \alpha_1! \cdots \alpha_n! = \prod_{i=1}^n \alpha_i!$, and the
4. *partial derivative*: $D_{\boldsymbol{\alpha}}^{|\boldsymbol{\alpha}|} = \partial_1^{\alpha_1} \cdots \partial_n^{\alpha_n}$.

Moving forward, any $\boldsymbol{\alpha}$ will denote a multi-index over $[n]$ of order $k \geq 1$ unless otherwise specified.

Sometimes we will need to express $\boldsymbol{\alpha}$ in what we will refer to as its *mutliset notation*. Recall that a multiset is a modification of a set that allows for multiple instances of each of its elements. More formally [8,18], a *multiset* M on a set S is a pair (S, ν), where ν is a function $\nu : S \to \mathbb{N}$ assigning each element $x \in S$ its positive multiplicity i.e. the number of times x is repeated in M. We consider both the multi-index (Def. 1) and the proposed multiset notation (Def. 2) for a label $\boldsymbol{\alpha}$, because

1. the *multi-index* version is standard in relevant previous literature [13,15,17], and
2. the *multiset* version can be easier to utilize, as later showcased in Theorem 2.

Definition 2 (Multiset notation). *For a multi-index $\boldsymbol{\alpha}$ over $[n]$ of order $|\boldsymbol{\alpha}| = k \geq 1$, let $s(\boldsymbol{\alpha})$ denote the map*

$$s(\boldsymbol{\alpha}) : [k] \to [n] \qquad s(\boldsymbol{\alpha})(\ell) := s(\boldsymbol{\alpha})_\ell = \min\{i \in [n] \mid \ell \leq \sum_{r=1}^{i} \alpha_r\},^3$$

which we call the multiset notation *for $\boldsymbol{\alpha}$. It is straightforward (see Lemma 1) to verify that*

$$[s(\boldsymbol{\alpha})_1, \ldots, s(\boldsymbol{\alpha})_k] = [\overbrace{1, \ldots, 1}^{\alpha_1 \textit{ times}}, \ldots, \overbrace{n \cdots n}^{\alpha_n \textit{ times}}] = [1^{\alpha_1}, \ldots, n^{\alpha_n}].$$

Sometimes we use $s(\boldsymbol{\alpha})$ to refer to the output *$[s(\boldsymbol{\alpha})_1, \ldots, s(\boldsymbol{\alpha})_k]$ of the map across its whole domain $[k]$ rather than to the map itself, with context making the distinction clear. The* elements *of $s(\boldsymbol{\alpha})$ are $s(\boldsymbol{\alpha})_1, \ldots, s(\boldsymbol{\alpha})_k$, and the* order *$|s(\boldsymbol{\alpha})|$ of $s(\boldsymbol{\alpha})$ is the total number of such elements k, which is also the order of $\boldsymbol{\alpha}$. When $k = 0$, so that $\boldsymbol{\alpha} = (0, \ldots, 0)$, we write $s(\boldsymbol{\alpha}) = \emptyset$.*

[3] Note that this set is nonempty, because always $\ell \leq k = \sum_{r=1}^n \alpha_i$, so $i = n$ always satisfies the condition that $\ell \leq \sum_{r=1}^i \alpha_i$.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77980-1_53

The name *multiset notation* is chosen, since the outputs of $s(\boldsymbol{\alpha})$ are reminiscent of a multiset when written as $[s(\boldsymbol{\alpha})_1 \dots s(\boldsymbol{\alpha})_k]$, with each $i \in [n]$ represented with multiplicity α_i in the array (Lem. 1). Note that $\alpha_i = 0$ indicates that an i is not present in the $s(\boldsymbol{\alpha})$ array.

As an example, if $\boldsymbol{\alpha} = (2, 1, 0, 0, 1)$, then $s(\boldsymbol{\alpha})$ has order $k = 2 + 1 + 1 = 4$ with $s(\boldsymbol{\alpha})_1 = s(\boldsymbol{\alpha})_2 = 1, s(\boldsymbol{\alpha})_3 = 2$, and $s(\boldsymbol{\alpha})_4 = 5$. Thus, we represent $\boldsymbol{\alpha}$ via the map $s(\boldsymbol{\alpha})$ by the multiset notation $s(\boldsymbol{\alpha}) = [1, 1, 2, 5]$.

Finally, the *partial derivative operator* for $s(\boldsymbol{\alpha})$ is defined as

$$D^k_{s(\boldsymbol{\alpha})} := \partial_{s(\boldsymbol{\alpha})_1} \cdots \partial_{s(\boldsymbol{\alpha})_k} = \partial_1^{\alpha_1} \cdots \partial_n^{\alpha_n}$$

so that $D^k_{s(\boldsymbol{\alpha})} = D^k_{\boldsymbol{\alpha}}$ as expected. With this notation in place, we present the multi-dimensional analog of the Hermite polynomial.

Definition 3 (Multivariate Hermite polynomial). *Let $\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \Sigma)$ such that $\Sigma \in \mathbb{R}^{n \times n}$ is symmetric positive definite (SPD) with joint density function $\phi(\boldsymbol{\xi}; \Sigma)$ given by Eq. 3. Then for any multi-index $\boldsymbol{\alpha}$ over $[n]$, the multivariate Hermite polynomial $H_{\boldsymbol{\alpha}}(\boldsymbol{\xi}; \Sigma)$ indexed by $\boldsymbol{\alpha}$ is a polynomial in $\boldsymbol{\xi}$ of degree $|\boldsymbol{\alpha}|$ defined as*

$$H_{\boldsymbol{\alpha}}(\boldsymbol{\xi}; \Sigma) = \begin{cases} \frac{(-1)^{|\boldsymbol{\alpha}|}}{\phi(\boldsymbol{\xi}; \Sigma)} D^{|\boldsymbol{\alpha}|}_{\boldsymbol{\alpha}}(\phi(\boldsymbol{\xi}; \Sigma)) & \text{if } |\boldsymbol{\alpha}| \geq 1 \\ 1 & \text{if } |\boldsymbol{\alpha}| = 0. \end{cases}$$

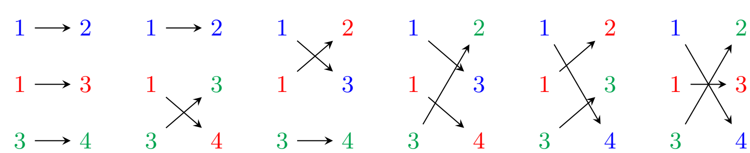
With multiset notation $s(\boldsymbol{\alpha})$, note that $H_{s(\boldsymbol{\alpha})}(\boldsymbol{\xi}; \Sigma) = H_{\boldsymbol{\alpha}}(\boldsymbol{\xi}; \Sigma)$, since $D^{|\boldsymbol{\alpha}|}_{\boldsymbol{\alpha}}$ and $D^{|s(\boldsymbol{\alpha})|}_{s(\boldsymbol{\alpha})}$ denote identical derivative operators.

To establish notation for the proof of Thm. 2, let

$$T_{\boldsymbol{\alpha}} = \{(\ell, s(\boldsymbol{\alpha})_\ell) \mid \ell \in [k]\} \tag{7}$$

be the set of k-tuples, one for each index $\ell \in [k]$. For multi-indices $\boldsymbol{\alpha}, \boldsymbol{\beta}$ of the same order k, we will consider *bijections between $T_{\boldsymbol{\alpha}}$ and $T_{\boldsymbol{\beta}}$*, i.e. the ways to pair-off the elements of $T_{\boldsymbol{\alpha}}$ and $T_{\boldsymbol{\beta}}$. As a heuristic, we can think of this as the number of ways to draw lines between the "entries" of $s(\boldsymbol{\alpha})$ and $s(\boldsymbol{\beta})$ such that each entry has a unique partner and all entries are covered. For instance, for $s(\boldsymbol{\alpha}) = [1, 1, 5]$ and $s(\boldsymbol{\beta}) = [1, 2, 4]$, we have the $3! = 6$ options depicted in Figure 1. Observe how the copies of entries (e.g. 1 in $s(\boldsymbol{\alpha}) = [1, 1, 5]$) are treated as distinct when drawing these lines; in particular, the number of such pairings for $|\boldsymbol{\alpha}| = |\boldsymbol{\beta}| = k$ will always be $k!$.

Fig. 1: All possible bijections between $T_{\boldsymbol{\alpha}}$ and $T_{\boldsymbol{\beta}}$ for $\boldsymbol{\alpha} = (2, 0, 1, 0)$ and $\boldsymbol{\beta} = (0, 1, 1, 1)$. Both $\boldsymbol{\alpha}, \boldsymbol{\beta}$ are multi-indices of order $k = 3$ over $[4]$, with $s(\boldsymbol{\alpha}) = [1, 1, 3]$ and $s(\boldsymbol{\beta}) = [2, 3, 4]$.



ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77980-1_53

If we are imagining pairings of the form $(s(\boldsymbol{\alpha})_\ell,\, s(\boldsymbol{\beta})_j)$, why discuss bijections between $T_{\boldsymbol{\alpha}}$ and $T_{\boldsymbol{\beta}}$ rather than bijections between the $s(\boldsymbol{\alpha})$ and $s(\boldsymbol{\beta})$ outputs directly? There are several reasons. For one, we would like to have "repeated" mappings counted with multiplicity rather than treated as single entities.[4] For instance, in Fig. 1 the first and third mappings are counted as distinct, even though they choose the same pairings $(1,2),(1,3),(3,4)$.

The bijections between $T_{\boldsymbol{\alpha}}$ and $T_{\boldsymbol{\beta}}$ are the foundation for our new and computationally efficient double-product formula (Theorem 2). As we shall see in Sect. 2, the proposed formula involves multiplying entries of the inverse covariance matrix Σ^{-1} — and the bijections between $T_{\boldsymbol{\alpha}}$ and $T_{\boldsymbol{\beta}}$ determine precisely which entries of Σ^{-1} are selected in this calculation.

2 Double Product Formulations

For both Theorems 1 and 2, let Σ^{-1} denote the known $n \times n$ SPD inverse matrix of the generic covariance Σ given in Def. 3. In this context, the second-order moments of multivariate Hermite polynomials were first proved in the comprehensive work of [15] to equal the expression in Thm. 1 below.

Theorem 1 (Proved in [15]). *Let $\boldsymbol{\theta} \in \mathbb{N}_0^{n\times n}$. Define $r(\boldsymbol{\theta})$ as the vector of row sums of $\boldsymbol{\theta}$; that is,*

$r(\boldsymbol{\theta}) = (r_1, \dots, r_n)^T$ *with* $r_i = \sum_{j=1}^n \theta_{ij} = \|\boldsymbol{\theta}_{i\bullet}\|_1 =$ ℓ_1 *norm of the ith row of* $\boldsymbol{\theta}$.

Similarly, let $c(\boldsymbol{\theta})$ be the $n \times 1$ vector such that $c_j = \|\boldsymbol{\theta}_{\bullet j}\|_1$. Then

$$\langle H_{\boldsymbol{\alpha}}, H_{\boldsymbol{\beta}} \rangle = \begin{cases} \boldsymbol{\alpha}!\boldsymbol{\beta}! \displaystyle\sum_{\substack{\boldsymbol{\theta} \in \mathbb{N}_0^{n\times n} \\ \boldsymbol{r}(\boldsymbol{\theta})=\boldsymbol{\alpha}, \boldsymbol{c}(\boldsymbol{\theta})=\boldsymbol{\beta}}} \frac{\prod_{p=1}^n \prod_{q=1}^n \left(\Sigma_{pq}^{-1}\right)^{\theta_{pq}}}{\boldsymbol{\theta}!}, & \text{if } |\boldsymbol{\alpha}| = |\boldsymbol{\beta}| \\ 0 & \text{else.} \end{cases}$$

where $\boldsymbol{\theta}! = \prod_{i=1}^n \prod_{\ell=1}^n \theta_{i\ell}!$.

We now propose a novel evaluation of the double product based on the multiset notation introduced in Def. 2.

Theorem 2. *Let $\boldsymbol{\alpha}, \boldsymbol{\beta}$ be multi-indices over $[n]$ with $|\boldsymbol{\alpha}| = |\boldsymbol{\beta}| = k \geq 1$. Then*

$$\langle H_{\boldsymbol{\alpha}}, H_{\boldsymbol{\beta}} \rangle = \sum_{p \in S_k} \prod_{i=1}^k \Sigma^{-1}_{s(\boldsymbol{\alpha})_i,\, s(\boldsymbol{\beta})_{p(i)}}$$

where $\Sigma^{-1}_{s(\boldsymbol{\alpha})_i,\, s(\boldsymbol{\beta})_{p(i)}}$ is the $\big(s(\boldsymbol{\alpha})_i,\, s(\boldsymbol{\beta})_{p(i)}\big)$th entry of Σ^{-1} and S_k is the symmetric group on $\{1, \dots, k\}$, recalling the multiset index notation $s(\boldsymbol{\alpha})$ given in Def. 2. If $|\boldsymbol{\alpha}| \neq |\boldsymbol{\beta}|$, then $\langle H_{\boldsymbol{\alpha}}, H_{\boldsymbol{\beta}} \rangle = 0$.

[4] Accordingly, we introduce $s(\boldsymbol{\alpha})$ as multiset *notation* rather than a *literal* multiset. An underlying philosophy of multisets is that copies of elements cannot be picked out or distinguished by (say) an indexing convention [8,10,16]. For our purposes, however, we want to treat such copies as distinct.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77980-1_53

Note that Thm. 2 only considers when $k > 0$, since the case $k = 0$ is trivial. To explore and illustrate the differences between these two formulas for the double product, we provide the following Ex. 1.

Example 1. Let $\boldsymbol{\alpha} = (2, 0, 1, 0)$ and $\boldsymbol{\beta} = (0, 1, 1, 1)$ over $[4]$, as they were in Fig. 1. To use Thm. 1, we are searching for $\boldsymbol{\theta} \in \mathbb{N}_0^{4\times 4}$ such that $r(\boldsymbol{\theta}) = (2, 0, 1, 0)$ and $c(\boldsymbol{\theta}) = (0, 1, 1, 1)$. Let $\boldsymbol{r}^{(i)}$ denote the ith row of $\boldsymbol{\theta}$. For the constraint $r(\boldsymbol{\theta}) = (2, 0, 1, 0)$, noting that rows 2 and 4 must be zeroes, we have $\binom{4+2-1}{2} = 10$ options for $\boldsymbol{r}^{(\mathbf{1})}$ and $\binom{4+1-1}{2} = 3$ options for $\boldsymbol{r}^{(\mathbf{3})}$ [18]. Naively, we could then check all $10 \times 3 = 30$ options for $\boldsymbol{\theta}$ and eliminate those that fail to satisfy the column constraint $c(\boldsymbol{\theta}) = \boldsymbol{\beta}$. To be clever, we can eliminate the $\boldsymbol{r}^{(\boldsymbol{i})}$ along the way whose entries $\boldsymbol{r}_j^{(\boldsymbol{i})} \geq \beta_j$, since these rows guarantee that some columns sums in $\boldsymbol{\theta}$ will be too large — which is indicated by the slashes in Fig. 2. Hence, we have $3 \times 3 = 9$ initial matrices $\boldsymbol{\theta}$ to iterate through to find those such that $c(\boldsymbol{\theta}) = \boldsymbol{\beta}$, from which there are 3 final candidates (Fig. 2).

Fig. 2: Determining the possible $\boldsymbol{\theta} \in \mathbb{N}_0^{4\times 4}$ such that $r(\boldsymbol{\theta}) = (2, 0, 1, 0)$ and $c(\boldsymbol{\theta}) = (0, 1, 1, 1)$ in Ex. 1 to use Thm. 1. First we find the options for $\boldsymbol{r}^{(\mathbf{1})}$ and $\boldsymbol{r}^{(\mathbf{3})}$ (rows 1 and 3) of $\boldsymbol{\theta}$ such that $\boldsymbol{r}^{(\mathbf{1})}$ sums to 2 and $\boldsymbol{r}^{(\mathbf{3})}$ sums to 1, eliminating the options along the way that have an entry $\boldsymbol{r}_j^{(\boldsymbol{i})} \geq \beta_j$, as indicated by the slashes. The $3\times 3 = 9$ possible $\boldsymbol{\theta}$ are iterated over to see which satisfy the column sums constraint $c(\boldsymbol{\theta}) = (0, 1, 1, 1)$. This leaves the 3 matrices denoted by $\boldsymbol{\theta}^{(\mathbf{1})}, \boldsymbol{\theta}^{(\mathbf{2})}, \boldsymbol{\theta}^{(\mathbf{3})}$ below.

$\boldsymbol{r}^{(\mathbf{1})}$ options: ~~(2, 0, 0, 0)~~, ~~(0, 2, 0, 0)~~, ~~(0, 0, 2, 0)~~, ~~(0, 0, 0, 2)~~, ~~(1, 1, 0, 0)~~, ~~(1, 0, 1, 0)~~, ~~(1, 0, 0, 1)~~, (0, 1, 1, 0), (0, 1, 0, 1), (0, 0, 1, 1)

$\boldsymbol{r}^{(\mathbf{3})}$ options: (0, 1, 0, 0), (0, 0, 1, 0), (0, 0, 0, 1).

$$\cancel{\begin{pmatrix}0&1&1&0\\0&0&0&0\\0&1&0&0\\0&0&0&0\end{pmatrix}}, \cancel{\begin{pmatrix}0&1&1&0\\0&0&0&0\\0&0&1&0\\0&0&0&0\end{pmatrix}}, \overbrace{\begin{pmatrix}0&1&1&0\\0&0&0&0\\0&0&0&1\\0&0&0&0\end{pmatrix}}^{\boldsymbol{\theta}^{(\mathbf{1})}} \quad \cancel{\begin{pmatrix}0&1&0&1\\0&0&0&0\\0&1&0&0\\0&0&0&0\end{pmatrix}}, \overbrace{\begin{pmatrix}0&1&0&1\\0&0&0&0\\0&0&1&0\\0&0&0&0\end{pmatrix}}^{\boldsymbol{\theta}^{(\mathbf{2})}}, \cancel{\begin{pmatrix}0&1&0&1\\0&0&0&0\\0&0&0&1\\0&0&0&0\end{pmatrix}} \quad \overbrace{\begin{pmatrix}0&0&1&1\\0&0&0&0\\0&1&0&0\\0&0&0&0\end{pmatrix}}^{\boldsymbol{\theta}^{(\mathbf{3})}}, \cancel{\begin{pmatrix}0&0&1&1\\0&0&0&0\\0&0&1&0\\0&0&0&0\end{pmatrix}}, \cancel{\begin{pmatrix}0&0&1&1\\0&0&0&0\\0&0&0&1\\0&0&0&0\end{pmatrix}}$$

Each of these $\boldsymbol{\theta}$ satisfies $\boldsymbol{\theta}! = 1$. Compute

$$\langle H_{\boldsymbol{\alpha}}, H_{\boldsymbol{\beta}} \rangle = 2[\Sigma^{-1}_{12}\Sigma^{-1}_{13}\Sigma^{-1}_{34} + \Sigma^{-1}_{12}\Sigma^{-1}_{14}\Sigma^{-1}_{33} + \Sigma^{-1}_{13}\Sigma^{-1}_{14}\Sigma^{-1}_{23}],$$

where we tacitly used that Σ^{-1} is symmetric.

By using the formulation in Thm. 2 instead, we have the $3! = 6$ terms to consider from the start, one for each $p \in S_3$. Once a computer obtains these S_3 entries, which often is elementary and trivially fast to do,[5] the $\Sigma^{-1}_{(s(\boldsymbol{\alpha})_i,\, s(\boldsymbol{\beta})_{p(i)})}$ can be evaluated directly. Alternatively, we evaluate Σ^{-1} at the pairs matched by the mappings drawn in Fig. 1. From the final row in Table 1, which sums the entries of the previous rows, we yield the same $\langle H_{\boldsymbol{\alpha}}, H_{\boldsymbol{\beta}} \rangle$ as was found with Thm. 1 previously.

What happens to the product in Thm. 2 when the ξ_i are uncorrelated? In this case, every Σ^{-1}_{ij} in which $i \neq j$ equals zero. *If* $\boldsymbol{\alpha} \neq \boldsymbol{\beta}$, then it is easy to show that for every $p \in S_k$ there is at least one $\ell \in [k]$ for which $s(\boldsymbol{\alpha})_\ell \neq s(\boldsymbol{\beta})_{p(\ell)}$. As expected, then Thm. 2 gives that $\mathbb{E}(H_{\boldsymbol{\alpha}} H_{\boldsymbol{\beta}}) = 0$.

[5] For instance, in `Python3`, the `combinatorics` module in `itertools` [20] suffices.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77980-1_53

Table 1: Using Thm. 2 to compute $\langle H_{\boldsymbol{\alpha}}, H_{\boldsymbol{\beta}}\rangle$ in Ex. 1. For each permutation p on $\{1,2,3\}$, we find the tuples $(s(\boldsymbol{\alpha})_i, s(\boldsymbol{\beta})_{p(i)})$ for $s(\boldsymbol{\alpha}) = [1,1,3]$ and $s(\boldsymbol{\beta}) = [2,3,4]$. Then the product $\prod_{i=1}^{3} \Sigma^{-1}_{s(\boldsymbol{\alpha})_i, s(\boldsymbol{\beta})_{p(i)}}$ is evaluated at these tuples. Equivalently, the $(s(\boldsymbol{\alpha})_i, s(\boldsymbol{\beta})_{p(i)})$ are precisely the pairings shown in the maps of Fig. 1 (left to right), matched by color to show correspondence across the two figures.

$p \in S_3$	$(s(\boldsymbol{\alpha})_i, s(\boldsymbol{\beta})_{p(i)})\ \forall_{i\in[3]}$, i.e. pairings in Fig. 1	$\prod_{i=1}^{3} \Sigma^{-1}_{s(\boldsymbol{\alpha})_i, s(\boldsymbol{\beta})_{p(i)}}$
(123)	$(1,2), (1,3), (3,4)$	$\Sigma^{-1}_{12}\Sigma^{-1}_{13}\Sigma^{-1}_{34}$
(132)	$(1,2), (1,4), (3,3)$	$\Sigma^{-1}_{12}\Sigma^{-1}_{14}\Sigma^{-1}_{33}$
(213)	$(1,3), (1,2), (3,4)$	$\Sigma^{-1}_{12}\Sigma^{-1}_{13}\Sigma^{-1}_{34}$
(231)	$(1,3), (1,4), (3,2)$	$\Sigma^{-1}_{13}\Sigma^{-1}_{14}\Sigma^{-1}_{23}$
(312)	$(1,4), (1,2), (3,3)$	$\Sigma^{-1}_{12}\Sigma^{-1}_{14}\Sigma^{-1}_{33}$
(321)	$(1,4), (1,3), (3,2)$	$\Sigma^{-1}_{13}\Sigma^{-1}_{14}\Sigma^{-1}_{23}$
$\langle H_{\boldsymbol{\alpha}}, H_{\boldsymbol{\beta}}\rangle$	$2[\Sigma^{-1}_{12}\Sigma^{-1}_{13}\Sigma^{-1}_{34} + \Sigma^{-1}_{12}\Sigma^{-1}_{14}\Sigma^{-1}_{33} + \Sigma^{-1}_{13}\Sigma^{-1}_{14}\Sigma^{-1}_{23}]$	

3 Comparing Computational Complexity

We assume that the inverse covariance matrix Σ^{-1} and $\boldsymbol{\alpha}, \boldsymbol{\beta}$ of order k are given. To use Thm. 1 to compute a single double product, the computation time is dominated by producing all the $\boldsymbol{\theta} \in \mathbb{N}_0^n$ such that $\boldsymbol{r}(\boldsymbol{\theta}) = \boldsymbol{\alpha}, \boldsymbol{c}(\boldsymbol{\theta}) = \boldsymbol{\beta}$.[6] A reasonable, albeit naive, algorithm for finding all such $\boldsymbol{\theta}$ is to first generate the possible $\boldsymbol{\theta}$ such that $r(\boldsymbol{\theta}) = \boldsymbol{\alpha}$ and then eliminate those which do not satisfy the column constraint; this was the process taken in Ex. 1. Along the way, perhaps we can eliminate possibilities for the ith row $\boldsymbol{r}^{(i)}$ based on whether $\boldsymbol{r}^{(i)}_j \leq \beta_j$, but in the worst case scenario, none of the possibilities for *any* of the $\boldsymbol{r}^{(i)}$ can be discarded based on $\boldsymbol{\beta}$. We do not claim that this procedure is the most efficient computation of $\langle H_{\boldsymbol{\alpha}}, H_{\boldsymbol{\beta}}\rangle$ via Thm. 1 — but we will use it as the straightforward benchmark for comparison against computing the double product via Thm. 2.

Before considering column constraints,

$$\begin{aligned}\#\text{ of options for row } i &= \#\text{ of } n\text{-tuples whose entries sum to } \alpha_i \\ &= \tbinom{n+\alpha_i-1}{n-1}\end{aligned}$$

by [8,10,15,16,18]. Following §4.5.1 in [16], there exist algorithms that output all of the options for $\boldsymbol{r}^{(i)}$ with computational complexity proportional to the number of options, i.e. $\binom{n+\alpha_i-1}{n-1}$. Repeating for all of the rows, there are at least

$$\tbinom{n+\alpha_1-1}{n-1}\cdots\tbinom{n+\alpha_n-1}{n-1} = \frac{(n-1)!^n \overbrace{[(n)(n+1)\cdots(n+\alpha_1-1)]}^{\alpha_1 \text{ terms}} \cdots \overbrace{[(n)(n+1)\cdots(n+\alpha_n-1)]}^{\alpha_n \text{ terms}}}{(n-1)!^n \boldsymbol{\alpha}!}$$

$$\geq \frac{n^{\alpha_1}\cdots n^{\alpha_n}}{\boldsymbol{\alpha}!} = \frac{n^{\sum_i \alpha_i}}{\boldsymbol{\alpha}!} = \frac{n^k}{\boldsymbol{\alpha}!} \geq \frac{n^k}{k!} \qquad [\text{since } \max_{|\boldsymbol{\alpha}|=k} \boldsymbol{\alpha}! = k!]$$

[6] Counting the number of such index matrices, which are often called *contingency tables with fixed margins* in statistics literature, is well-studied [3,6] and can be done in poly(n) time [4]. This does not mean that the *number* of contingency tables is poly(n) but that algorithms can produce the total count of them in poly(n) time.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77980-1_53

options for $\boldsymbol{\theta} \in \mathbb{N}_0^{n \times n}$ in Eq. 1 such that $r(\boldsymbol{\theta}) = \boldsymbol{\alpha}$. Thus, producing the necessary $\boldsymbol{\theta}$ to sum over in Eq. 1 involves iterating over at least $\frac{n^k}{k!}$ matrices in terms of asymptotic complexity. Per $\boldsymbol{\theta}$, computing $\prod_{p=1}^{n} \prod_{q=1}^{n} (\Sigma_{pq}^{-1})^{\theta_{pq}}$ involves a total of at least $\sum_{p,q=1}^{n} \theta_{p,q} = k$ multiplications. Then the computational complexity of implementing Eq. 1 in this direct manner is $\Omega(\frac{n^k}{(k-1)!})$, which is exponential in k.

When computing the double product via Thm. 2, there are $k!$ terms in the summation, and each summand is the product of k entries of Σ^{-1}. So the cost for computing $\langle H_{\boldsymbol{\alpha}}, H_{\boldsymbol{\beta}} \rangle$ in this case is factorial in k, namely $O(kk!) = O(k!)$, for $\boldsymbol{\alpha}, \boldsymbol{\beta}$ of order k.

4 Conclusion

Polynomial chaos (PC) expansions are effective for incorporating and quantifying uncertainties in problems governed by partial differential equations. In some contexts, they offer significant computational advantages to classic Monte Carlo sampling methods (for example) [2,15], whose converge rates are especially hindered when each sample evaluation of the PDE is expensive [2,12]. However, when multiple input uncertainties are considered without transformations, PC approaches cannot be generalized in a simple fashion unless the uncertainties are represented in terms of independent variables. Unlike when $\boldsymbol{\xi}$ is one dimensional or uncorrelated, many of the double product coefficients $\langle H_{\boldsymbol{\alpha}}, H_{\boldsymbol{\beta}} \rangle := \mathbb{E}_{\boldsymbol{\xi}}(H_{\boldsymbol{\alpha}} H_{\boldsymbol{\beta}})$ that appear from the Galerkin projections are nonzero and therefore *essential* to compute a priori in order to solve the resulting system numerically.

In this paper, we prove a new formula (Thm. 2) for the double product of two multivariate Hermite polynomials whose n-dimensional input Gaussian random variable has an arbitrary SPD covariance matrix. To do so, we introduce what we call *multiset notation* (Def. 2) for the label indices $\boldsymbol{\alpha}$. Calculating the double product is computationally more efficient and (arguably) simpler with the proposed approach than doing so with the classical formula in [15] given by Thm. 1. In particular, Sect. 3 analyzes the computational complexity of the two formulations; the implementations considered for each were purposely straightforward and already showcase the reduced cost achieved by the use of the multiset notation.

From the foundational work in this paper, the authors plan to explore the triple product $\langle H_{\boldsymbol{\alpha}}, H_{\boldsymbol{\beta}}, H_{\boldsymbol{\gamma}} \rangle$ calculations in terms of these double product constituents. As demonstrated in inviscid Burgers' equation (Eq. 5), the triple products can arise when the original PDE has quadratic *nonlinear* terms. Establishing these triple product values will be a pivotal building block for handling nonlinear PDEs that incorporate uncertainties in a general setting.

5 Appendix A: Proof of Theorem 2

The proof of Thm. 2 relies on the following Thms. 3 and 4. In the following discussion, assume $\boldsymbol{\alpha}$ is a multi-index over $[n]$ of order $k \geq 1$ unless otherwise specified.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77980-1_53

Recall the definition of $T_{\boldsymbol{\alpha}}$ in Eq. 7 in Sect. 1.2. We assign an ordering to the elements of $T_{\boldsymbol{\alpha}}$ (or any subset of $T_{\boldsymbol{\alpha}}$) based on their first components ascending. That is, $(T_{\boldsymbol{\alpha}})_1 = (1, s(\boldsymbol{\alpha})_1), \ldots, (T_{\boldsymbol{\alpha}})_k = (k, s(\boldsymbol{\alpha})_k)$, and when $A \subseteq T_{\boldsymbol{\alpha}}$, we label $A_1 = (\ell_1, s(\boldsymbol{\alpha})_{\ell_1}), \ldots, A_{|A|} = (\ell_{|A|}, s(\boldsymbol{\alpha})_{\ell_{|A|}})$ such that $\ell_1 < \cdots < \ell_{|A|}$. Before fretting about the specifics, realize that this ordering follows intuition. For example, if $\boldsymbol{\alpha} = (2, 1, 0, 0, 1)$, then $s(\boldsymbol{\alpha}) = [1, 1, 2, 5]$, and $T_{\boldsymbol{\alpha}} = \{(1, 1), (2, 1), (3, 2), (4, 5)\}$. Now, $(T_{\boldsymbol{\alpha}})_1 = (1, 1), (T_{\boldsymbol{\alpha}})_2 = (2, 1), (T_{\boldsymbol{\alpha}})_3 = (3, 2)$, and $(T_{\boldsymbol{\alpha}})_4 = (4, 5)$. For the subset $A = T_{\boldsymbol{\alpha}} \setminus \{(2, 1)\} = \{(1, 1), (3, 2), (4, 5)\}$ of $T_{\boldsymbol{\alpha}}$, we have $A_1 = (1, 1), A_2 = (3, 2)$, and $A_3 = (4, 5)$. In fact, we will use the shorthand

$$(T_{\boldsymbol{\alpha}})^{-j} := T_{\boldsymbol{\alpha}} \setminus \{\, (j, s(\boldsymbol{\alpha})_j)\, \} \qquad \text{for any } j \in [k], \tag{8}$$

where $(T_{\boldsymbol{\alpha}})^{-j}_{\ell}$ is the ℓth element of $(T_{\boldsymbol{\alpha}})^{-j}$ according to this ordering by first components ascending.

For any $i \in [n]$ such that $\alpha_i > 0$, define

$$i^* = \min\left(s(\boldsymbol{\alpha})^{-1}(\{i\})\right) = \min\{\ell \in [k] \mid s(\boldsymbol{\alpha})_\ell = i\}. \tag{9}$$

By Lemma 2, we have

$$(T_{\boldsymbol{\alpha}})^{-i^*}_{\ell} = \begin{cases} (\ell, s(\boldsymbol{\alpha})_\ell) & \text{if } \ell < i^* \\ (\ell + 1, s(\boldsymbol{\alpha})_{\ell+1}) & \text{else} \end{cases} = \begin{cases} (\ell, s(\boldsymbol{\alpha} - \boldsymbol{e_i})_\ell) & \text{if } \ell < i^* \\ (\ell + 1, s(\boldsymbol{\alpha} - \boldsymbol{e_i})_\ell) & \text{else.} \end{cases}$$

Note that the elements of $T_{\boldsymbol{\alpha}-\boldsymbol{e_i}}$ have the form $(\ell, s(\boldsymbol{\alpha} - \boldsymbol{e_i})_\ell)$ for $\ell \in [k-1]$. Therefore, *the second coordinate of the ℓth element of $(T_{\boldsymbol{\alpha}})^{-i^*}_{\ell}$ is identical to the second coordinate of the ℓth element of $T_{\boldsymbol{\alpha}-\boldsymbol{e_i}}$.*

Define the projector operator

$$\text{proj} : [k] \times [n] \to [n] \qquad \text{proj}(t_1, t_2) = t_2 \tag{10}$$

that simply ignores the first coordinate of its input. Then for any $\ell \in [k-1]$ and $i \in [n]$ such that $\alpha_i > 0$,

$$\text{proj}((T_{\boldsymbol{\alpha}-\boldsymbol{e_i}})_\ell) = \text{proj}((T^{-i^*}_{\boldsymbol{\alpha}})_\ell), \qquad i^* = \min s(\boldsymbol{\alpha})^{-1}(\{i\}). \tag{11}$$

Equation 11, along with the definitions in Equations 7, 8, 10, will be utilized in the proof of Theorem 3.

Theorem 3. *Let $\boldsymbol{\alpha}, \boldsymbol{\beta}$ be two multi-indices over $[n]$ such that $|\boldsymbol{\alpha}| = |\boldsymbol{\beta}| = k \geq 1$. Then*

$$\frac{\partial^{|\boldsymbol{\beta}|} H_{\boldsymbol{\alpha}}}{\partial \xi_{\boldsymbol{\beta}}} = \sum_{p \in S_k} \prod_{i=1}^{k} \Sigma^{-1}_{s(\boldsymbol{\alpha})_i,\, s(\boldsymbol{\beta})_{p(i)}}.$$

Proof. We proceed by induction on k. When $k = 1$, $\boldsymbol{\alpha} = \boldsymbol{e_r}$ and $\boldsymbol{\beta} = \boldsymbol{e_s}$ for some $r, s \in [n]$, so the base case is proved by Lemma 4. For the inductive step, let $j = \min\{i \in [n] \mid \beta_i > 0\}$, where we know j exists since $|\boldsymbol{\beta}| = k \geq 1$. Let $g_{\boldsymbol{\alpha},j,\Sigma}$ be a function on $[n]$ such that $g_{\boldsymbol{\alpha},j,\Sigma}(i) = \Sigma^{-1}_{ij} H_{\boldsymbol{\alpha}-\boldsymbol{e_i}}$. Then

$$\begin{aligned} \frac{\partial}{\partial \xi_j}(H_{\boldsymbol{\alpha}}) &= \sum_{i=1}^{n} \alpha_i \Sigma^{-1}_{ij} H_{\boldsymbol{\alpha}-\boldsymbol{e_i}} = \sum_{i=1}^{n} \alpha_i\, g_{\boldsymbol{\alpha},j,\Sigma}(i) && [\text{Lem. 5 \& def. of } g_{\boldsymbol{\alpha},j,\Sigma}] \\ &= \sum_{i=1}^{k} g_{\boldsymbol{\alpha},j,\Sigma}(s(\boldsymbol{\alpha})_i) = \sum_{i=1}^{k} \Sigma^{-1}_{s(\boldsymbol{\alpha})_i,j} H_{\boldsymbol{\alpha}-\boldsymbol{e}_{s(\boldsymbol{\alpha})_i}} && [\text{Lem. 6}]. \end{aligned}$$

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77980-1_53

Substituting,

$$D^k_{\boldsymbol{\beta}}(H_{\boldsymbol{\alpha}}) = D^{k-1}_{\boldsymbol{\beta}-\boldsymbol{e}_j}\frac{\partial}{\partial \xi_j}(H_{\boldsymbol{\alpha}}) = \sum_{i=1}^{k} \Sigma^{-1}_{s(\boldsymbol{\alpha})_i,j} D^{k-1}_{\boldsymbol{\beta}-\boldsymbol{e}_j}\left(H_{\boldsymbol{\alpha}-\boldsymbol{e}_{s(\boldsymbol{\alpha})_i}}\right)$$
$$= \sum_{i=1}^{k} \Sigma^{-1}_{s(\boldsymbol{\alpha})_i,j} \sum_{p\in S_{k-1}} \prod_{\ell=1}^{k-1} \Sigma^{-1}_{s(\boldsymbol{\alpha}-\boldsymbol{e}_{s(\boldsymbol{\alpha})_i})_\ell,\, s(\boldsymbol{\beta}-\boldsymbol{e}_j)_{p(\ell)}} \qquad \text{[ind. hypothesis]}.$$

Let

$$i^* = \min\{\ell \in [k] \mid s(\boldsymbol{\alpha})_\ell = s(\boldsymbol{\alpha})_i\}, \quad j^* = \{\ell \in [k] \mid s(\boldsymbol{\beta})_\ell = j\}.$$

For all $\ell \in [k-1]$, Equations 7, 8, 10, and 11 give

$$\operatorname{proj}\left((T^{-i^*}_{\boldsymbol{\alpha}})_\ell\right) = \operatorname{proj}\left((T_{\boldsymbol{\alpha}-\boldsymbol{e}_{s(\boldsymbol{\alpha})_i}})_\ell\right) = s(\boldsymbol{\alpha}-\boldsymbol{e}_{s(\boldsymbol{\alpha})_i})_\ell$$
$$\operatorname{proj}\left((T^{-j^*}_{\boldsymbol{\beta}})_\ell\right) = \operatorname{proj}\left((T_{\boldsymbol{\beta}-\boldsymbol{e}_j})_\ell\right) = s(\boldsymbol{\alpha}-\boldsymbol{e}_j)_\ell.$$

Let $h : T_{\boldsymbol{\alpha}} \times T_{\boldsymbol{\beta}} \to \mathbb{R}$ such that $h(\boldsymbol{r},\boldsymbol{t}) = \Sigma^{-1}_{\operatorname{proj}(\boldsymbol{r}),\operatorname{proj}(\boldsymbol{t})}$, noting that $T^{-i^*}_{\boldsymbol{\alpha}} \subset T_{\boldsymbol{\alpha}}$ and $T^{-j^*}_{\boldsymbol{\beta}} \subset T_{\boldsymbol{\beta}}$. Then

$$D^k_{\boldsymbol{\beta}}(H_{\boldsymbol{\alpha}}) = \sum_{i=1}^{k} \Sigma^{-1}_{s(\boldsymbol{\alpha})_i,j} \sum_{p\in S_{k-1}} \prod_{\ell=1}^{k-1} h\left((T^{-i^*}_{\boldsymbol{\alpha}})_\ell, (T^{-j^*}_{\boldsymbol{\beta}})_{p(\ell)}\right) \quad \text{[def. of } h\text{]}$$
$$= \sum_{i=1}^{k} \Sigma^{-1}_{s(\boldsymbol{\alpha})_i,j} \sum_{b:T^{-i^*}_{\boldsymbol{\alpha}} \hookrightarrow\!\!\!\!\to T^{-j^*}_{\boldsymbol{\beta}}} \prod_{\boldsymbol{t}\in T^{-i^*}_{\boldsymbol{\alpha}}} h(\boldsymbol{t}, b(\boldsymbol{t})) \qquad \text{[by Lemma 7, Eq. 12]}$$

where we can match notation from Eq. 12 in Lem. 7 by setting $A = T^{-i^*}_{\boldsymbol{\alpha}}$ and $B = T^{-j^*}_{\boldsymbol{\beta}}$. Now,

$$s(\boldsymbol{\alpha})_i = \operatorname{proj}(\,(i^*, s(\boldsymbol{\alpha})_i)\,) \qquad \text{[def. of proj map in Eq. 10]}$$
$$= \operatorname{proj}(\,(i^*, s(\boldsymbol{\alpha})_{i^*})\,) \qquad \text{[since } s(\boldsymbol{\alpha})_{i^*} = s(\boldsymbol{\alpha})_i \text{ by def. of } i^*\text{]}$$
$$= \operatorname{proj}((T_{\boldsymbol{\alpha}})_{i^*}) \qquad \text{[labeling of } T_{\boldsymbol{\alpha}} \text{ elements]}.$$

By a similar argument, $j = \operatorname{proj}((T_{\boldsymbol{\beta}})_{j^*})$. Therefore,

$$D^k_{\boldsymbol{\beta}}(H_{\boldsymbol{\alpha}}) = \sum_{i=1}^{k} h((T_{\boldsymbol{\alpha}})_{i^*}, (T_{\boldsymbol{\beta}})_{j^*}) \sum_{b:T^{-i^*}_{\boldsymbol{\alpha}} \hookrightarrow\!\!\!\!\to T^{-j^*}_{\boldsymbol{\beta}}} \prod_{\boldsymbol{t}\in T^{-i^*}_{\boldsymbol{\alpha}}} h(\boldsymbol{t}, b(\boldsymbol{t}))$$
$$= \sum_{b:T_{\boldsymbol{\alpha}} \hookrightarrow\!\!\!\!\to T_{\boldsymbol{\beta}}} \prod_{\boldsymbol{t}\in T_{\boldsymbol{\alpha}}} h(\boldsymbol{t}, b(\boldsymbol{t})) \qquad \text{[Lem. 7, Eq. 13]}$$
$$= \sum_{p\in S_k} \prod_{\ell=1}^{k} h((T_{\boldsymbol{\alpha}})_\ell, (T_{\boldsymbol{\beta}})_{p(\ell)}) \qquad \text{[Lem. 7, Eq. 12]}$$
$$= \sum_{p\in S_k} \prod_{\ell=1}^{k} \Sigma^{-1}_{s(\boldsymbol{\alpha})_\ell, s(\boldsymbol{\beta})_{p(\ell)}} \qquad \text{[def. of } h\text{]}$$

as desired. □

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77980-1_53

Theorem 4. *Let $\boldsymbol{\alpha}, \boldsymbol{\beta}$ be two multi-indices over $[n]$. Then*

$$\mathbb{E}(H_{\boldsymbol{\alpha}} H_{\boldsymbol{\beta}}) = \begin{cases} \frac{\partial^{|\boldsymbol{\beta}|} H_{\boldsymbol{\alpha}}}{\partial \xi_{\boldsymbol{\beta}}} & \text{if } |\boldsymbol{\alpha}| = |\boldsymbol{\beta}| \\ 0 & \text{else.} \end{cases}$$

Proof. It is proved in [15] that $|\boldsymbol{\alpha}| \neq |\boldsymbol{\beta}|$ implies that $\mathbb{E}(H_{\boldsymbol{\alpha}} H_{\boldsymbol{\beta}}) = 0$. So suppose $|\boldsymbol{\alpha}| = |\boldsymbol{\beta}| = k$, and we proceed by induction on k. The $k = 1$ base case is a straightforward consequence of Lem. 3 and the fact that the expected value of any multivariate Hermite polynomial is zero by [15]. For the inductive step, assume $|\boldsymbol{\alpha}| = |\boldsymbol{\beta}| = k$. Let $j = \min\{\ell \in [n] \mid \beta_\ell > 0\}$, where we know such a j exists since $|\boldsymbol{\beta}| > 0$. By Lemma 5 and linearity of expectation,

$$\begin{aligned}
\mathbb{E}\left(D^k_{\boldsymbol{\alpha}}(H_{\boldsymbol{\beta}})\right) &= \sum_{i=1}^{n} \alpha_i \Sigma^{-1}_{ij} \mathbb{E}\left(D^{k-1}_{\boldsymbol{\beta}-\boldsymbol{e_j}}(H_{\boldsymbol{\alpha}-\boldsymbol{e_i}})\right) \\
&= \sum_{i=1}^{n} \Sigma^{-1}_{ij} \mathbb{E}\left(H_{\boldsymbol{\alpha}-\boldsymbol{e_i}} H_{\boldsymbol{\beta}-\boldsymbol{e_j}}\right) && \text{[inductive hypothesis]} \\
&= \mathbb{E}\left(H_{\boldsymbol{\beta}-\boldsymbol{e_j}} \sum_{i=1}^{n} \alpha_i \Sigma^{-1}_{ij} H_{\boldsymbol{\alpha}-\boldsymbol{e_i}}\right) \\
&= \mathbb{E}\left(H_{\boldsymbol{\beta}-\boldsymbol{e_j}} D_{\boldsymbol{e_j}}(H_{\boldsymbol{\alpha}})\right) && \text{[Lem. 5]} \\
&= \mathbb{E}\left(H_{\boldsymbol{\beta}-\boldsymbol{e_j}} H_{\boldsymbol{\alpha}} H_{\boldsymbol{e_j}}\right) - \mathbb{E}\left(H_{\boldsymbol{\beta}-\boldsymbol{e_j}} H_{\boldsymbol{\alpha}+\boldsymbol{e_j}}\right) && \text{[Lem. 3 for } H_{\boldsymbol{\alpha}+\boldsymbol{e_j}}\text{].}
\end{aligned}$$

From $|\boldsymbol{\beta} - \boldsymbol{e_j}| \neq |\boldsymbol{\alpha} + \boldsymbol{e_j}|$, $\mathbb{E}(H_{\boldsymbol{\beta}-\boldsymbol{e_j}} H_{\boldsymbol{\alpha}+\boldsymbol{e_j}}) = 0$. Therefore, $\mathbb{E}(D^k_{\boldsymbol{\alpha}}(H_{\boldsymbol{\beta}})) = \mathbb{E}(H_{\boldsymbol{\alpha}} H_{\boldsymbol{\beta}-\boldsymbol{e_j}} H_{\boldsymbol{e_j}})$. Applying Lem. 3 to $H_{\boldsymbol{\beta}}$, $\mathbb{E}(H_{\boldsymbol{\alpha}} H_{\boldsymbol{\beta}-\boldsymbol{e_j}} H_{\boldsymbol{e_j}}) = \mathbb{E}(H_{\boldsymbol{\alpha}} H_{\boldsymbol{\beta}}) + \mathbb{E}(H_{\boldsymbol{\alpha}} D_{\boldsymbol{e_j}}(H_{\boldsymbol{\beta}-\boldsymbol{e_j}}))$. By Lem. 5, we know that $D_{\boldsymbol{e_j}}(H_{\boldsymbol{\beta}-\boldsymbol{e_j}})$ is a linear combination of polynomials of the form $H_{\boldsymbol{\beta}-\boldsymbol{e_j}-\boldsymbol{e_r}}$. Hence, $\mathbb{E}(H_{\boldsymbol{\alpha}} D_{\boldsymbol{e_j}}(H_{\boldsymbol{\beta}-\boldsymbol{e_j}}))$ is a linear combination of such terms $\mathbb{E}(H_{\boldsymbol{\alpha}} H_{\boldsymbol{\gamma}})$ for $|\boldsymbol{\alpha}| \neq |\boldsymbol{\gamma}|$, each of which is zero. Thus, $\mathbb{E}(D^k_{\boldsymbol{\beta}}(H_{\boldsymbol{\alpha}})) = \mathbb{E}\left(H_{\boldsymbol{\alpha}} H_{\boldsymbol{\beta}}\right)$. Finally, we know from Thm. 3 that $D_{\boldsymbol{\beta}}(H_{\boldsymbol{\alpha}})$ is deterministic (since it is independent of $\boldsymbol{\xi}$), so $\mathbb{E}(H_{\boldsymbol{\alpha}} H_{\boldsymbol{\beta}}) = \mathbb{E}(D^k_{\boldsymbol{\beta}}(H_{\boldsymbol{\alpha}})) = \frac{\partial^k H_{\boldsymbol{\alpha}}}{\partial \xi_{\boldsymbol{\beta}}}$. □

Proof (of Theorem 2). Combining Theorems 3 and 4 immediately gives the desired result.

6 Appendix B

For brevity, several proofs are omitted, but we outline them here. Lems. 1 and 2 are straightforward. Lems. 3 and 4 involve differentiating the density ϕ in Def. 3 directly. Lem. 5 is proved by induction and applying Lem. 3. Lem. 6 follows from decomposing $[k]$ into the preimage sets $s(\boldsymbol{\alpha})^{-1}(\{\ell\})$ for all $\ell \in [n]$. Lem. 7 is a specific application of an elementary combinatorial argument that regards every bijection between two sets as an extension of a bijection on two smaller subsets [11].

Lemma 1. *For multi-index $\boldsymbol{\alpha}$ over $[n]$ of order $k > 0$,*

1. *$s(\boldsymbol{\alpha})$ is non-decreasing in its indices, i.e. $s(\boldsymbol{\alpha})_\ell \leq s(\boldsymbol{\alpha})_{\ell+1}$ for all $\ell \in [k-1]$,*

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77980-1_53

2. *for fixed* $j \in [n]$ *such that* $\alpha_j > 0$, $\min s(\boldsymbol{\alpha})^{-1}(\{j\}) = \sum_{r=1}^{j-1} \alpha_r + 1$,
3. *each* $j \in [n]$ *appears* α_j *total times in* $[s(\boldsymbol{\alpha})_1, \ldots, s(\boldsymbol{\alpha})_k]$.

Lemma 2. *Let* $\boldsymbol{\alpha}$ *be a multi-index over* $[n]$ *such that* $|\boldsymbol{\alpha}| = k > 0$. *Let* $i \in [n]$ *such that* $\alpha_i > 0$. *Define* $i^* = \min\{\ell \in [k] \mid s(\boldsymbol{\alpha})_\ell = i\}$. *Then for* $\ell \in [k-1]$, $s(\boldsymbol{\alpha} - \boldsymbol{e_i})_\ell = s(\boldsymbol{\alpha})_\ell$ *if* $\ell < i^*$ *and* $s(\boldsymbol{\alpha} - \boldsymbol{e_i})_\ell = s(\boldsymbol{\alpha})_{\ell+1}$ *otherwise.*

Lemma 3. *Let* $\boldsymbol{\alpha}$ *be a multi-index over* $[n]$. *Then for any* $i \in [n]$ *such that* $\alpha_i > 0$, $H_{\boldsymbol{\alpha}} = H_{\boldsymbol{\alpha}-\boldsymbol{e_i}} H_{\boldsymbol{e_i}} - \frac{\partial}{\partial \xi_i} H_{\boldsymbol{\alpha}-\boldsymbol{e_i}}$.

Lemma 4. *Let* $\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \Sigma)$ *be* $\mathbb{R}^n$*-valued. Then for any* $\ell \in [n]$, $H_{\boldsymbol{e_\ell}}(\boldsymbol{\xi}; \Sigma) = (\Sigma^{-1})_{\ell\bullet}\, \boldsymbol{\xi}$, *where* $(\Sigma^{-1})_{\ell\bullet}$ *is the* ℓ*th row of the inverse covariance matrix* Σ, *and* $\boldsymbol{e_\ell}$ *is the* ℓ*th standard basis vector written as a multi-index. Thus,* $\frac{\partial H_{\boldsymbol{e_\ell}}(\xi;\Sigma)}{\partial \xi_j} = \Sigma^{-1}_{\ell j}$ *for any* $j \in [n]$.

Lemma 5. *Let* $\boldsymbol{\alpha}$ *be a multi-index over* $[n]$ *such that* $|\boldsymbol{\alpha}| = k \geq 1$. *Then for any* $j \in [n]$, $D^1_{\boldsymbol{e_j}}(H_{\boldsymbol{\alpha}}) = \frac{\partial H_{\boldsymbol{\alpha}}}{\partial \xi_j} = \sum_{i=1}^n \alpha_i \Sigma^{-1}_{ij} H_{\boldsymbol{\alpha}-\boldsymbol{e_i}}$.

Lemma 6. *Let* $\boldsymbol{\alpha}$ *be an order-*k *multi-index for* $k \geq 1$ *over* $[n]$. *Let* f *be a generic function of the indices* $[n]$. *Then* $\sum_{i=1}^n \alpha_i f(i) = \sum_{i=1}^k f(s(\boldsymbol{\alpha})_i)$.

Lemma 7. *Let* A, B *be finite sets such that* $|A| = |B| = k \geq 1$. *Let* $M(A, B)$ *denote the set of bijections between* A *and* B. *Then for fixed* $b \in B$ *and an arbitrary* $h : A \times B \to \mathbb{R}$,

$$\sum_{p \in S_k} \prod_{\ell=1}^k h(A_\ell, B_{p(\ell)}) = \sum_{f \in M(A,B)} \prod_{a \in A} h(a, f(a)) \tag{12}$$

$$= \sum_{\ell=1}^k h(A_\ell, b) \sum_{g \in M(A \setminus \{A_\ell\}, B \setminus \{b\})} \prod_{a \in A \setminus \{A_\ell\}} h(a, g(a)) \tag{13}$$

where S_k *is the symmetric group of permutations on* $[k] = \{1, \ldots, k\}$.

References

1. Cameron, R.H., Martin, W.T.: The orthogonal development of non-linear functionals in series of Fourier-Hermite functionals. Annals of Mathematics **48**(2), 385–392 (1947)
2. Constantine, P.: Spectral methods for parametrized matrix equations. Ph.D. thesis, Stanford University (2009)
3. Diaconis, P., Gangolli, A.: Rectangular arrays with fixed margins. In: Discrete probability and algorithms, pp. 15–41. Springer (1995)
4. Dittmer, S.: Counting linear extensions and contingency tables. Ph.D. thesis, University of California, Los Angeles (2019)
5. Ernst, Oliver G., Mugler, Antje, Starkloff, Hans-Jörg, Ullmann, Elisabeth: On the convergence of generalized polynomial chaos expansions. ESAIM: M2AN **46**(2), 317–339 (2012)
6. Gail, M., Mantel, N.: Counting the number of $r \times c$ contingency tables with fixed margins. Journal of the American Statistical Association **72**(360), 859–862 (1977)
7. Ghanem, R.G., Spanos, P.D.: Stochastic Finite Elements: A Spectral Approach. Springer-Verlag, Berlin, Heidelberg (1991)

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77980-1_53

8. Hickman, J.: A note on the concept of multiset. Bulletin of the Australian Mathematical Society **22**(2), 211–217 (1980)
9. Janson, S.: Gaussian Hilbert Spaces. Cambridge Tracts in Mathematics, Cambridge University Press (1997). https://doi.org/10.1017/CBO9780511526169
10. Knuth, D.E.: The Art of Computer Programming: A Draft of Section 7.2.1.1, Generating all n-Tuples. Addison-Wesley (06 2004)
11. Lovasz, L., Pelikan, J., Vesztergombi, K.: Discrete Mathematics: Elementary and Beyond. Springer-Verlag (2003)
12. Lyman, L., Iaccarino, G.: Extending bluff-and-fix estimates for polynomial chaos expansions. Journal of Computational Science **50**, 101287 (2021)
13. Noreddine, S., Nourdin, I.: On the gaussian approximation of vector-valued multiple integrals. Journal of Multivariate Analysis **102**(6), 1008–1017 (2011)
14. Owen, A.B.: Monte Carlo theory, methods and examples (2013), https://https://statweb.stanford.edu/~owen/mc/
15. Rahman, S.: Wiener–Hermite polynomial expansion for multivariate Gaussian probability measures. Journal of Mathematical Analysis and Applications **454**(1), 303 – 334 (2017)
16. Ruskey, F.: Combinatorial Generation. Preliminary working draft. University of Victoria, Victoria, BC, Canada, pp. 71 – 73, §4.5.1 (2003)
17. Slepian, D.: On the symmetrized kronecker power of a matrix and extensions of Mehler's formula for hermite polynomials. SIAM Journal on Mathematical Analysis **3**(4), 606–616 (1972)
18. Stanley, R.P.: Enumerative Combinatorics, Vol. 2. Cambridge University Press (2001)
19. Szegő, G.: Orthogonal Polynomials. American Math. Soc: Colloquium publ, American Mathematical Society (1975)
20. Van Rossum, G.: The Python Library Reference, release 3.8.2. Python Software Foundation (2020)
21. Wiener, N.: The homogeneous chaos. American Journal of Mathematics **60**(4), 897–936 (1938)
22. Xiu, D.: Generalized (Wiener-Askey) Polynomial Chaos. Ph.D. thesis, Brown University (2004)

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77980-1_53