

The algebra of interpolatory cubature formulæ for generic nodes

Claudia Fassino · Giovanni Pistone · Eva Riccomagno

Submitted November 12, 2018

Abstract We consider the classical problem of computing the expected value of a real function f of the d -variate random variable X using cubature formulæ. We use in synergy tools from Commutative Algebra for cubature rulæ, from elementary orthogonal polynomial theory and from Probability.

Keywords Design of experiments · Cubature formulæ · Algebraic Statistics · Orthogonal polynomials · Evaluation of expectations

1 Introduction

Consider the classical problem of computing the expected value of a real function f of the d -variate random variable X as a linear combination of its values $f(z)$ at a finite set of points $z \in \mathcal{D} \subset \mathbb{R}^d$. The general cubature problem is: determine classes of functions $f: \mathbb{R}^d \rightarrow \mathbb{R}$, finite set of n nodes $\mathcal{D} \subset \mathbb{R}^d$ and positive weights $\{\lambda_z\}_{z \in \mathcal{D}}$ such that

$$\mathbb{E}(f(X)) = \int_{\mathbb{R}^d} f(x) d\lambda(x) = \sum_{z \in \mathcal{D}} f(z) \lambda_z \quad (1)$$

where λ is the probability distribution of the random vector X . In the univariate case, $d = 1$, the set \mathcal{D} is the set of zeros of a *node polynomial*, e.g. the n -th orthogonal polynomial for λ , see e.g. [5, Sec. 1.4]. Not much is known in the multivariate case, unless the set of nodes is product of one-dimensional sets.

Claudia Fassino · Eva Riccomagno
Dipartimento di Matematica, Università di Genova,
Via Dodecaneso 35, 16141 Genova, Italy
E-mail: fassino@dim.unige.it
E-mail: riccomagno@dim.unige.it

Giovanni Pistone
Collegio Carlo Alberto, Via Real Collegio 30, 10024 Moncalieri, Italy
E-mail: giovanni.pistone@carloalberto.org

A similar setting appears in statistical Design of Experiment (DoE) where one considers a finite set of *treatments* \mathcal{D} and the experimental outputs as function of the treatment. The set of treatments and the set of nodes are both described efficiently as zeros of systems of polynomial equations, i.e. as what is called a *0-dimensional variety* in Commutative Algebra. This framework is systematic for Algebraic Statistics where tools from modern Computational Commutative Algebra are used to address problems in statistical inference and modelling, see e.g. [3, 6, 10]. In DoE the set \mathcal{D} is called a *design* and the *affine* structure of the ring of real functions on \mathcal{D} is analyzed in detail because it represents the set of real responses to treatments in \mathcal{D} . However, in the algebraic setting the *euclidean* structure, such as the computation of mean values, is missing. In algebraic design of experiment the computation of mean values has been obtained by considering very special sets called *factorial designs*, e.g. $\{+1, -1\}^d$, see e.g. [4] and [10, Ch. 5]. Note that $\{+1, -1\}$ is the zero set of the polynomial $x^2 - 1$.

The purpose of the present paper is to discuss how the above comes together by considering orthogonal polynomials. In particular, we consider algorithms from Commutative Algebra for the cubature problem in (1) by mixing tools from elementary orthogonal polynomial theory and Probability. Vice versa, Formula (1) provides an interesting interpretation of the RHS term as expected value.

We proceed by steps of increasing degree of generality. In Section 2 we consider the univariate case and take λ to admit an orthogonal system of polynomials. Let $g(x) = \prod_{z \in \mathcal{D}} (x - z)$ and by univariate division given a polynomial p there exist unique q and r such that $p = qg + r$ and r has degree smaller than the number of points in \mathcal{D} , that is the degree of g . Furthermore, r can be written as $\sum_{z \in \mathcal{D}} r(z) l_z(x)$ where l_z is the Lagrange polynomial for $z \in \mathcal{D}$. Then we show that

1. the expected values of p and r coincide if and only if the n -coefficients of the Fourier expansion of q with respect to the orthogonal polynomial system is zero,
2. the weights λ_z in (1) are the expected values of the Lagrange polynomials l_z , for $z \in \mathcal{D}$.

The case when the design \mathcal{D} is a proper subset of the zero set of the n -th orthogonal polynomial is developed in Section 4.

In Section 3, λ is a standard Gaussian probability law and \mathcal{D} the zero set of the n -th Hermite polynomial H_n . By applying Stein-Markov theory we give a representation of some Hermite polynomials, including those of degree $2n - 1$, as sum of an element in the polynomial ideal generated by H_n and of a remainder. See Theorem 4 and the following discussion, in particular Equation (8) which, unsurprisingly, is reminiscent of other formulæ for iterated Itô integrals e.g. [9, Eq. (6.4.17)]. The point is to describe a ring structure of the space generated by Hermite polynomials up to a certain order. This ring structure is essentially the aliasing on functions induced by limiting observations to \mathcal{D} . The particular form of the recurrence relationship for Hermite polynomials makes this possible and we suspect that the study of the ring structure over \mathcal{D} for other systems of orthogonal polynomials will require different tools from those we use here.

This result implies a system of equations in Theorem 5 (extended to the multidimensional case in Section 5) which gives an implicit description of design and weights via two polynomial equations. We envisage applicability of this in the choice of \mathcal{D} for suitable classes of functions but have not developed this here.

Section 6 contains our most general set-up: we restrict ourselves to product probability measures on \mathbb{R}^d but consider any set of n distinct points in \mathbb{R}^d . Then a Buchberger-Möller type of algorithm is provided that works exclusively over vector-space generated by orthogonal polynomials up to a suitable degree. It gives a generating set of the vanishing ideal of \mathcal{D} expressed in terms of orthogonal polynomials. This is used to determine sufficient and necessary conditions on a polynomial function f for which (1) holds for suitably defined weights. Furthermore, exploiting the Fourier expansion of a Gröbner basis of the vanishing ideal of \mathcal{D} , some results about the exactness of the cubature formulæ are shown. Of course it will be of interest to determine generalisations of our results to the cases where λ is not a product measure and still admits an orthogonal system of polynomials.

1.1 Basic commutative algebra

We start with some notation on polynomials: $\mathbb{R}[x]$ is the ring of polynomials with real coefficients and in the d -variables (or indeterminate) $x = (x_1, \dots, x_d)$; for $\alpha = (\alpha_1, \dots, \alpha_d)$ a d -

dimensional vector with non-negative integer entries, $x^\alpha = x_1^{\alpha_1} \dots x_d^{\alpha_d}$ indicates a monomial; τ indicates a term-ordering on the monomials in $\mathbb{R}[x]$. If $d = 1$ there is only one term ordering, this is not the case for $d \geq 2$. Designs of product form share some commonalities with the one dimension case. Because of this, term orders are not much used in standard quadrature theory. We will see that refining the division partial order to a proper term-order is actually relevant in some, but not all, multivariate cases.

The total degree of the monomials x^α is $\sum_{i=1}^d \alpha_i$. The symbol $\mathbb{R}[x]_k$ indicates the set of polynomials of total degree k and $\mathbb{R}[x]_{\leq k}$ the vector space of all polynomials of at most total degree k . Let \mathcal{D} be a finite set of distinct points in \mathbb{R}^d , λ a probability measure over \mathbb{R}^d and X a real-valued random vector with probability distribution λ so that the expected value of the random variable $f(X)$ is $\mathbb{E}(f(X)) = \int f(x) d\lambda(x)$.

Given a term ordering τ , let $f_1, \dots, f_t \in \mathbb{R}[x]$ form a *Gröbner basis* with respect to τ (see [2, Ch. 2]) of the ideal $\mathcal{I}(\mathcal{D})$ of polynomials vanishing over \mathcal{D} . For each $p \in \mathbb{R}[x]$ there exist $h_i, i = 1, \dots, t$ and an unique $r \in \mathbb{R}[x]$ such that

$$p(x) = \sum_{i=1}^t h_i(x) f_i(x) + r(x), \quad (2)$$

and r has its largest term in τ not divisible by the largest term of $f_i, i = 1, \dots, t$. Note that the $h_i(x) \in \mathbb{R}[x]$ are *not* necessarily unique. For all $p \in \mathbb{R}[x]$, the polynomial r above is referred to as *remainder* or *normal form*. It is often indicated with the symbol $\text{NF}_\tau(p, \{f_1, \dots, f_t\})$, or the shorter version $\text{NF}(p)$, while $\langle f_1, \dots, f_t \rangle$ indicates the polynomial ideal generated by f_1, \dots, f_t . Moreover, monomials not divisible by the largest terms of $f_i, i = 1, \dots, t$, form a vector basis of monomial functions for the vector space $\mathcal{L}(\mathcal{D})$ of real functions on \mathcal{D} . The polynomials $g = \sum_{i=1}^t h_i f_i$ and r in (2) are fundamental in the applications of Algebraic Geometry to finite spaces. Various general purpose softwares, including Maple, Mathematica, Matlab and computer algebra softwares, like CoCoA, Macaulay, Singular, allow manipulation with polynomial ideals, in particular compute remainders and monomial bases.

The polynomial $r(x)$ in (2) can be written uniquely as

$$r(x) = \sum_{z \in \mathcal{D}} p(z) l_z(x) \quad (3)$$

where l_z is the indicator polynomial of the point z in \mathcal{D} , i.e. for $x \in \mathcal{D}$ it is $l_z(x) = 1$ if $x = z$ and $l_z(x) = 0$ if $x \neq z$. Equation (3) follows from the fact that $\{l_z : z \in \mathcal{D}\}$ is a \mathbb{R} -vector space basis of $\mathcal{L}(\mathcal{D})$.

The expected value of the random polynomial function $p(X)$ with respect of λ is

$$\mathbb{E}(p(X)) = \mathbb{E}(g(X)) + \mathbb{E}(r(X)) = \mathbb{E}(g(X)) + \sum_{z \in \mathcal{D}} p(z) \mathbb{E}(l_z(X))$$

by linearity. In this paper we discuss classes of polynomials p and design points \mathcal{D} for which

$$\mathbb{E}(p(X)) = \sum_{z \in \mathcal{D}} p(z) \mathbb{E}(l_z(X))$$

equivalently $\mathbb{E}(g(X)) = 0$.

In one dimension, the polynomial f vanishing over \mathcal{D} and of degree $n = |\mathcal{D}|$ forms a Gröbner basis for $\mathcal{S}(\mathcal{D})$. Here $|A|$ indicates the number of elements of a set A . Furthermore, r satisfies three main properties:

1. r is a polynomial of degree less or equal to $n - 1$,
2. $p(x) = g(x) + r(x) = q(x)f(x) + r(x)$ for a suitable $q \in \mathbb{R}[x]$ and $g, f \in \langle f \rangle$,
3. $r(x) = p(x)$ if x is such that $f(x) = 0$. Such an x is in \mathcal{D} .

In Section 2 we consider the algebra of orthogonal polynomials in one variable.

2 Orthogonal polynomials and their algebra

In this section let $d = 1$ and \mathcal{D} be the zero set of a polynomial which is orthogonal to the constant functions with respect to λ . We next recall the basics on orthogonal polynomials we use, see e.g. [5].

Let I be a finite or infinite interval of \mathbb{R} and λ a positive measure over I such that all moments $\mu_j = \int_I x^j d\lambda(x)$, $j = 0, 1, \dots$, exist and are finite. In particular, each polynomial function is square integrable on I and the $L^2(\lambda)$ scalar product is defined by

$$\langle f, g \rangle_\lambda = \int_I f(x)g(x) d\lambda(x)$$

We consider only λ whose related inner product is definite positive, i.e. $\|f\| = \sqrt{\langle f, f \rangle_\lambda} > 0$ if $f \neq 0$. In this case there is a unique infinite sequence of monic orthogonal polynomials with respect to λ and we denote them as

$$\pi_0(x) = 1, \pi_1(x) = x + \dots, \pi_2(x) = x^2 + \dots, \dots$$

Furthermore we have $\pi_k \in \mathbb{R}[x]_k$; π_0, \dots, π_k form a real vector space basis of $\mathbb{R}[x]_{\leq k}$; π_k is orthogonal to all polynomials of total degree smaller than k ; for $p \in \mathbb{R}[x]$ and $n \in \mathbb{Z}_{\geq 0}$ there exists unique $c_n(p) \in \mathbb{R}$, called n -th Fourier coefficient of p , such that $p(x) = \sum_{n=0}^{+\infty} c_n(p) \pi_n(x)$ and only a finite number of $c_n(p)$ are not zero.

Since the inner product satisfies the *shift property*, i.e.

$$\langle xp(x), q(x) \rangle_\lambda = \langle p(x), xq(x) \rangle_\lambda \text{ for } p, q \in \mathbb{R}[x]$$

then the corresponding orthogonal polynomial system satisfies a *three-term recurrence* relationship. More precisely, all orthogonal polynomial systems on the real line satisfy a three-term recurrence relationships. Conversely, Favard's theorem holds [11].

Theorem 1 (Favard's theorem) Let $\gamma_n, \alpha_n, \beta_n$ be sequences of real numbers and for $n \geq 0$ let

$$\pi_{n+1}(x) = (\gamma_n x - \alpha_n) \pi_n(x) - \beta_n \pi_{n-1}(x)$$

be defined recurrently with $\pi_0(x) = 1, \pi_{-1}(x) = 0$. The $\pi_n(x)$, $n = 0, 1, \dots$ form a system of orthogonal polynomials if and only if $\gamma_n \neq 0, \alpha_n \neq 0$ and $\alpha_n \gamma_n \gamma_{n-1} > 0$ for all $n \geq 0$. If $\gamma_n = 1$ for all n then the system is of monic orthogonal polynomials.

In the monic case,

$$\alpha_k = \frac{\langle x\pi_k, \pi_k \rangle}{\langle \pi_k, \pi_k \rangle} \quad \text{and} \quad \beta_k = \frac{\langle \pi_k, \pi_k \rangle}{\langle \pi_{k-1}, \pi_{k-1} \rangle}$$

hold true and therefore the norm of π_n is computed from the β 's as $\|\pi_n\|^2 = \beta_n \beta_{n-1} \dots \beta_0$.

For orthonormal polynomials $\tilde{\pi}_k = \pi_k / \|\pi_k\|$ the Christoffel-Darboux formulæ hold

$$\begin{aligned} \sum_{k=0}^{n-1} \tilde{\pi}_k(x) \tilde{\pi}_k(t) &= \sqrt{\beta_n} \frac{\tilde{\pi}_n(x) \tilde{\pi}_{n-1}(t) - \tilde{\pi}_{n-1}(x) \tilde{\pi}_n(t)}{x-t} \\ \sum_{k=0}^{n-1} \tilde{\pi}_k(t)^2 &= \sqrt{\beta_n} (\tilde{\pi}'_n(t) \tilde{\pi}_{n-1}(t) - \tilde{\pi}'_{n-1}(t) \tilde{\pi}_n(t)) \end{aligned} \quad (4)$$

Example 1 Inner products of the Sobolev type, namely $\langle u, v \rangle_S = \langle u, v \rangle_{\lambda_0} + \langle u', v' \rangle_{\lambda_1} \dots + \langle u^{(s)}, v^{(s)} \rangle_{\lambda_s}$ where λ_i are positive measures possibly having different support, do not satisfy the shift condition. Neither do the complex Hermitian inner products.

Theorem 2 Let $\mathcal{D} = \{x \in \mathbb{R} : \pi_n(x) = 0\}$ be the zero set of the n -th orthogonal polynomial with respect to the distribution λ of the real random variable X . Consider the division of $p \in \mathbb{R}[x]$ by π_n giving $p(x) = q(x)\pi_n(x) + r(x)$ as above. Then there exist weights $\lambda_z, z \in \mathcal{D}$, such that the expected value of $p(X)$ is

$$\mathbb{E}(p(X)) = c_n(q) \|\pi_n\|_\lambda^2 + \sum_{z \in \mathcal{D}} p(z) \lambda_z$$

where $\lambda_z = \mathbb{E}(l_z(X))$ and $c_n(q)$ is the n -th Fourier coefficient of the polynomial q .

Remark 1 This theorem is a version of a well known result, see e.g. [5, Sec. 1.4]. We include the proof to underline a particular form of the error in the quadrature formula, to be used again in Theorem 3 and in Section 6. Applying Theorem 2 to $p = 1$ we have that $1 = \sum_{z \in \mathcal{D}} \lambda_z$.

Proof The set \mathcal{D} contains n distinct points. For a univariate polynomial p , we can write uniquely $p(x) = q(x)\pi_n(x) + r(x)$ with $\deg(r) < n$ and $\deg(q) = \max\{\deg(p) - n, 0\}$. Furthermore, the indicator functions in the expression $r(x) = \sum_{z \in \mathcal{D}} p(z) l_z(x)$ are the Lagrange polynomials for \mathcal{D} : namely

$$l_z(x) = \prod_{w \in \mathcal{D}: w \neq z} \frac{x-w}{z-w}, \quad z \in \mathcal{D}.$$

Hence we have

$$\begin{aligned}\mathbb{E}(p(X)) &= \mathbb{E}(q(X)\pi_n(X)) + \sum_{z \in \mathcal{D}} p(z) \mathbb{E}(l_z(X)) \\ &= \sum_{k=0}^{+\infty} c_k(q) \mathbb{E}(\pi_k(X)\pi_n(X)) + \sum_{z \in \mathcal{D}} p(z) \lambda_z \\ &= c_n(q) \|\pi_n\|_\lambda^2 + \sum_{z \in \mathcal{D}} p(z) \lambda_z\end{aligned}$$

as $\mathbb{E}(\pi_k(X)\pi_n(X)) = \delta_{k,n}$ with $\delta_{k,n} = 1$ if $k = n$ and 0 otherwise. \square

A particular case of Theorem 2 occurs if p has degree less than $2n$. In this case q has degree at most $n-1$ and $c_n(q) = 0$. This shows that the quadrature rule with n nodes given by the zeros of π_n and weights $\{\lambda_z\}_{z \in \mathcal{D}}$ is a Gaussian quadrature rule and it is exact for all polynomial functions of degree smaller or equal to $2n-1$. For notes on quadrature rules see for example [5, Ch. 1].

Example 2 (Identification) For f polynomial of degree $N \leq 2n-1$ we can write $f(x) = \sum_{k=0}^N c_k(f) \pi_k(x)$. The constant term is given by

$$c_0(f) = \mathbb{E}(f(X)) = \sum_{z \in \mathcal{D}} f(z) \lambda_z$$

and for all i such that $N+i \leq 2n-1$

$$\|\pi_i\|_\lambda^2 c_i(f) = \mathbb{E}(f(X)\pi_i(X)) = \sum_{z \in \mathcal{D}} f(z) \pi_i(z) \lambda_z$$

In particular, if $\deg f = n-1$ then all coefficients in the Fourier expansion of f can be computed with an evaluation on \mathcal{D} .

In general for a polynomial of degree N possibly larger than $2n-1$, Theorem 2 gives the Fourier expansion of its remainder by π_n , indeed

$$\begin{aligned}\sum_{z \in \mathcal{D}} f(z) \pi_i(z) \lambda_z &= \sum_{z \in \mathcal{D}} \text{NF}(f \pi_i)(z) \lambda_z \\ &= \mathbb{E}(\text{NF}(f(X) \pi_i(X))) = \|\pi_i\|_\lambda^2 c_i(\text{NF}(f))\end{aligned}$$

Theorem 3 below generalises Theorem 2 to a generic finite set of n distinct points in \mathbb{R} , say \mathcal{D} . As above, the indicator function of $z \in \mathcal{D}$ is $l_z(x) = \prod_{w \in \mathcal{D}: w \neq z} \frac{x-w}{z-w}$. Let $g(x) = \prod_{z \in \mathcal{D}} (x-z)$ be the unique monic polynomial vanishing over \mathcal{D} and of degree n . Write a polynomial $p \in \mathbb{R}[x]$ uniquely as $p(x) = q(x)g(x) + r(x)$ and consider the Fourier expansions of q and g : $q(x) = \sum_{k=0}^{+\infty} c_k(q) \pi_k(x)$ and $g(x) = \sum_{k=0}^n c_k(g) \pi_k(x)$.

Theorem 3 *With the above notation,*

$$\mathbb{E}_\lambda(p(X)) = \sum_{k=0}^{+\infty} c_k(q) c_k(g) \|\pi_k\|_\lambda^2 + \sum_{z \in \mathcal{D}} p(z) \lambda_z.$$

Proof From

$$\begin{aligned}p(x) &= q(x)g(x) + r(x) \\ &= \sum_{k=0}^{+\infty} c_k(q) \pi_k(x) \sum_{j=0}^n c_j(g) \pi_j(x) + \sum_{z \in \mathcal{D}} p(z) l_z\end{aligned}$$

we have

$$\begin{aligned}\mathbb{E}(p(X)) &= \sum_{k=0}^{+\infty} \sum_{j=0}^n c_k(q) c_j(g) \mathbb{E}(\pi_k(X) \pi_j(X)) + \sum_{z \in \mathcal{D}} p(z) \lambda_z \\ &= \sum_{k=0}^n c_k(q) c_k(g) \|\pi_k\|_\lambda^2 + \sum_{z \in \mathcal{D}} p(z) \lambda_z\end{aligned}$$

and this proves the theorem. \square

The condition in Theorem 3 is linear in the Fourier coefficients of q , which is found easily from f by polynomial division. The first $|\mathcal{D}|$ Fourier coefficients of q appearing in the conditions of the theorem are determined by solving the system of linear equations

$$M[c_k(q)]_{k=0, \dots, |\mathcal{D}|-1} = [q(z)]_{k=0, \dots, |\mathcal{D}|-1} \quad (5)$$

where $M = [\pi_k(z)]_{z \in \mathcal{D}, k=0, \dots, |\mathcal{D}|-1}$ is the design/evaluation matrix for the first $|\mathcal{D}|$ orthogonal polynomials.

Theorem 3 can be used in two ways at least. If p is known, the condition in the theorem can be checked to verify if the expected value of p can be determined by Gaussian quadrature rule with nodes \mathcal{D} and weights

$$\lambda_z = \mathbb{E} \left(\prod_{w \in \mathcal{D}: w \neq z} \frac{X-w}{z-w} \right) = \frac{\sum_0^{n-1} \alpha(z, k) \mathbb{E}(X^k)}{\sum_0^{n-1} \alpha(z, k) z^k}$$

for $z \in \mathcal{D}$, where $\alpha(z, k)$ is the k -th symmetric function of the polynomial $\pi_n(x)/(x-z)$. The Fourier coefficients of g can be computed analogously to those of q adapting Equation (5).

If p is an unknown polynomial and $p(x) = \sum_\alpha p_\alpha x^\alpha$ for a finite number of non-zero, unknown, real coefficients p_α , Theorem 3 characterizes all the polynomials for which the Gaussian quadrature rule is exact, namely $\mathbb{E}_\lambda(p(X)) = \sum_{z \in \mathcal{D}} p(z) \lambda_z$. Furthermore, the characterization is a linear expression in the unknown p_α . This is because in Equation (5) the $q(z)$ are linear combinations of the coefficients of p .

In Section 3 we shall specialise our study to Hermite polynomials, while in Section 5 we shall generalise Theorem 3 to higher dimension. To conclude this section, we discuss the remainder r vs the orthogonal projection.

Remark 2 Let $p(x) \in \mathbb{R}[x]$ and write $p(x) = q(x)\pi_n(x) + r(x)$ where r has degree less than n . Then

1. q is the unique polynomial such that $p - q\pi_n$ is orthogonal to all π_m with $m \geq n$. This is a rephrasing of the characteristic property of the remainder: r belongs to $\mathbb{R}[x]_{\leq n-1}$ if, and only if, r is orthogonal to all π_m with $m \geq n$. Should it exist two such q 's, then $(q_1 - q_2)\pi_n$ would be in the same space, hence null.
2. If $\deg(p) = n$, then r is the orthogonal projection of p on $\mathbb{R}[x]_{\leq n-1}$. In fact, q is the leading coefficient of p , therefore $p - r$ is a multiple of π_n and indeed orthogonal to $\mathbb{R}[x]_{\leq n-1}$.
3. If $\deg(p) = N \geq n$, then the orthogonal projection of p on $\mathbb{R}[x]_{\leq n-1}$ differs from r , unless the projection of $q\pi_n$ is zero.

Example 3 Substituting the Fourier expansions of q and p in the division above, for $m \geq n$ we find that the m -th coefficient in the Fourier expansion of p can be written as

$$\begin{aligned} \mathbb{E}(p(X)\pi_m(X)) &= \mathbb{E}(q(X)\pi_n(X)\pi_m(X)) \\ \sum_{k=0}^{+\infty} c_k(p) \mathbb{E}(\pi_k(X)\pi_m(X)) &= \sum_{j=0}^{+\infty} c_j(q) \mathbb{E}(\pi_j(X)\pi_n(X)\pi_m(X)) \\ c_m(p) \|\pi_m\|^2 &= \sum_{j=0}^{+\infty} c_j(q) \mathbb{E}(\pi_j(X)\pi_n(X)\pi_m(X)) \end{aligned}$$

For Hermite polynomials it can be simplified by e.g. using the product formula in Theorem 4 of Section 3.

3 Hermite polynomials

There is another way to look at the algebra of orthogonal polynomials that we discuss here in the case of Hermite polynomials. The reference measure λ is the normal distribution and $d\lambda(x) = w(x)dx$, with $w(x) = e^{-x^2/2}/\sqrt{2\pi}$, $x \in \mathbb{R}$.

3.1 Stein-Markov operators for standard normal distribution

For a real valued, differentiable function f , define

$$\delta f(x) = xf(x) - \frac{d}{dx}f(x) = -e^{x^2/2} \frac{d}{dx} \left(f(x)e^{-x^2/2} \right),$$

$d^n = \frac{d^n}{dx^n}$, and consider $Z \sim \lambda$. The following identity holds

$$\mathbb{E}(\phi(Z) \delta^n \psi(Z)) = \mathbb{E}(d^n \phi(Z) \psi(Z)) \quad (6)$$

if ϕ, ψ are such that $\lim_{x \rightarrow \pm\infty} \phi(x)\psi(x)e^{-x^2/2} = 0$ and are square integrable, see [7, Ch. V Lemma 1.3.2 and Proposition 2.2.3]). Polynomials satisfy these conditions and δ is also called the Stein-Markov operator for the standard normal distribution. It is a shift-equivariant linear operator on $\mathbb{R}[x]$, namely $\delta f(x+a) = (x+a)f(x+a) - f'(x+a) =$

$(\delta f)(x+a)$ holds for $a \in \mathbb{R}$ and the constant one is mapped into x .

The n -th Hermite polynomial can be defined as $H_n(x) = \delta^n 1$. Direct computation using δ proves the following well-known facts:

1. The first Hermite polynomials are

$$\begin{aligned} H_0 &= 1 \\ H_1(x) &= x \\ H_2(x) &= x^2 - 1 \\ H_3(x) &= x^3 - 3x \\ H_4(x) &= x^4 - 6x^2 + 3 \\ H_5(x) &= x^5 - 10x^3 + 15x \end{aligned}$$

2. $H_n(x) = (-1)^n e^{x^2/2} d^n (e^{-x^2/2})$ (Rodrigues' formula)
3. $d\delta - \delta d$ is the identity operator. From this the relationships $dH_n = nH_{n-1}$, $d^m H_n = \frac{n!}{m!} H_{n-m}$ for $m \leq n$ and the three-term recurrence relationship

$$H_{n+1} = xH_n - nH_{n-1} \quad (7)$$

are deduced.

4. Hermite polynomials are orthogonal with respect to the standard normal distribution λ . Indeed from Equation (6) we have $\mathbb{E}(H_n(Z)H_m(Z)) = n!\delta_{n,m}$ where $\delta_{n,m} = 0$ if $n \neq m$ and $\delta_{n,m} = 1$ if $n = m$.

We already mentioned that $\{H_n(x) : n \leq d\}$ spans $\mathbb{R}[x]_{\leq d}$ and that H_n is orthogonal to any polynomial of degree different from n . The ring structure of the space generated by the Hermite polynomials is described in Theorem 4.

Theorem 4 *The Fourier expansion of the product $H_k H_n$ is*

$$H_k H_n = H_{n+k} + \sum_{i=1}^{n \wedge k} \binom{n}{i} \binom{k}{i} i! H_{n+k-2i}$$

Proof Note that $\langle \phi, \psi \rangle = \mathbb{E}(\phi(Z)\psi(Z))$ is a scalar product on the obvious space and let $n \leq k$ with $Z \sim \mathcal{N}(0, 1)$ and ψ, ϕ square integrable functions for which identity (6) holds. Then

$$\begin{aligned} \langle H_k H_n, \psi \rangle &= \langle \delta^n 1, H_k \psi \rangle \\ &= \langle 1, d^n (H_k \psi) \rangle = \sum_{i=0}^n \langle 1, \binom{n}{i} d^i H_k d^{n-i} \psi \rangle \\ &= \langle 1, H_k d^n \psi \rangle + \sum_{i=1}^n \langle 1, \binom{n}{i} d^i H_k d^{n-i} \psi \rangle \\ &= \langle H_{n+k}, \psi \rangle + \sum_{i=1}^n \binom{n}{i} k(k-1) \dots (k-i+1) \langle H_{n+k-2i}, \psi \rangle \\ &= \langle H_{n+k}, \psi \rangle + \left\langle \sum_{i=1}^n \binom{n}{i} \binom{k}{i} i! H_{n+k-2i}, \psi \right\rangle \end{aligned}$$

□

Example 4 (Aliasing) As an application of Theorem 4, observe that the three-term recurrence relation for Hermite polynomials, Equation (7)

$$H_{n+1} = xH_n - nH_{n-1}$$

evaluated on the zeros of $H_n(x)$, say \mathcal{D}_n , becomes $H_{n+1}(x) \equiv -nH_{n-1}(x)$ where \equiv indicates that equality holds for $x \in \mathcal{D}_n$. In general let $H_{n+k} \equiv \sum_{j=0}^{n-1} h_j^{n+k} H_j$ be the Fourier expansion of the normal form of H_{n+k} at \mathcal{D}_n , where we simplified the notation for the Fourier coefficients. Substitution in the product formula in Theorem 4 gives the formula to write h_j^{n+k} in terms of Fourier coefficients of lower order Hermite polynomials:

$$\begin{aligned} \text{NF}(H_{n+k}) &\equiv -\sum_{i=1}^{n \wedge k} \binom{n}{i} \binom{k}{i} i! \text{NF}(H_{n+k-2i}) \\ &\equiv -\sum_{i=1}^{n \wedge k} \binom{n}{i} \binom{k}{i} i! \sum_{j=0}^{n-1} h_j^{n+k-2i} H_j \end{aligned} \quad (8)$$

Equating coefficients gives a closed formula

$$h_j^{n+k} = -\sum_{i=1}^{n \wedge k} \binom{n}{i} \binom{k}{i} i! h_j^{n+k-2i}$$

In Table 1 the normal form of H_{k+n} with respect to H_n is written in terms of Hermite polynomials of degree smaller than n . For example, $H_{n+3}(x) = -n(n-1)(n-2)H_{n-3}(x) + 3nH_{n-1}(x)$ for those values of x such that $H_n(x) = 0$.

k	$H_{n+k} \equiv$
1	$-nH_{n-1}$
2	$-n(n-1)H_{n-2}$
3	$-n(n-1)(n-2)H_{n-3} + 3nH_{n-1}$
4	$-n(n-1)(n-2)(n-3)H_{n-4} + 8n(n-1)H_{n-2}$
5	$-\frac{n!}{(n-5)!}H_{n-5} + 5nH_{n-1} + 15n(n-1)(n-2)H_{n-3}$
6	$-\frac{n!}{(n-6)!}H_{n-6} + 24n(n-1)(n-2)(n-3)H_{n-4} + 10n(n-1)(2n-5)H_{n-2}$

Table 1 Aliasing of H_{n+k} , $k = 1, \dots, 6$ over $\mathcal{D} = \{H_n(x) = 0\}$

Example 5 Observe that if f has degree $n+1$ equivalently $k = 1$ then

$$\begin{aligned} f &= \sum_{i=0}^{n-1} c_i(f)H_i + \underline{c_n(f)H_n} + c_{n+1}(f)H_{n+1} \\ &\equiv \sum_{i=0}^{n-2} c_i(f)H_i + (c_{n-1}(f) - nc_{n+1}(f))H_{n-1} \end{aligned}$$

and all coefficients up to degree $n-2$ are “clean”.

We give another proof of Theorem 2 for Hermite polynomials.

Corollary 1 Let $\mathcal{D}_n = \{x : H_n(x) = 0\}$ and $p \in \mathbb{R}[x]$. Let $p(x) = q(x)H_n(x) + r(x)$ with the degree of r smaller than n and let $Z \sim \mathcal{N}(0, 1)$. Then

$$\mathbb{E}(p(Z)) = \sum_{z \in \mathcal{D}_n} p(z)\lambda_z \text{ if and only if } \mathbb{E}(d^n q(Z)) = 0$$

$$\text{with } \lambda_z = \mathbb{E}(l_z(Z)) \text{ and } l_z(x) = \prod_{w \in \mathcal{D}_n: w \neq z} \frac{x-w}{z-w}, z \in \mathcal{D}_n.$$

Proof From Equation (6) we have

$$\mathbb{E}(q(Z)H_n(Z)) = \mathbb{E}(q(Z)\delta^n 1) = \mathbb{E}(d^n q(Z))$$

Now by the same steps followed in the proof of Theorem 2 we conclude that

$$\mathbb{E}(p(Z)) = \mathbb{E}(d^n q(Z)) + \sum_{z \in \mathcal{D}_n} p(z)\lambda_z$$

□

3.2 Algebraic characterisation of the weights

Theorem 5 gives two polynomial equations whose zeros are the design points and the weights. This is a particular case of Formula (1.17) in [12]. We provide the proof to highlight the algebraic nature of the result and of its proof.

Theorem 5 Let $\mathcal{D} = \{x : H_n(x) = 0\}$.

1. There exists only one polynomial λ of degree $n-1$ such that $\lambda(x) = \lambda_x$ for all $x \in \mathcal{D}$,
2. furthermore $\lambda_x = \frac{(n-1)!}{n} H_{n-1}^{-2}(x)$. Equivalently
3. the polynomial λ satisfies

$$\begin{cases} H_n(x) = 0 \\ \lambda(x)H_{n-1}^2(x) = \frac{(n-1)!}{n} \end{cases}$$

Proof 1. The univariate polynomial λ is the interpolation polynomial of the values λ_x 's at the n distinct points in \mathcal{D}_n and hence it exists, unique of degree $n-1$.

2. Observe that for Hermite polynomials $\alpha_n = 0$, $\beta_n = n$, $\tilde{H}_n(x) = H_n(x)/\sqrt{n!}$ and $\tilde{H}'_n(x) = \sqrt{n}\tilde{H}_{n-1}(x)$. Substitution in the Christoffel-Darboux formulæ and evaluation at $\mathcal{D}_n = \{x_1, \dots, x_n\}$ give

$$\begin{aligned} \sum_{k=0}^{n-1} \tilde{H}_k(x_i)\tilde{H}_k(x_j) &= 0 \text{ if } i \neq j \\ \sum_{k=0}^{n-1} \tilde{H}_k(x_i)^2 &= n\tilde{H}_{n-1}(x_i)^2 \end{aligned} \quad (9)$$

In matrix form Equations (9) becomes

$$\mathbb{H}_n \mathbb{H}_n^t = n \text{diag}(\tilde{H}_{n-1}(x_i)^2 : i = 1, \dots, n)$$

where \mathbb{H}_n is the square matrix $\mathbb{H}_n = \left[\tilde{H}_j(x_i) \right]_{i=1, \dots, n; j=0, \dots, n-1}$ and diag indicates a diagonal matrix. Observe that \mathbb{H}_n is invertible and

$$\mathbb{H}_n^{-1} = \mathbb{H}_n^t n^{-1} \text{diag}(\tilde{H}_{n-1}^{-2}(x_i) : i = 1, \dots, n)$$

Now, let f be a polynomial of degree at most $n-1$, that is a typical remainder by division for H_n , then $f(x) = \sum_{j=0}^{n-1} c_j \tilde{H}_j(x)$. Write $\underline{f} = \mathbb{H}_n \underline{c}$ where $\underline{f} = [f(x_i)]_{i=1, \dots, n}$ and $\underline{c} = [c_j]_j$. Furthermore note that

$$\begin{aligned} \underline{c} &= \mathbb{H}_n^{-1} \underline{f} = \mathbb{H}_n^t n^{-1} \text{diag}(\tilde{H}_{n-1}^{-2}(x_i) : i = 1, \dots, n) \underline{f} \\ &= \mathbb{H}_n^t n^{-1} \text{diag}(\tilde{H}_{n-1}^{-2}(x_i) f(x_i) : i = 1, \dots, n) \\ c_j &= \frac{1}{n} \sum_{i=1}^n \tilde{H}_j(x_i) f(x_i) \tilde{H}_{n-1}^{-2}(x_i) \end{aligned} \quad (10)$$

Apply this to the k -th Lagrange polynomial, $f(x) = l_k(x)$, whose Fourier expansion is $f(x) = \sum_{j=0}^{n-1} c_{kj} \tilde{H}_j(x)$. Using $l_k(x_i) = \delta_{ik}$ in Equation (10), obtain

$$c_{kj} = \frac{1}{n} \tilde{H}_j(x_k) \tilde{H}_{n-1}^{-2}(x_k) \quad (11)$$

The expected value of $l_k(Z)$ is

$$\lambda_k = \mathbb{E}(l_k(Z)) = \sum_{j=0}^{n-1} c_{kj} \mathbb{E}(\tilde{H}_j(x)) = c_{k0}$$

Substitution in Equation (11) for $j=0$ gives

$$\lambda_k = \frac{1}{n} \tilde{H}_{n-1}^{-2}(x_k) = \frac{(n-1)!}{n} H_{n-1}^{-2}(x_k)$$

This holds for all $k = 1, \dots, n$.

- The system of equations is a rewriting of the previous parts of the theorem because the first equation $H_n(x) = 0$ states that only values of $x \in \mathcal{D}_n$ are to be considered and the second equation is what we have just proven. \square

Item 2 in Theorem 5 states that the weights are strictly positive. Theorem 2 applied to the constant polynomial $p(x) = 1$ shows that they sum to one. In other words, the mapping that associates z to λ_z , $z \in \mathcal{D}_n$, is a discrete probability density. Theorem 2 states that the expected value of the polynomial functions of $Z \sim \mathcal{N}(0, 1)$ for which $c_n(q) = 0$, is equal to the expected value of a discrete random variables X given by $P_n(X = x_k) = \mathbb{E}(l_k(Z)) = \lambda_k$, $k = 1, \dots, n$

$$\mathbb{E}(p(Z)) = \sum_{k=1}^n p(x_k) \lambda_k = \mathbb{E}_n(p(X))$$

Example 6 For $n = 3$ the polynomial λ in Theorem 5 can be determined by-hand. For larger values of n an algorithm is

provided in Section 3.3. The polynomial system to be considered is

$$\begin{aligned} 0 &= H_3(x) = x^3 - 3x \\ 2/3 &= \lambda(x) H_2^2 = (\theta_0 + \theta_1 x + \theta_2 x^2)(x^2 - 1)^2 \end{aligned}$$

where $\lambda(x) = \theta_0 + \theta_1 x + \theta_2 x^2$. The degree of $\lambda(x) H_2^2$ is reduced to 2 by using $x^3 = 3x$

$$2/3 = \lambda(x) H_2^2 = \theta_0 + \theta_1 4x + (\theta_0 + 4\theta_2)x^2 \quad (12)$$

Coefficients in Equation (12) are equated to give $\lambda(x) = \frac{2}{3} - \frac{x^2}{6}$.

In some situations, e.g. the design of an experimental plan or of a Gaussian quadrature rule, the exact computation of the weights might not be necessary and $\lambda(x)$ is all we need. When the explicit values of the weights are required, the computation has to be done outside a symbolic computation setting as we need to solve, e.g., $H_3(x) = 0$ to get $\mathcal{D}_3 = \{-\sqrt{3}, 0, \sqrt{3}\}$ and evaluate $\lambda(x)$ to find $\lambda_{-\sqrt{3}} = \lambda(-\sqrt{3}) = \frac{1}{6} = \lambda_{\sqrt{3}}$ and $\lambda_0 = \lambda(0) = \frac{2}{3}$.

Example 7 Let a positive integer N be given and let k and n be positive integers such that $kn < 2N$ then

$$\mathbb{E}(H_n(Z)^k) = \sum_{z \in \mathcal{D}_N} \frac{(n-1)!}{n} \frac{H_n(z)^k}{H_{N-1}^2(z)}$$

by Theorem 5. The issue is then the evaluation of H_{N-1} and H_n at the zeros of H_N , for which the recurrence relationship can be used when the values are not tabulated.

3.3 Code for the weighing polynomial

The polynomial $\lambda(x)$ in Theorem 5 is called the *weighing polynomial*. Table 2 gives a code written in the specialised software for symbolic computation called CoCoA [1] to compute the Fourier expansion of $\lambda(x)$ based on Theorem 5.

Line 1 specifies the number of nodes N . Line 2 establishes that the working environment is a polynomial ring whose variables are the first $(N-1)$ -Hermite polynomials plus an extra variable w which encodes the weighing polynomial; here it is convenient to work with a elimination term-ordering of w , called `Elim(w)`, so that the variable w will appear as least as possible. Lines 3, 4, 5 construct Hermite polynomials up-to-order N by using the recurrence relationships (7). Specifically they provide the expansion of H_j over H_k with $k < j$ for $k = 0, \dots, N-1$. Line 6 states that $H_N = H_1 H_{N-1} - (N-1) H_{N-2} = 0$, ‘giving’ the nodes of the quadrature. Line 7 is the polynomial in the second equation in the system in Item 3 of Theorem 5 and ‘gives’ the weights. There are N equations which are collected in an ideal whose Gröbner basis is computed in Line 8. In our application it is

interesting that the Gröbner bases contains a polynomial in which w appears alone as a term of degree one. This element of the Gröbner basis relates explicitly to the desired weighing polynomial w to the first $(n - 1)$ -Hermite polynomials.

```
L1 N:=4;
L2 Use R:=Q[w,h[1..(N-1)]], Elim(w);
L3 Eqs:=[h[2]-h[1]*h[1]+1];
L4 For I:=3 To N-1 Do
L5 Append(Eqs,h[I]-h[1]*h[I-1]+(I-1)*h[I-2]) EndFor;
L6 Append(Eqs,h[1]*h[N-1]-(N-1)*h[N-2]);
L7 Append(Eqs,N*w*h[N-1]^2-Fact(N-1));
L8 J:=Ideal(Eqs); GB_J:=GBasis(J); Last(GB_J);
L9 3w + 1/4h[2] - 5/4
```

Table 2 Computation of the Fourier expansion of the weighing polynomial using Theorem 5

Line 9 in Table 2 gives the polynomial obtained for $N = 4$, as set in Line 1, namely

$$\lambda(x) = \left(-\frac{1}{4}H_2(x) + \frac{5}{4} \right) \frac{1}{3} = \frac{6-x^2}{12}$$

The nodes are $\pm\sqrt{3 \pm \sqrt{6}}$ and the values of the weights are $\frac{3 \pm \sqrt{6}}{12}$, showing that both nodes and weights are algebraic numbers but not rational numbers. On a Mac OS X with an Intel Core 2 Duo processor (at 2.4 GHz) using CoCoA (release 4.7) the result is obtained for $N = 10$ in Cpu time = 0.08, User time = 0; for $N = 20$ in Cpu time = 38.40, User time = 38; for $N = 25$ in Cpu time = 141.28, User time = 142 and for $N = 30$ in Cpu time = 5132.71, User time = 5186. Observe that this computations can be done once for all and the results stored. Observe furthermore that in Line 8 the CoCoA commands `GB_J:=GBasis(J); Last(GB_J);` could be substituted by `NF(w, J)`. This does not improve on computational cost as NF requires the computation of a Gröbner basis and a reduction. As a minor point we observe that the symbol w would not appear in Line 9.

4 Fractional design

In this section we return to the case of general orthogonal polynomials, $\{\pi_n\}_n$, and positive measure, $d\lambda$. We assume that the nodes are a proper subset \mathcal{F} of $\mathcal{D}_n = \{x \in \mathbb{R} : \pi_n(x) = 0\}$ with m points, $0 < m < n$. We work within two different settings, in one the ambient design \mathcal{D}_n is considered while in the other one it is not.

Consider the indicator function of \mathcal{F} as subset of \mathcal{D}_n , namely $1_{\mathcal{F}}(x) = 1$ if $x \in \mathcal{F}$ and 0 if $x \in \mathcal{D}_n \setminus \mathcal{F}$. It can be represented by a polynomial of degree n because it is a function defined over \mathcal{D}_n [2, 10]. Let p be a polynomial of degree at most $n - 1$ so that the product $p(x)1_{\mathcal{F}}(x)$ is a

polynomial of degree at most $2n - 1$. Then from Theorem 2 we have

$$\begin{aligned} \mathbb{E}((p1_{\mathcal{F}})(X)) &= \sum_{z \in \mathcal{F}} p(z)\lambda_z = \mathbb{E}_n(p(Y)1_{\mathcal{F}}(Y)) \\ &= \mathbb{E}_n(p(Y)|Y \in \mathcal{F})P_n(Y \in \mathcal{F}) \end{aligned}$$

where X is a random variable with probability law λ and Y is a discrete random variable taking value $z \in \mathcal{F}$ with probability $P_n(Y = z) = \lambda_z$. The first equality follows from the fact that $p(x)1_{\mathcal{F}}(x)$ is zero for $x \in \mathcal{D} \setminus \mathcal{F}$ and the last equality from the definition of conditional expectation.

Another approach is to consider the polynomial whose zeros are the elements of \mathcal{F} , say $\omega_{\mathcal{F}}(x) = \prod_{z \in \mathcal{F}} (x - z)$. Now consider the Lagrange polynomials for \mathcal{F} , namely $l_z^{\mathcal{F}}(x) = \prod_{\substack{w \neq z \\ w \in \mathcal{F}}} \frac{x - w}{z - w}$ for $z \in \mathcal{F}$.

Lemma 1 *Let $\mathcal{F} \subset \mathcal{D}_n$. The Lagrange polynomial for $z \in \mathcal{F}$ is the remainder of the Lagrange polynomial for $z \in \mathcal{D}_n$ with respect to $\omega_{\mathcal{F}}(x)$, namely*

$$l_z^{\mathcal{F}}(x) = \text{NF}(l_z(x), \langle \omega_{\mathcal{F}}(x) \rangle)$$

Proof There exists unique $\text{NF}(l_z)(x)$, polynomial of degree small than m , such that

$$l_z(x) = q(x)\omega_{\mathcal{F}}(x) + \text{NF}(l_z)(x)$$

Furthermore, for $a \in \mathcal{F}$ we have $l_z(a) = \text{NF}(l_z)(a) = \delta_{z,a} = l_z^{\mathcal{F}}(a)$. The two polynomials $l_z^{\mathcal{F}}(x)$ and $\text{NF}(l_z)(x)$ have degree smaller than m and coincide on m points, by interpolation they must be equal. \square

For a polynomial p of degree N , write $p(x) = q(x)\omega_{\mathcal{F}}(x) + r(x)$ with $f(z) = r(z)$ if $z \in \mathcal{F}$ and $r(x) = \sum_{z \in \mathcal{F}} p(z)l_z^{\mathcal{F}}(x)$. Let $q(x) = \sum_{j=0}^{N-m} b_j\pi_j(x)$ and $\omega_{\mathcal{F}}(x) = \sum_{i=0}^m c_i\pi_i(x)$ as $\omega_{\mathcal{F}}$ has degree m . Then

$$\begin{aligned} \mathbb{E}(p(X)) &= \mathbb{E}\left(\sum_{j=0}^{N-m} b_j\pi_j(X) \sum_{i=0}^m c_i\pi_i(X)\right) + \mathbb{E}(r(X)) = \\ &b_0c_0\|\pi_0\|_{\lambda}^2 + b_1c_1\|\pi_1\|_{\lambda}^2 + \dots \\ &+ b_{(N-m) \wedge m}c_{(N-m) \wedge m}\|\pi_{(N-m) \wedge m}\|_{\lambda}^2 + \sum_{z \in \mathcal{F}} p(z)\lambda_z^{\mathcal{F}} \end{aligned}$$

where $\lambda_z^{\mathcal{F}} = \mathbb{E}(\text{NF}(l_z(X), \langle \omega_{\mathcal{F}}(X) \rangle))$, $z \in \mathcal{F}$.

Note that the error of the Gaussian quadrature rule,

$$\begin{aligned} &b_0c_0\|\pi_0\|_{\lambda}^2 + b_1c_1\|\pi_1\|_{\lambda}^2 + \dots + \\ &b_{(N-m) \wedge m}c_{(N-m) \wedge m}\|\pi_{(N-m) \wedge m}\|_{\lambda}^2 \end{aligned}$$

is linear in the Fourier coefficients b_j , and also in the Fourier coefficients c_j of the node polynomial. This is generalised in Section 5. If the fraction \mathcal{F} coincides with the ambient design \mathcal{D}_n and hence contains n points and if p is a polynomial

of degree at most $2n - 1$, then we obtain the well known result of zero error because $(N - n) \wedge n \leq n - 1$ and the only non-zero Fourier coefficient of the node polynomial π_n is of order n . In general one should try to determine pairs of \mathcal{F} and sets of polynomials for which the absolute value of the errors is minimal.

5 Higher dimension: zero set of orthogonal polynomials as design support

In this section we return to the higher dimensional set-up of Section 1.1 but we restrict ourselves to consider the product measure $\lambda^d = \times_{i=1}^d \lambda$ and X_1, \dots, X_d independent random variables each one of which is distributed according to the probability law λ . As design we take a product grid of zeros of orthogonal polynomials with respect to λ , more precisely our design points or interpolation nodes are

$$\mathcal{D}_{n_1, \dots, n_d} = \left\{ x \in \mathbb{R}^d : \pi_{n_1}(x_1) = \pi_{n_2}(x_2) = \dots = \pi_{n_d}(x_d) = 0 \right\}$$

where π_{n_k} is the orthogonal polynomial with respect to λ of degree n_k .

The Lagrange polynomial of the point $y = (y_1, \dots, y_d) \in \mathcal{D}_{n_1, \dots, n_d}$ is defined as $l_y(x_1, \dots, x_d) = \prod_{k=1}^d l_{y_k}^{n_k}(x_k)$, the apex n_k indicates that $l_{y_k}^{n_k}(x_k)$ is the univariate Lagrange polynomial for $y_k \in \{x_k : \pi_{n_k}(x_k) = 0\} = \mathcal{D}_{n_k} \subset \mathbb{R}$.

The Span $(l_y : y \in \mathcal{D}_{n_1, \dots, n_d})$ is equal to the linear space generated by the monomials whose exponents lie on the integer grid $\{0, \dots, n_1 - 1\} \times \dots \times \{0, \dots, n_d - 1\}$. Any polynomial $f \in \mathbb{R}[x]$ can be written as

$$f(x_1, \dots, x_d) = \sum_{k=1}^d q_k(x_1, \dots, x_d) \pi_{n_k}(x_k) + r(x_1, \dots, x_d)$$

where r is unique, its degree in the variable x_k is smaller than n_k , for $k = 1, \dots, d$, and belongs to that Span.

The coefficients of the Fourier expansion of q_k with respect to the variable x_k are functions of x_1, \dots, x_d but not of x_k . Let x_{-k} denote the $(d - 1)$ -dimensional vector obtained from (x_1, \dots, x_d) removing the k -th component and write

$$f(x_1, \dots, x_d) = \sum_{k=1}^d \left(\sum_{j=0}^{+\infty} c_j(q_k)(x_{-k}) \pi_j(x_k) \right) \pi_{n_k}(x_k) + r(x_1, \dots, x_d)$$

Only a finite number of $c_j(q_k)(x_{-k})$ are not zero.

From the independence of X_1, \dots, X_n , the expected value of the Lagrange polynomial l_y is

$$\mathbb{E}_{\lambda^d} (l_y(X_1, \dots, X_d)) = \prod_{k=1}^d \mathbb{E}_{\lambda} (l_{y_k}^{n_k}(X_k)) = \prod_{k=1}^d \lambda_k^{n_k}$$

where $\lambda_k^{n_k} = \mathbb{E} (l_{y_k}^{n_k}(X_k))$ is the expected value of a univariate random Lagrange polynomial as in the previous sections.

Theorem 6 It holds

$$\begin{aligned} \mathbb{E}_{\lambda^d} (f(X_1, \dots, X_d)) &= \\ \sum_{k=1}^d \mathbb{E}_{\lambda^{d-1}} (c_k(q_k)(X_{-k})) \|\pi_k\|_{\lambda}^2 &+ \\ \sum_{(x_1, \dots, x_n) \in \mathcal{D}_{n_1, \dots, n_d}} f(x_1, \dots, x_d) \lambda_{x_1}^{n_1} \dots \lambda_{x_d}^{n_d} & \end{aligned}$$

Proof The proof is very similar to that of Theorem 2 and we do it for $d = 2$ only. In a simpler notation the design is the $n \times m$ grid given by $\mathcal{D}_{nm} = \{(x, y) : \pi_n(x) = 0 = \pi_m(y)\}$ and X and Y are independent random variables distributed according to λ . The polynomial f is decomposed as

$$\begin{aligned} f(x, y) &= \\ q_1(x, y) \pi_n(x) + q_2(x, y) \pi_m(y) &+ \sum_{(a, b) \in \mathcal{D}_{n, m}} f(a, b) l_a^n(x) l_b^m(y) = \\ \sum_{j=0}^{+\infty} c_j(q_1)(y) \pi_j(x) \pi_n(x) &+ \sum_{j=0}^{+\infty} c_j(q_2)(x) \pi_j(y) \pi_m(y) + \\ \sum_{(a, b) \in \mathcal{D}_{n, m}} f(a, b) l_a^n(x) l_b^m(y) & \end{aligned}$$

Taking expectation, using independence of X and Y and orthogonality of the π_i , we have

$$\begin{aligned} \mathbb{E}_{\lambda^2} (f(X, Y)) &= \\ \mathbb{E}_{\lambda} (c_n(q_1)(Y)) \|\pi_n\|_{\lambda}^2 &+ \mathbb{E}_{\lambda} (c_m(q_2)(X)) \|\pi_m\|_{\lambda}^2 + \\ \sum_{(a, b) \in \mathcal{D}_{n, m}} f(a, b) \lambda_a^n \lambda_b^m & \end{aligned}$$

□

Note in the proof above that a sufficient condition for $\mathbb{E}_{\lambda} (c_n(q_1)(Y))$ being zero is that f has degree in x smaller than $2n - 1$, similarly for $\mathbb{E}_{\lambda} (c_m(q_2)(X))$. We retrieve the well-known results that if for each i the degree in x_i of f is smaller than $2n_i - 1$, then

$$\mathbb{E}_{\lambda^d} (f(X_1, \dots, X_d)) = \sum_{(x_1, \dots, x_n) \in \mathcal{D}_{n_1, \dots, n_d}} f(x_1, \dots, x_d) \lambda_{x_1}^{n_1} \dots \lambda_{x_d}^{n_d}$$

In the Gaussian set-up, by Theorem 5 applied to each variable, weights and nodes satisfy the polynomial system

$$\left\{ \begin{array}{l} H_{n_1}(x_1) = 0 \\ \lambda_1(x_1) H_{n_1-1}(x_1)^2 = \frac{(n_1 - 1)!}{n_1} \\ \vdots \\ H_{n_d}(x_d) = 0 \\ \lambda_d(x_d) H_{n_d-1}(x_d)^2 = \frac{(n_d - 1)!}{n_d} \end{array} \right. \quad (13)$$

For the grid set-up of this section and for the Gaussian case, in analogy to Example 4 some Fourier coefficients of polynomials of low enough degree can be determined exactly from the values of the polynomials on the grid points as shown in Example 8 below.

$$\begin{aligned}
1_{(0,0) \in \mathcal{F}}(x,y) &= \frac{2}{3}H_0 - \frac{1}{3}H_2(y) \\
1_{(\sqrt{3},\sqrt{3}) \in \mathcal{F}}(x,y) &= \frac{1}{12}H_0 + \frac{1}{12}\sqrt{3}H_1(x) + \frac{1}{12}\sqrt{3}H_1(y) + \frac{1}{12}H_1(x)H_1(y) + \frac{1}{12}H_2(y) \\
1_{(\sqrt{3},-\sqrt{3}) \in \mathcal{F}}(x,y) &= \frac{1}{12}H_0 - \frac{1}{12}\sqrt{3}H_1(x) + \frac{1}{12}\sqrt{3}H_1(y) - \frac{1}{12}H_1(x)H_1(y) + \frac{1}{12}H_2(y) \\
1_{(-\sqrt{3},\sqrt{3}) \in \mathcal{F}}(x,y) &= \frac{1}{12}H_0 + \frac{1}{12}\sqrt{3}H_1(x) - \frac{1}{12}\sqrt{3}H_1(y) - \frac{1}{12}H_1(x)H_1(y) + \frac{1}{12}H_2(y) \\
1_{(-\sqrt{3},-\sqrt{3}) \in \mathcal{F}}(x,y) &= \frac{1}{12}H_0 - \frac{1}{12}\sqrt{3}H_1(x) - \frac{1}{12}\sqrt{3}H_1(y) + \frac{1}{12}H_1(x)H_1(y) + \frac{1}{12}H_2(y)
\end{aligned}$$

Table 3 Indicator functions for Example 9

Example 8 Consider a square grid of size n , \mathcal{D}_{nm} , and a polynomial f of degrees in x and in y smaller than n , the Hermite polynomials and the standard normal distribution. Then we can write

$$f(x,y) = \sum_{i,j=0}^{n-1} c_{ij}H_i(x)H_j(y)$$

As both the degree in x of fH_k and the degree in y of fH_h are smaller than $2n-1$, we have

$$\mathbb{E}(f(Z_1, Z_2)H_k(X_1)H_h(X_2)) = c_{hk} \|H_k(X_1)\|^2 \|H_h(X_2)\|^2$$

$$c_{kh} = \frac{1}{k!h!} \sum_{(x,y) \in \mathcal{D}_{nm}} f(x,y)H_k(x)H_h(y)\lambda_x\lambda_y$$

Note if f is the indicator function of a fraction $\mathcal{F} \subset \mathcal{D}_{nm}$ then

$$c_{kh} = \frac{1}{k!h!} \sum_{(x,y) \in \mathcal{F}} H_k(x)H_h(y)\lambda_x\lambda_y \quad \text{with } 0 \leq h, k < n$$

Example 9 deals with a general design and introduces the more general theory of Section 6.

Example 9 Let \mathcal{F} be the zero set of

$$\begin{cases}
g_1 = x^2 - y^2 = H_2(x) - H_2(y) = 0 \\
g_2 = y^3 - 3y = H_3(y) = 0 \\
g_3 = xy^2 - 3x = H_1(x)(H_2(y) - 2H_0) = 0
\end{cases}$$

namely \mathcal{F} is given by the five points $(0,0)$, $(\pm\sqrt{3}, \pm\sqrt{3})$. Write a polynomial $f \in \mathbb{R}[x,y]$ as $f = \sum q_i g_i + r$ where $r(x,y) = f(x,y)$ for $(x,y) \in \mathcal{F}$ and

$$r \in \text{Span}(H_0, H_1(x), H_1(y), H_1(x)H_1(y), H_2(y)) = \text{Span}(1, x, y, xy, y^2).$$

If, furthermore, f is such that

$$\begin{aligned}
q_1(x,y) &= a_0 + a_1H_1(x) + a_2H_1(y) + a_3H_1(x)H_1(y) \\
q_2 &= \theta_1(x) + \theta_2(x)H_1(y) + \theta_3(x)H_2(y) \\
q_3 &= a_4 + a_5H_1(y)
\end{aligned}$$

with $a_i, \theta_j \in \mathbb{R}$ for $i = 0, \dots, 5$ and $j = 1, \dots, 3$, then

$$\mathbb{E}(g_i(Z_1, Z_2)q_i(Z_1, Z_2)) = 0, \quad i = 1, 2, 3$$

for Z_1 and Z_2 independent normally distributed random variables. Write r as a linear combination of the indicator functions of the points in \mathcal{F} , i.e.

$$r(x,y) = \sum_{(a,b) \in \mathcal{F}} f(a,b)1_{(a,b) \in \mathcal{F}}(x,y)$$

Each indicator function $1_{(a,b) \in \mathcal{F}}$ belongs to

$$\text{Span}(H_0, H_1(x), H_1(y), H_1(x)H_1(y), H_2(y))$$

and are shown in Table 3. Their expected values are given by the H_0 -coefficients. Furthermore, by linearity

$$\mathbb{E}(f(Z_1, Z_2)) = \mathbb{E}(r(Z_1, Z_2)) = \sum_{(a,b) \in \mathcal{F}} f(a,b) \mathbb{E}(1_{(a,b) \in \mathcal{F}}(Z_1, Z_2))$$

and we can conclude

$$\mathbb{E}(f(Z_1, Z_2)) = \mathbb{E}(r(Z_1, Z_2)) = 2 \frac{f(0,0)}{3} + \frac{f(\sqrt{3}, \sqrt{3}) + f(\sqrt{3}, -\sqrt{3}) + f(-\sqrt{3}, \sqrt{3}) + f(-\sqrt{3}, -\sqrt{3})}{12}$$

The key points in Example 9 are

1. determine the class of polynomial functions for which $\mathbb{E}(g_i(Z_1, Z_2)q_i(Z_1, Z_2)) = 0$ and
2. determine the H_0 -coefficients of the indicator functions of the points in \mathcal{F} .

In Section 6 we give algorithms to do this for any fraction \mathcal{F} .

6 Higher dimension: general design support

In the previous sections we considered particular designs whose sample points were zeros of orthogonal polynomials. In the Gaussian case we exploited the ring structure of the set of functions defined over the design in order to obtain recurrence formula and to write Fourier coefficients of higher order Hermite polynomials in terms of those of lower order Hermite polynomials (Example 4). Also we deduced a system of polynomial equations whose solution gives the weights of a quadrature formula. The mathematical tools that allowed this are Equation (6) and the particular structure it implies for Hermite polynomials on the recurrence relation for general, orthogonal polynomials

$$\pi_{k+1}(x) = (\gamma_k x - \alpha_k) \pi_k(x) - \beta_k \pi_{k-1}(x) \quad x \in \mathbb{R} \quad (14)$$

with $\gamma_k, \alpha_k \neq 0$ and $\alpha_k \gamma_k \gamma_{k-1} > 0$ (cf. Theorem 1).

In this section we switch focus and consider a generic set of points in \mathbb{R}^d as a design, or nodes for a cubature formula, and a generic set of orthogonal polynomials. We gain something and lose something. The essential computations are linear: such is the computation of a Gröbner basis for a finite set of distinct points [8]; the Buchberger Möller type of algorithm in Table 4 is based on finding solutions of linear systems of equations; in Section 6.1 we give a characterisation of polynomials with the same expected values which is a linear expression of some Fourier coefficients and a square free polynomial of degree two in a larger set of Fourier coefficients (see Equation 18).

Given a set of points and a term-ordering the algorithm in Table 4 returns the reduced Gröbner basis of the design ideal expressed as linear combination of orthogonal polynomial of low enough degree. It does so directly; that is, it computes the Gröbner basis by working only in the space of orthogonal polynomials.

We lose the equivalent of Theorem 4 for Hermite polynomials, in particular we do not know yet how to impose a ring structure on $\text{Span}(\pi_0, \dots, \pi_n)$ for generic orthogonal polynomials π and we miss a general formula to write the product $\pi_k \pi_n$ as linear combination of π_i with $i = 0, \dots, n \wedge k, n + k$, which is fundamental for the aliasing structure discussed for Hermite polynomials.

For multivariate cubature formulæ we refer e.g. to [13] and [14] which, together with [8], are basic references for this section. For clarity we repeat some basics and notation. Let λ be a one-dimensional probability measure and $\{\pi_n\}_{n \in \mathbb{Z}_{\geq 0}}$ be its associated orthogonal polynomial system. To a multi-index $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{Z}_{\geq 0}^d$ we associate the monomial $x^\alpha = x_1^{\alpha_1} \dots x_d^{\alpha_d}$ and the product of polynomials $\pi_\alpha(x) = \pi_{\alpha_1}(x_1) \dots \pi_{\alpha_d}(x_d)$. Note that $\{\pi_\alpha\}_\alpha$ is a system of orthogonal polynomials for the product measure λ^d . Theorem 7 describes the one-to-one correspondence between the x^α 's and the $\pi_\alpha(x)$'s.

Theorem 7 1. For $d = 1$ and $k \in \mathbb{Z}_{\geq 0}$, in the notation of Equation (14) we have that

$$x^k = \sum_{j=0}^k c_j(x^k) \pi_j(x)$$

where $c_0(x^0) = 1$, $c_{-1}(x^0) = c_1(x^0) = 0$, and, for $k = 1, 2, \dots$ and $j = 0, \dots, k-1$

$$c_{-1}(x^k) = c_{k+1}(x^k) = 0$$

$$c_j(x^k) = \frac{c_{j-1}(x^{k-1})}{\gamma_{j-1}} + \frac{c_j(x^{k-1})\alpha_j}{\gamma_j} + \frac{c_{j+1}(x^{k-1})\beta_{j+1}}{\gamma_{j+1}}$$

$$c_k(x^k) = \frac{1}{\gamma_0 \dots \gamma_{k-1}}$$

2. For $d > 1$, the monomial x^α is a linear combination of π_β , with $\beta \leq \alpha$ component wise, and vice versa. In formulae

$$\pi_\alpha = \sum_{\beta \leq \alpha} a_\beta x^\beta \quad \text{and} \quad x^\alpha = \sum_{\beta \leq \alpha} b_\beta \pi_\beta \quad (15)$$

where $\beta \leq \alpha$ holds component wise.

Proof The proof of Item 1 is by induction and that of Item 2 follows by rearranging the coefficients in the product. They are given in Appendix 8. \square

Example 10 If π_j is the j -th Hermite polynomial, then Item 1 of Theorem 7 gives the well known result

$$\begin{aligned} c_j(x^k) &= 0 && \text{if } k+j \text{ is odd} \\ c_j(x^k) &= \binom{k}{j} (k-j-1)!! && \text{if } k+j \text{ is even} \end{aligned}$$

Direct application of Theorem 7 is cumbersome and we need only to characterise the polynomial functions for which the cubature formula is exact. So we proceed by another way. The finite set of distinct points $\mathcal{D} \subset \mathbb{R}^d$ is associated to its vanishing polynomial ideal

$$\mathcal{I}(\mathcal{D}) = \{f \in \mathbb{R}[x] : f(z) = 0 \text{ for all } z \in \mathcal{D}\}$$

Let $LT_\sigma(f)$ or $LT(f)$ denote the largest term in a polynomial f with respect to a term-ordering σ . Let $[f(z)]_{z \in \mathcal{D}}$ be the evaluation vector of the polynomial f at \mathcal{D} and for a finite set of polynomials $G \subset \mathbb{R}[x]$ let $[g(z)]_{z \in \mathcal{D}, g \in G}$ be the evaluation matrix whose columns are the evaluation vectors at \mathcal{D} of the polynomials in G . In DoE often this matrix is called the X -matrix of \mathcal{D} and G .

As mentioned at the end of Section 1.1, the space $\mathcal{L}(\mathcal{D})$ of real valued functions defined over \mathcal{D} is a linear space and particularly important vector space bases can be constructed as follows. Let $LT(\mathcal{I}(\mathcal{D})) = \langle LT_\sigma(f) : f \in \mathcal{I}(\mathcal{D}) \rangle$. If G is the σ -reduced Gröbner basis of $\mathcal{I}(\mathcal{D})$, then $LT(\mathcal{I}(\mathcal{D})) = \langle LT_\sigma(f) : f \in G \rangle$. Now we can define two interesting vector

space bases of $\mathcal{L}(\mathcal{D})$. Let $L = \{\alpha \in \mathbb{Z}_{\geq 0}^d : x^\alpha \notin LT(\mathcal{I}(\mathcal{D}))\}$ and define

$$\mathcal{B} = \{x^\alpha : \alpha \in L\} \quad \text{and} \quad \mathcal{O}\mathcal{B} = \{\pi_\alpha : \alpha \in L\}$$

Example 11 For $d = 2$ and $L = \{(0,0), (1,0), (0,1), (2,0)\}$ we have $\mathcal{B} = \{1, x, y, x^2\}$ and $\mathcal{O}\mathcal{B} = \{1, \pi_1(x), \pi_1(y), \pi_2(x)\}$, since $\pi_0(x) = \pi_0(y) = 1$.

The sets L , \mathcal{B} and $\mathcal{O}\mathcal{B}$ depend on σ . It is well known that if $t \in \mathcal{B}$ and r divides t , then $r \in \mathcal{B}$; it follows that if $\alpha \in L$ and $\beta \leq \alpha$ component wise then also β belongs to L and π_β to $\mathcal{O}\mathcal{B}$. Note that σ induces a total ordering also on the orthogonal polynomials: $\pi_\alpha <_\sigma \pi_\beta$ if and only if $x^\alpha <_\sigma x^\beta$; analogously $\alpha <_\sigma \beta$ if and only if $x^\alpha <_\sigma x^\beta$ for each $\alpha, \beta \in \mathbb{Z}_{\geq 0}^d$. Here we used the same symbol to indicate related orderings over the α 's, the x^α 's and the π_α 's. Further, given $\alpha \leq \beta$ componentwise, since x^α divides x^β and since $1 \leq_\sigma x^{\beta-\alpha}$, we have $x^\alpha \leq_\sigma x^\beta$, that is $\alpha \leq_\sigma \beta$.

Now, given a term-ordering σ , any $g \in G$ can be uniquely written as its leading term, $x^\alpha = LT(g)$, and tail which is a linear combination of terms in \mathcal{B} preceding $LT(g)$ in σ , that is $g = x^\alpha + \sum_{\beta \in L, \beta <_\sigma \alpha} a_\beta x^\beta$ with $a_\beta \in \mathbb{R}$.

Theorem 8 provides an alternative to the classical method of rewriting a polynomial f in terms of orthogonal polynomials. It does so by applying Theorem 7 and by substituting each monomial in f . Theorem 8 gives linear rules to write the elements of G and the remainder of a polynomial divided by G as linear combinations of orthogonal polynomials of low enough order. The proof is in Appendix 8.

Theorem 8

1. $\text{Span}(\mathcal{B}) = \text{Span}(\mathcal{O}\mathcal{B})$;
2. Let G be the reduced σ -Gröbner basis of $\mathcal{I}(\mathcal{D})$. Each $g \in G$ with $LT(g) = x^\alpha$ is uniquely written as

$$g = \pi_\alpha - \sum_{\beta \in L, \beta <_\sigma \alpha} b_\beta \pi_\beta$$

where $b = [b_\beta]_{\beta \in L, \beta <_\sigma \alpha}$ solves the linear system

$$[\pi_\beta(z)]_{z \in \mathcal{D}, \beta \in L, \beta <_\sigma \alpha} b = [\pi_\alpha(z)]_{z \in \mathcal{D}};$$

in words the coefficient matrix is the evaluation matrix over \mathcal{D} of the orthogonal polynomials π_β with x^β in tail of g and the vector of constant terms is the evaluation vector of π_α .

3. Let $p \in \mathbb{R}[x]$ be a polynomial and $[p(z)]_{z \in \mathcal{D}}$ its evaluation vector. The polynomial p^* defined as

$$p^* = \sum_{\beta \in L} a_\beta \pi_\beta \quad (16)$$

where $a = [a_\beta]_{\beta \in L}$ solves the linear system

$$[\pi_\beta(z)]_{z \in \mathcal{D}, \beta \in L} a = [p(z)]_{z \in \mathcal{D}}$$

is the unique polynomial belonging to $\text{Span}(\mathcal{O}\mathcal{B})$ such that $p^*(z) = p(z)$ for all $z \in \mathcal{D}$.

Theorem 8 provides a pseudo-algorithm to compute a Gröbner basis for $\mathcal{I}(\mathcal{D})$ and the interpolating polynomial at \mathcal{D} in terms of orthogonal polynomials of low order directly from \mathcal{D} and $\mathcal{O}\mathcal{B}$. Table 4 gives the algorithm which is a variation of the Buchberger-Möller algorithm [8]. It starts with a finite set of distinct points \mathcal{D} and a term-ordering σ and it returns L and the expressions $g = \pi_\alpha - \sum_{\beta \in L, \beta <_\sigma \alpha} b_\beta \pi_\beta$ for g in the reduced σ -Gröbner basis of $\mathcal{I}(\mathcal{D})$. It does so by performing linear operations. If the real vector $[p(z)]_{z \in \mathcal{D}}$ is assigned, then the expression $p^* = \sum_{\beta \in L} a_\beta \pi_\beta$ can now be found using Item 3 in Theorem 8. This permits to rewrite every polynomial $p \in \mathbb{R}[x]$ as a linear combination of orthogonal polynomials.

The algorithm in Table 4 returns the σ -reduced Gröbner basis as linear combination of orthogonal polynomials. It performs operations only with orthogonal polynomials and in particular it does not involve at any step the x^α monomials. This is computationally faster than first computing a classical Gröbner basis in the x^α and next substituting the π_α . Furthermore working with only one vector space basis, and not switching between the x^α and the π_α , is conceptually appealing.

Input: a set \mathcal{D} of distinct points in \mathbb{R}^d , a term-ordering σ and any vector norm $\|\cdot\|$.

Output: the reduced σ -Gröbner basis G of $\mathcal{I}(\mathcal{D})$ as linear combination of orthogonal polynomials and the set L .

Step 1 Let $L = \{0 \in \mathbb{Z}_{\geq 0}^d\}$, $\mathcal{O}\mathcal{B} = [1]$, $G = []$ and $M = [x_1, \dots, x_d]$.

Step 2 If $M = []$ stop; else set $x^\alpha = \min_\sigma(M)$ and delete x^α from M .

Step 3 Solve in b the overdetermined linear system $[\pi_\beta(z)]_{z \in \mathcal{D}, \beta \in L} b = [\pi_\alpha(z)]_{z \in \mathcal{D}}$ and compute the residual

$$\rho = [\pi_\alpha(z)]_{z \in \mathcal{D}} - [\pi_\beta(z)]_{z \in \mathcal{D}, \beta \in L} b$$

Step 4

1. If $\|\rho\| > 0$, then include α in L , and include in M those elements of $\{x_1 x^\alpha, \dots, x_d x^\alpha\}$ which are not multiples of an element in M or of $LT(g)$, $g \in G$. Return to Step 2.
2. If $\|\rho\| = 0$, then include in G the polynomial

$$g = \pi_\alpha - \sum_{\beta \in L} b_\beta \pi_\beta$$

where the values b_β , $\beta \in L$, are the components of the solutions b of the linear system in Step 3. Delete from M all multiples of x^α .

Table 4 Buchberger-Möller algorithm using orthogonal polynomials

Summarising: given a function f , a finite set of distinct points $\mathcal{D} \subset \mathbb{R}^d$ and a term-ordering σ , a probability product measure λ^d over \mathbb{R}^d , its system of product orthogonal polynomials, and a random vector with probability distribution λ^d , then the expected value of f with respect to λ^d can be approximated by

1. computing L with the algorithm in Table 4 and
2. determining, by solving the linear system

$$[\pi_\beta(z)]_{z \in \mathcal{D}, \beta \in L} a = [f(z)]_{z \in \mathcal{D}},$$

the unique polynomial p^* such that $p^*(z) = f(z)$ for all $z \in \mathcal{D}$. The polynomial p^* is expressed as linear combination of orthogonal polynomials.

3. The coefficient a_0 of π_0 is the required approximation.

Recall that $p^*(x) = \sum_{z \in \mathcal{D}} f(z) l_z(x)$ is a linear combination of the indicator functions of the points in \mathcal{D} (Lagrange polynomials) and hence $a_0 = \sum_{z \in \mathcal{D}} f(z) \mathbb{E}(l_z(X))$. In particular, $\mathbb{E}(l_z(X)) = \lambda_z$, $z \in \mathcal{D}$, can be computed by applying the above to $f = l_z$. Notice however that as λ^d is a product measure, the λ_z can be obtained from the one-dimensional ones as noticed before Theorem 6. It would be interesting to generalise this section to non-product measures.

Here an algorithm has been provided to approximate the expected value of polynomials. Next the set of polynomials whose expected value coincides with the value of the cubature formula is characterised. In Section 6.1 we provide a characterisation of the full set via linear relationships that Fourier coefficients of suitable polynomials have to satisfy, while in Section 6.2 a possibly proper subset has been characterised via a simple condition on the total degree of the polynomials.

6.1 Characterisation of polynomial functions with zero expectation

In this section we characterise the set of polynomials whose expected value coincides with the value of the cubature formula. As mentioned in Section 1.1 given $\mathcal{D} \in \mathbb{R}^d$, its vanishing ideal $\mathcal{I}(\mathcal{D})$, a term-ordering σ and the Gröbner basis G of $\mathcal{I}(\mathcal{D})$ with respect to σ , then any polynomial $f \in \mathbb{R}[x]$ can be written as

$$f(x) = \sum_{g \in G} q_g(x) g(x) + r_\sigma(x)$$

where $r_\sigma(x)$ is unique in $\text{Span}(\mathcal{B})$ such that $r_\sigma(z) = f(z)$ for all $z \in \mathcal{D}$ and can be written as $r_\sigma(x) = \sum_{z \in \mathcal{D}} f(z) l_z(x)$, where the l_z 's are the product Lagrange polynomials in Section 5. Theorem 8 states how to write r_σ over $\mathcal{O}\mathcal{B}$.

If $f \in \mathbb{R}[x]$ is such that $\mathbb{E}_\lambda(f(X)) = \mathbb{E}_\lambda(r_\sigma(X))$ then we have $\mathbb{E}_\lambda(f(X) - r_\sigma(X)) = 0$, where $f - r_\sigma \in \mathcal{I}(\mathcal{D})$. Hence to study the set

$$\mathcal{E}_\sigma = \{f \in \mathbb{R}[x] : \mathbb{E}_\lambda(f(X)) = \mathbb{E}_\lambda(r_\sigma(X))\}$$

is equivalent to characterize the set

$$\mathcal{E}_0 = \{p \in \mathcal{I}(\mathcal{D}) : \mathbb{E}_\lambda(p(X)) = 0\}$$

We do this in two ways. First we study the Fourier expansion of the elements of \mathcal{E}_0 , next we present some results about the degree of the elements belonging to \mathcal{E}_σ .

The elements of \mathcal{E}_0 are characterized in Theorem 9. Note that if $f \in \mathbb{R}[x]$ is such that $f = p + r_\sigma$ with $p \in \mathcal{E}_0$ and $r_\sigma \in \text{Span}(\mathcal{B})$ then by linearity and independence

$$\mathbb{E}_\lambda(f) = \sum_{(z_1, \dots, z_n) \in \mathcal{D}} f(z_1, \dots, z_n) \lambda_{z_1}^{n_1} \cdots \lambda_{z_d}^{n_d}$$

Theorem 9 *Let λ be a product probability measure with product orthogonal polynomials $\pi_\alpha(x)$, $\alpha \in \mathbb{Z}_{\geq 0}^d$ and let X be a random vector with distribution λ^d . Let $\mathcal{D} \subset \mathbb{R}^d$ be a set of distinct points, σ a term-ordering, G the σ -reduced Gröbner basis of $\mathcal{I}(\mathcal{D})$ whose elements as linear combinations of orthogonal polynomials. Thus for $g \in G$ and $x^\alpha = LT(g)$ we write*

$$g = \pi_\alpha - \sum_{\alpha >_\sigma \beta \in L} c_\beta(g) \pi_\beta$$

where $\alpha >_\sigma \beta \in L$ stands for $\alpha >_\sigma \beta$ and $\beta \in L$.

Let $p = \sum_{g \in G} q_g g \in \mathcal{I}(\mathcal{D})$ for suitable $q_g \in \mathbb{R}[x]$, and consider the Fourier expansion of each q_g , $g \in G$,

$$q_g = \sum_{\beta \in \mathbb{Z}_{\geq 0}^d} c_\beta(q_g) \pi_\beta \quad (17)$$

Then $\mathbb{E}_\lambda(p(X)) = 0$ if and only if

$$\sum_{g \in G} \|\pi_\alpha\|_\lambda^2 c_\alpha(q_g) - \sum_{g \in G} \sum_{\alpha >_\sigma \beta \in L} \|\pi_\beta\|_\lambda^2 c_\beta(q_g) c_\beta(g) = 0 \quad (18)$$

Proof The key observation is that $\mathbb{E}_\lambda(\pi_m \pi_n) = 0$ if $n \neq m$ and then linearity of \mathbb{E}_λ is used. The proof can be found in Appendix 8. \square

Importantly, only terms of low enough Fourier order in Equation (17) matter for the computation of the expectation.

Example 12 Consider Z_1 and Z_2 two independent standard normal random variables and hence the Hermite polynomials. Consider also the five point design

$$\mathcal{D} = \{(-6, -1), (-5, 0), (-2, 1), (3, 2), (10, 3)\}$$

and the $\sigma = \text{DegLex}$ term-ordering over the monomials in $\mathbb{R}[x, y]$. The algorithm in Table 4 gives $G = \{g_1, g_2, g_3\}$ where

$$g_1 = H_2(y) - H_1(x) + 2H_1(y) - 4$$

$$g_2 = H_2(x)H_1(y) - 9H_2(x) + 47H_1(x)H_1(y) - 123H_1(x) + 271H_1(y) - 399$$

$$g_3 = H_3(x) - 47H_2(x) + 300H_1(x)H_1(y) - 845H_1(x) + 2040H_1(y) - 2987$$

and $L = \{(0, 0), (1, 0), (0, 1), (1, 1), (0, 2)\}$, that is

$$\mathcal{O}\mathcal{B} = \{1, H_1(y), H_1(x), H_1(x)H_1(y), H_2(x)\}$$

$$\begin{array}{ccccccc}
c_0^{(1)} = -34 & c_{(1,0)}^{(1)} = 0 & c_{(0,1)}^{(1)} = -2 & c_{(0,0)}^{(1)} = -8 & & & \\
c_0^{(2)} = 0 & c_{(2,0)}^{(2)} = 0 & c_{(1,1)}^{(2)} = 1 & c_{(1,0)}^{(2)} = -2 & c_{(0,1)}^{(2)} = 1 & c_{(0,0)}^{(2)} = 2 & \\
c_0^{(3)} = 10 & c_{(2,0)}^{(3)} = 2 & c_{(1,1)}^{(3)} = 0 & c_{(1,0)}^{(3)} = -5 & c_{(0,1)}^{(3)} = 1 & c_{(0,0)}^{(3)} = 5863/2987 &
\end{array}$$

Table 5 A solution for (18)

By Theorem 9 for the purpose of computing its expectation a polynomial $p = q_1g_1 + q_2g_2 + q_3g_3 \in \mathcal{S}(\mathcal{D})$ can be simplified to have the form

$$\begin{aligned}
p = & \left(c_0^{(1)}H_2(y) + c_{(1,0)}^{(1)}H_1(x) + c_{(0,1)}^{(1)}H_1(y) + c_{(0,0)}^{(1)} \right) g_1 + \\
& \left(c_0^{(2)}H_2(x)H_1(y) + c_{(2,0)}^{(2)}H_2(x) + c_{(1,1)}^{(2)}H_1(x)H_1(y) + \right. \\
& \left. c_{(1,0)}^{(2)}H_1(x) + c_{(0,1)}^{(2)}H_1(y) + c_{(0,0)}^{(2)} \right) g_2 + \left(c_0^{(3)}H_3(x) + \right. \\
& \left. c_{(2,0)}^{(3)}H_2(x) + c_{(1,1)}^{(3)}H_1(x)H_1(y) + c_{(1,0)}^{(3)}H_1(x) + \right. \\
& \left. c_{(0,1)}^{(3)}H_1(y) + c_{(0,0)}^{(3)} \right) g_3
\end{aligned}$$

and furthermore by Equation (18)

$$\begin{aligned}
& c_0^{(1)}2! - c_{(1,0)}^{(1)} + 2c_{(0,1)}^{(1)} - 4c_{(0,0)}^{(1)} + c_0^{(2)}2! - 9c_{(2,0)}^{(2)}2! + 47c_{(1,1)}^{(2)} \\
& - 123c_{(1,0)}^{(2)} + 271c_{(0,1)}^{(2)} - 399c_{(0,0)}^{(2)} + c_0^{(3)}3! - 47c_{(2,0)}^{(3)}2! \\
& + 300c_{(1,1)}^{(3)} - 845c_{(1,0)}^{(3)} + 2040c_{(0,1)}^{(3)} - 2987c_{(0,0)}^{(3)} = 0
\end{aligned}$$

In practice, for $i = 1, 2, 3$, put coefficients of g_i and q_i in two vectors, multiply them component wise and sum the result. There are infinite polynomials that satisfy the above equations, one such polynomial is given in Table 5. The above equation involves only a finite number of Fourier coefficients, namely $c_\beta(q_g)$'s is relevant for the equation if and only if $\beta \in L$ and x^β is smaller in σ than the leading terms of $g \in G$. Hence if to q_g we add a polynomial of the form $\sum_{\beta > \alpha \text{ or } \beta \notin L} c_\beta H_\beta$ we still obtain a zero mean polynomial. That is, we can modify q_g by adding high enough terms without changing the mean value.

For example by adding $H_4(x)$ to q_1 and $H_4(y)$ to q_2 we obtain the following zero mean polynomial

$$\begin{aligned}
p(Z_1, Z_2) = & Z_1^2Z_2^5 + 10Z_1^6 + Z_1^4Z_2^2 - 9Z_1^2Z_2^4 + 47Z_1Z_2^5 - 469Z_1^5 + \\
& 3002Z_1^4Z_2 + Z_1^3Z_2^2 - 6Z_1^2Z_2^3 - 123Z_1Z_2^4 + 270Z_2^5 - \\
& 8614Z_1^4 + 20990Z_1^3Z_2 + 96Z_1^2Z_2^2 - 282Z_1Z_2^3 - 424Z_2^4 \\
& - 87898560/2987Z_1^3 - 6700Z_1^2Z_2 + 1389Z_1Z_2^2 - 1690Z_2^3 + \\
& \frac{71785814}{2987}Z_1^2 - \frac{218275468}{2987}Z_1Z_2 + 4845Z_2^2 + \\
& \frac{307862660}{2987}Z_1 - \frac{5937584}{2987}Z_2 - \frac{5931425}{2987}
\end{aligned}$$

6.2 On exactness of cubature formulae

Here we adopt another viewpoint and characterise the set \mathcal{E}_σ . Instead of studying the Fourier expansion of its polynomials, we focus our attention on their degree. Given a degree compatible term ordering σ , we show how to compute the maximum degree s such that $\mathbb{R}[x]_{\leq s}$ is in \mathcal{E}_σ , that is the degree of the cubature formula with nodes \mathcal{D} . Our strategy is based on the definition of s -orthogonal polynomials [14].

Definition 1 A polynomial $g \in \mathbb{R}[x]$ is s -orthogonal if $s \in \mathbb{Z}_{>0}$ is the maximum integer such that

$$fg \in \mathbb{R}[x]_{\leq s} \quad \text{implies} \quad \mathbb{E}_\lambda(f(X)g(X)) = 0$$

Furthermore, a set G of polynomials is s -orthogonal if each $g \in G$ is $s(g)$ -orthogonal and $s = \min_{g \in G} s(g)$.

Theorem 10 reformulates and summarizes two theorems about the degree of a cubature formula presented in [13] and [14] (where H-bases are considered).

Theorem 10 Given a set \mathcal{D} and a degree compatible term ordering σ the following conditions are equivalent.

1. $\mathbb{R}[x]_{\leq s} \subset \mathcal{E}_\sigma$;
2. $\mathbb{E}_\lambda(f(X)) = 0$ for all $f \in \mathcal{S}(\mathcal{D}) \cap \mathbb{R}[x]_{\leq s}$;
3. the σ -Gröbner basis G of $\mathcal{S}(\mathcal{D})$ is s -orthogonal.

Proof 1 \Rightarrow 2. Let $f \in \mathcal{S}(\mathcal{D}) \cap \mathbb{R}[x]_{\leq s}$. Since by hypothesis $\mathbb{R}[x]_{\leq s} \subset \mathcal{E}_\sigma$, then $f \in \mathcal{E}_\sigma$ that is $\mathbb{E}_\lambda(f(X)) = \mathbb{E}_\lambda(r_\sigma(X))$. Since $f \in I(\mathcal{D})$, we have $r_\sigma(x) \equiv 0$ and so $\mathbb{E}_\lambda(f(X)) = 0$.

2 \Rightarrow 3. For each $g \in G$ let f be such that $fg \in \mathbb{R}[x]_{\leq s}$. Since $fg \in \mathcal{S}(\mathcal{D}) \cap \mathbb{R}[x]_{\leq s}$ then $\mathbb{E}_\lambda(f(X)g(X)) = 0$ and so g is s -orthogonal.

3 \Rightarrow 1. For $p \in \mathbb{R}[x]_{\leq s}$, we have that

$$p = \sum_{g \in G} gq_g + r_\sigma$$

where each gq_g is such that $LT_\sigma(gq_g) \leq LT_\sigma(p)$ and so, since σ is degree compatible, $\deg(gq_g) \leq \deg(p) \leq s$. It follows that, since G is s -orthogonal, $\mathbb{E}_\lambda(gq_g) = 0$. By linearity we obtain $\mathbb{E}_\lambda(p(X)) = \mathbb{E}_\lambda(r_\sigma(X))$, that is $p \in \mathcal{E}_\sigma$. \square

Remark 3 The maximum integer s such that $\mathbb{E}_\lambda(f(X)) = 0$ for each $f \in \mathcal{S}(\mathcal{D}) \cap \mathbb{R}[x]_{\leq s}$ is the degree of the cubature formula with nodes \mathcal{D} and with respect to σ .

Theorem 10 shows that the maximum integer s such that all polynomials of total degree s are in \mathcal{E}_σ coincides with the maximum s such that G is s -orthogonal. Hence we focus our attention on the s -orthogonality of the elements of G .

Theorem 11 *Each polynomial $g \in \mathbb{R}[x]$ is s -orthogonal with $\deg(g) - 1 \leq s < 2 \deg(g)$.*

Proof Since $fg \in \mathbb{R}[x]_{\leq (\deg(g)-1)}$ if and only if f is the identically zero polynomial, then $fg \in \mathbb{R}[x]_{\leq (\deg(g)-1)}$ implies $\mathbb{E}_\lambda(f(X)g(X)) = 0$, that is g is always s -orthogonal with $s \geq \deg(g) - 1$. Moreover, $s < 2 \deg(g)$; in fact, g^2 belongs to $\mathbb{R}[x]_{\leq 2 \deg(g)}$ and, from the orthogonality of the polynomials π_α , we have

$$\mathbb{E}_\lambda(g(X)g(X)) = \sum_{\alpha} (c_{\alpha}(g))^2 \|\pi_{\alpha}\|^2$$

As g is not identically zero, then $\mathbb{E}_\lambda(g(X)g(X)) \neq 0$ and so g cannot be $2 \deg(g)$ -orthogonal. \square

The following theorem shows how to detect the s -orthogonality of a polynomial analysing its Fourier coefficients.

Theorem 12 *Let $g = \sum_{\alpha} c_{\alpha}(g)\pi_{\alpha}$ the Fourier expansion of a polynomial $g \in \mathbb{R}[x]$.*

1. *If $c_0(g) \neq 0$, the polynomial g is $(\deg(g) - 1)$ -orthogonal.*
2. *If $c_0(g) = 0$, the polynomial g is s -orthogonal where s is such that $c_{\alpha}(g) = 0$ for all α s.t. $\sum_{i=1}^d \alpha_i \leq s - \deg(g)$.*

Proof 1. Let f be a constant polynomial, $f \equiv 1$. We have that $fg \in \mathbb{R}[x]_{\leq \deg(g)}$ implies

$$\mathbb{E}_\lambda(f(X)g(X)) = \mathbb{E}_\lambda(g(X)) = c_0(g) \neq 0$$

and so g is not $\deg(g)$ -orthogonal. From Theorem 11 we conclude that g is $(\deg(g) - 1)$ -orthogonal.

2. Let $s < 2 \deg(g)$ and let f be a polynomial such that $fg \in \mathbb{R}[x]_{\leq s}$. Since $\deg(f) \leq s - \deg(g)$ the Fourier expansion of f is such that $f = \sum_{|\alpha| \leq s - \deg(g)} c_{\alpha}(f)\pi_{\alpha}$. From the orthogonality of the polynomials π_{α} and from $s - \deg(g) < \deg(g)$ we have

$$\mathbb{E}_\lambda(f(X)g(X)) = \sum_{|\alpha| \leq s - \deg(g)} c_{\alpha}(f)c_{\alpha}(g)\|\pi_{\alpha}\|^2$$

The generality of f implies that $\mathbb{E}_\lambda(f(X)g(X)) = 0$ only if $c_{\alpha}(g) = 0$ for each α such that $|\alpha| \leq s - \deg(g)$. \square

Corollary 2 *Given a finite set of distinct points $\mathcal{D} \in \mathbb{R}^d$ and a degree compatible term ordering σ on $\mathbb{R}[x]$, let G be the reduced σ -Gröbner basis of $\mathcal{S}(\mathcal{D})$. Then the maximum integer s_c such that $\mathbb{R}[x]_{\leq s_c}$ is in \mathcal{E}_σ is*

$$s_c = \min_{g \in G} s(g)$$

where $s(g)$ is such that the Fourier expansion of each $g \in G$ is given by

$$g = \sum_{\alpha} c_{\alpha}(g)\pi_{\alpha}$$

with $c_{\alpha}(g) = 0$ for all α such that $\sum_{i=1}^d \alpha_i \leq s(g) - \deg(g)$.

Proof By Theorem 10 the thesis follows if G is s_c -orthogonal. But this is true since, from Theorem 12, we have that each $g \in G$ is $s(g)$ -orthogonal. \square

Example 13 Given the set of points

$$\mathcal{D} = \{(-1, 0), (-1, -2), (1, -1 + \sqrt{3}), (1, -1 - \sqrt{3}), (2, 1)\}$$

and the degree compatible term ordering $\sigma = \text{DegLex}$, then by e.g. the algorithm in Table 4 the reduced σ -Gröbner basis G of $\mathcal{S}(\mathcal{D})$ can be written as

$$G = \begin{cases} g_1 = H_2(y) - H_1(x) + 2H_1(y) \\ g_2 = H_2(x)H_1(y) - H_2(x) \\ g_3 = H_3(x) - 2H_2(x) + 2H_1(x) \end{cases}$$

Since for each $g \in G$, $s(g)$ is such that $c_{\alpha}(g) = 0$ for each α s.t. $\sum_{i=1}^d \alpha_i \leq s(g) - \deg(g)$, we have that $s(g_1) = 2$, $s(g_2) = 4$ and $s(g_3) = 3$. It follows that G is s_c -orthogonal, with $s_c = \min\{s(g_1), s(g_2), s(g_3)\} = 2$ and so, from Corollary 2 it follows that the maximum integer s such that $\mathbb{R}[x]_{\leq s} \subset \mathcal{E}_\sigma$ is 2.

Example 14 For \mathcal{D} of Example 12 and $\sigma = \text{DegLex}$, the reduced σ -Gröbner basis $G = \{g_1, g_2, g_3\}$ of $\mathcal{S}(\mathcal{D})$ is such that $s(g_1) = 1$, $s(g_2) = 2$ and $s(g_3) = 3$ since $c_0(g) \neq 0$ for all $g \in G$. It follows that G is 1-orthogonal and so the cubature formula w.r.t. σ and \mathcal{D} is exact for all the polynomials with degree 0 and 1. Nevertheless let us remark that the cubature formula is exact for a much larger class of polynomials as shown in Theorem 9 and in Example 12.

7 Conclusion

In this paper we mixed tools from Computational Commutative Algebra, orthogonal polynomial theory and Probability to address the recurrent statistical problem of estimation of mean values of polynomial functions. Our work shares great similarity with applications of computational algebra to design and analysis of experiments, which inspired us with a non-classical viewpoint to cubature formulae.

We obtained two main results. In the Gaussian case we obtained a system of polynomial equations whose solution gives the weights of a quadrature formula (Theorem 13). For a finite product measure which admits an orthogonal system of polynomials, we characterise the set of polynomials with the same mean value. This depends substantially from Equation (18) and it is in terms of Fourier coefficients of particular polynomials obtained by adapting Gröbner basis theory.

8 Appendix: proofs

Theorem 7:

Proof 1. The proof is by induction on the monomial degree k . From the three terms recurrence formula $\pi_{j+1} = (\gamma_j x - \alpha_j)\pi_j - \beta_j \pi_{j-1}$ we have

$$x\pi_j = \frac{\pi_{j+1}}{\gamma_j} + \frac{\alpha_j}{\gamma_j}\pi_j + \frac{\beta_j}{\gamma_j}\pi_{j-1}$$

For $k = 0$ we have $x^0 = \pi_0(x) = c_0(x^0)\pi_0$. For $k = 1$ from the three terms recurrence formula we have

$$x = x\pi_0 = \frac{\pi_1}{\gamma_0} + \frac{\alpha_0}{\gamma_0}\pi_0 = c_1(x)\pi_1 + c_0(x)\pi_0$$

In the inductive step the thesis holds for k and we prove it for $k + 1$. From the three terms recurrence formula we have

$$\begin{aligned} x^{k+1} &= x x^k = \sum_{j=0}^k c_j(x^k) x \pi_j \\ &= \sum_{j=0}^k c_j(x^k) \left(\frac{\pi_{j+1}}{\gamma_j} + \frac{\alpha_j}{\gamma_j} \pi_j + \frac{\beta_j}{\gamma_j} \pi_{j-1} \right) \\ &= \sum_{j=1}^{k+1} \frac{c_{j-1}(x^k)}{\gamma_{j-1}} \pi_j + \sum_{j=0}^k c_j(x^k) \frac{\alpha_j}{\gamma_j} \pi_j + \sum_{j=0}^{k-1} c_{j+1}(x^k) \frac{\beta_{j+1}}{\gamma_{j+1}} \pi_j \\ &= \sum_{j=1}^{k-1} \left(\frac{c_{j-1}(x^k)}{\gamma_{j-1}} + c_j(x^k) \frac{\alpha_j}{\gamma_j} + c_{j+1}(x^k) \frac{\beta_{j+1}}{\gamma_{j+1}} \right) \pi_j \\ &\quad + \frac{c_{k-1}(x^k)}{\gamma_{k-1}} \pi_k + \frac{c_k(x^k)}{\gamma_k} \pi_{k+1} \\ &\quad + \frac{c_0(x^k) \alpha_0}{\gamma_0} \pi_0 + \frac{c_k(x^k) \alpha_k}{\gamma_k} \pi_k + \frac{c_1(x^k) \beta_1}{\gamma_1} \pi_0 \\ &= \sum_{j=1}^{k-1} c_j(x^{k+1}) \pi_j + c_{k+1}(x^{k+1}) \pi_{k+1} \\ &\quad + \left(\frac{c_{k-1}(x^k)}{\gamma_{k-1}} + \frac{c_k(x^k) \alpha_k}{\gamma_k} \right) \pi_k \\ &\quad + \left(\frac{c_0(x^k) \alpha_0}{\gamma_0} + \frac{c_1(x^k) \beta_1}{\gamma_1} \right) \pi_0 \end{aligned}$$

This concludes the proof of the first part of the theorem.

To prove the second part we apply what we just proved and unfold the multiplication. Given $x^\alpha = x_1^{\alpha_1} \cdots x_d^{\alpha_d}$, the polynomial $\pi_\alpha = \pi_{\alpha_1}(x_1) \cdots \pi_{\alpha_d}(x_d)$ is the product of d univariate polynomials π_{α_j} each of degree α_j in x_j , $j = 1, \dots, d$. Clearly if $\alpha_j = 0$ then $\pi_{\alpha_j} = 1$ and x_j does not divide x^α . Furthermore we have

$$\pi_\alpha = \prod_{j=1}^d \sum_{k=0}^{\alpha_j} d_k^{(j)} x_j^k$$

We deduce that π_α is a linear combination of x^α and of the power products which divide x^α , that is of power products

x^β with $\beta \leq \alpha$ component wise. Vice versa, applying the first part of the theorem we have

$$x^\alpha = \prod_{k=1}^d x_k^{\alpha_k} = \prod_{k=1}^d \left[\sum_{j_k=0}^{\alpha_k} c_{j_k}(x_k^{\alpha_k}) \pi_{j_k}(x_k) \right]$$

and commuting product with sum shows that x^α is a linear combination of products of $\pi_{\beta_i}(x_i)$ where $\beta = (\beta_1, \dots, \beta_d)$ is such that $\beta \leq \alpha$ component wise, that is x^β divides x^α . \square

Theorem 8:

Proof Recall that \mathcal{B} and $\mathcal{O}\mathcal{B}$ are defined in terms of a common set L of d -dimensional vectors with non-negative integer entries satisfying the property of ‘factor-closeness’, that is if $(\alpha_1, \dots, \alpha_d) \in L$ and if $\beta_i \leq \alpha_i$ for all $i = 1, \dots, d$ then $(\beta_1, \dots, \beta_d) \in L$.

1. If $x^\alpha \in \mathcal{B}$ for some α , then from Theorem 7 $x^\alpha = \sum_{\beta \leq \alpha} b_\beta \pi_\beta$ follows. Since $\beta \leq \alpha$ then $\beta \in L$ and so each $\pi_\beta \in \mathcal{O}\mathcal{B}$: we have that x^α belongs to $\text{Span}(\mathcal{O}\mathcal{B})$. The vice versa is proved analogously.

2. The matrix $[\pi_\beta(z)]_{z \in \mathcal{D}, \beta \in L}$ is a square matrix since L has as many elements as \mathcal{D} and has full rank. The linear independence of the columns of this matrix follows from the fact that each linear combination of its columns corresponds to a polynomial in $\text{Span}(\mathcal{O}\mathcal{B})$ which coincides with $\text{Span}(\mathcal{B})$.

Any polynomial $g \in G$ can be written as

$$g = x^\alpha - \sum_{\alpha >_\sigma \beta \in L} c_\beta x^\beta$$

where $x^\alpha = LT(g)$ is a multiple by some x_j of an element of \mathcal{B} . By Theorem 7 we have

$$g = \sum_{\gamma \leq \alpha} a_\gamma^{(g)} \pi_\gamma - \sum_{\alpha >_\sigma \beta \in L} c_\beta \sum_{\gamma \leq \beta} d_\gamma^{(g)} \pi_\gamma$$

The polynomial π_α appears only in the first sum with the coefficient 1. For the other terms in the first sum observe that as $\gamma < \alpha$ then $\gamma \in L$ and also $\gamma <_\sigma \alpha$. Analogously, for the second sum consider $\gamma \leq \beta < \alpha$; since $\beta \in L$ then $\gamma \in L$ and since $\gamma < \alpha$ then $\gamma <_\sigma \alpha$. And so, with obvious notation,

$$g = \pi_\alpha - \sum_{\alpha >_\sigma \beta \in L} b_\beta \pi_\beta$$

Since $g(z) = 0$ for $z \in \mathcal{D}$, then the vector $b = [b_\beta]_\beta$ of the coefficients in the identity above solves the linear system $[\pi_\beta(z)]_{z \in \mathcal{D}, \alpha >_\sigma \beta \in L} b = [\pi_\alpha(z)]_{z \in \mathcal{D}}$. Furthermore, since $[\pi_\beta(z)]_{z \in \mathcal{D}, \alpha >_\sigma \beta \in L}$ is a full rank matrix, then b is the unique solution of such a system.

3. Let $p^* = \sum_{\beta \in L} a_\beta \pi_\beta$ be the polynomial whose coefficients are the solution of the linear system

$$[\pi_\beta(z)]_{z \in \mathcal{D}, \beta \in L} a = [p(z)]_{z \in \mathcal{D}}.$$

Such a polynomial obviously interpolates the values $p(z)$, $z \in \mathcal{D}$, and, since the columns of $[\pi_\beta(z)]_{z \in \mathcal{D}, \beta \in L}$ are the evaluation vectors of the elements of $\mathcal{O}\mathcal{B}$ at \mathcal{D} , it belongs to $\text{Span}(\mathcal{O}\mathcal{B})$. We conclude that p^* is the unique polynomial belonging to $\text{Span}(\mathcal{O}\mathcal{B})$ which interpolates the values $p(z)$, $z \in \mathcal{D}$. \square

Theorem 9:

Proof As G is a basis of $\mathcal{I}(\mathcal{D})$, then for every $p \in \mathcal{I}(\mathcal{D})$ and $g \in G$ there exist $q_g \in \mathbb{R}[x]$ such that $p = \sum_{g \in G} q_g g$. Since by linearity

$$\mathbb{E}_\lambda \left(\sum_{g \in G} q_g g \right) = \sum_{g \in G} \mathbb{E}_\lambda (q_g g)$$

the thesis follows once we show that, for each $g \in G$ and $x^\alpha = LT(g)$

$$\mathbb{E}_\lambda (q_g g) = \|\pi_\alpha\|_\lambda^2 c_\alpha(q_g) - \sum_{\alpha >_\sigma \beta \in L} c_\beta(q_g) c_\beta(g) \|\pi_\beta\|_\lambda^2$$

holds. From Equation (17) we have

$$q_g g = \sum_{\beta} c_\beta(q_g) \pi_\beta g$$

and we substitute the Fourier expansion of g given in Theorem 8

$$g = \pi_\alpha - \sum_{\alpha >_\sigma \beta \in L} c_\beta(g) \pi_\beta$$

In computing the expectation we use the fact that $\mathbb{E}_\lambda (\pi_h \pi_k) = 0$ for different h and k . Then the expectation of $c_\beta(q_g) \pi_\beta g$ vanishes if $\beta >_\sigma \alpha$ or $\beta <_\sigma \alpha$, $\beta \notin L$, the expectation of $c_\alpha(q_g) \pi_\alpha g$ gives $\|\pi_\alpha\|_\lambda^2 c_\alpha(q_g)$ and, if $\alpha >_\sigma \beta \in L$, the expectation of $c_\beta(q_g) \pi_\beta g$ gives $-c_\beta(q_g) c_\beta(g) \|\pi_\beta\|_\lambda^2$. \square

Acknowledgments

G. Pistone is supported by de Castro Statistics Initiative, Collegio Carlo Alberto, Moncalieri Italy. E. Riccomagno worked on this paper while visiting the Department of Statistics, University of Warwick, and the Faculty of Statistics at TU-Dortmund on a DAAD grant. Financial support is gratefully acknowledged. The authors thank Prof. H. P. Wynn, Prof. G. Monegato (Politecnico di Torino) and Prof. Dr. Hans Michael Möller (Technische Universität – Dortmund) for their useful suggestions.

References

- [1] CoCoATeam, CoCoA: a system for doing Computations in Commutative Algebra. Available at <http://cocoa.dima.unige.it>.

- [2] David Cox, John Little, and Donal O’Shea, *Ideals, varieties, and algorithms*, Third, Undergraduate Texts in Mathematics, Springer, New York, 2007.
- [3] Mathias Drton, Bernd Sturmfels, and Seth Sullivant, *Lectures on Algebraic Statistics*, Birkhäuser, Basel, 2009.
- [4] Roberto Fontana, Giovanni Pistone, and Maria Piera Rogantin, *Classification of two-level factorial fractions*, J. Statist. Plann. Inference **87** (2000), no. 1, 149–172.
- [5] Walter Gautschi, *Orthogonal polynomials: computation and approximation*, Numerical Mathematics and Scientific Computation, Oxford University Press, New York, 2004.
- [6] Paolo Gibilisco, Eva Riccomagno, Maria Piera Rogantin, and Henry P. Wynn (eds.), *Algebraic and geometric methods in statistics*, Cambridge University Press, Cambridge, 2010.
- [7] Paul Malliavin, *Integration and probability*, Graduate Texts in Mathematics, vol. 157, Springer-Verlag, New York, 1995.
- [8] H. M. Möller and B. Buchberger, *The construction of multivariate polynomials with preassigned zeros*, Computer algebra (Marseille, 1982), 1982, pp. 24–31.
- [9] Giovanni Peccati and Murad S. Taqqu, *Wiener Chaos: Moments, Cumulants and Diagrams*, Bocconi & Springer-Verlag, Italia, 2011.
- [10] Giovanni Pistone, Eva Riccomagno, and Henry P. Wynn, *Algebraic statistics*, Monographs on Statistics and Applied Probability, Chapman & Hall/CRC, Boca Raton, FL, 2001.
- [11] Wim Schoutens, *Orthogonal polynomials in Stein’s method*, J. Math. Anal. Appl. **253** (2001), no. 2, 515–531.
- [12] A. H. Stroud and Don Secrest, *Gaussian quadrature formulas*, Prentice-Hall Inc., Englewood Cliffs, N.J., 1966.
- [13] Yuan Xu, *Polynomial interpolation in several variables, cubature formulae, and ideals*, Adv. Comput. Math. **12** (2000), no. 4, 363–376.
- [14] H. Michael Möller, *On the construction of cubature formulae with few nodes using Groebner bases*, Numerical integration (Halifax, N.S., 1986), NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci., vol. 203, Reidel, Dordrecht, 1987, pp. 177–192.