THE ALGEBRA OF QUADRATURE FORMULÆ FOR GENERIC NODES

CLAUDIA FASSINO, GIOVANNI PISTONE, AND EVA RICCOMAGNO

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 points $z \in \mathcal{D} \subset \mathbb{R}^d$. The general quadrature problem is: determine classes of functions $f \colon \mathbb{R}^d \to \mathbb{R}$, finite set of n *nodes* $\mathcal{D} \subset \mathbb{R}^d$ and positive weights $\{\lambda_z\}_{z\in\mathcal{D}}$ such that

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

where λ is the probability distribution of the random variable X. In the simplest 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. [4, Sec. 1.4]. Not much is known in the multivariate case, unless the set of nodes is a product of one-dimensional set.

A similar setting appears in statistical Design of Experiment (DoE) where one considers a finite set of *treatments* $\mathcal{D}$ and experimental outputs as funtion of the treatment. The set of treatments and the set of nodes are both described efficiently as zeros of systems of polinomial equations, i.e. as what is called in Commutative Algebra a 0-*dimensional variety*. Such a framework has been used in a systematic way in the literature on Algebraic Statistics using the tools of modern Computational Commutative Algebra, see e.g. [8] and [5]. In such studies 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, the *euclidean* structure, such as the computation of mean values, is missing in the algebraic setting. In algebric 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. [3]. Note that $\{+1,-1\}$ is the zero set of the polynomial $x^2 - 1$.

The purpose of the present paper is to discuss how both worlds can be treated together by considering orthogonal polynomials. In particular, we consider algorithms from the world of Commutative Algebra for the cubature problem in (1) by mixing tools from elementary orthogonal polynomial theory and from probability. Viceversa, the formula (1) provides an interesting interpretation of the equation in the RHS as expected value.

We proceed in sequence by increased degree of difficulty.

In Section 2 we consider the univariate case and take λ to admit an orthogonal system of polynomials. Let $g(x) = \prod_{d\in\mathcal{D}}(x-d)$ and by univariate division given a polynomial p there exist unique q and r such that $p = q\,g + r$ and r has degree smaller than the number of points in $\mathcal{D}$. Furthermore r can be written as $\sum_{d\in\mathcal{D}} r(d)l_d(x)$ where l_d is the Lagrange polynomial for $d \in \mathcal{D}$. 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 polynomials is zero
(2) the weights are the expected values of the Lagrange polynomials.

When λ is a standard Gaussian probability law and $\mathcal{D}$ the zero set of the n-th Hermite polynomial, the application of Stein-Markov theory premits a representation of some Hermite polynomials, including those of degree $2n-1$, as sum of an element in the polynomial ideal generated by the roots of the Hermite polynomial of degree n and of a reminder, suggests a folding of multivariate polynomials over a finite set of points.

The point is to describe a ring structure of the space generated by Hermite polynomials up to a certain order because it is essentially the aliasing on functions induced by limiting observations to $\mathcal{D}$. The particular structure of the recurrence relationship for Hermite polynomials makes this possibile 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 3.5 (extented to the multidimensional case in Section 5) which gives an implicitly 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. The case when the design is a proper subset of the zero set of the n-th Hermite polynomial is developed in Section 4.

Section 6 contains our most general set-up: we restrict to 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 with orthogonal polynomials. It gives a generating set of the vanishing ideal of $\mathcal{D}$ written in terms of orthogonal polynomials. Then it is applied to compute the coefficients of (1) in the Fourier expansion of the interpolatory polynomial at $\mathcal{D}$ of the function whose expectation is wanted. Of course it will is 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) $x = (x_1, \ldots, x_d)$; for $\alpha = (\alpha_1, \ldots, \alpha_d)$ a d-dimensional vector with non-negative integer entries, $x^\alpha = x_1^{\alpha_1} \ldots x_d^{\alpha_d}$ indicates a monomial; $\prec_\tau$ indicates a term-ordering on the monomials in $\mathbb{R}[x]$ i.e., a total well ordering on $\mathbb{Z}^d_{\geq 0}$ such that $x^\alpha \prec_\tau x^\beta$ if x^α divides x^β. If $d = 1$ there is only one term ordering of the monomials, which is definitely not the case for $d \geq 2$. Design of product form share some features with the dimension one. Because of that 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 cases, but not in 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 variable 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 $\prec_\tau$, for $p \in \mathbb{R}[x]$ there exists unique g and $r \in \mathbb{R}[x]$ such that

$$p(x) = g(x) + r(x), \quad g(x) = 0 \text{ if } x \in \mathcal{D}, \tag{2}$$

r is a linear combination of less monomials than points in $\mathcal{D}$, and r is smaller than g with respect to $\prec_\tau$. It is a consequence of the multivariate polynomial division, see e.g. [2, Sec. 2.3]. Furthermore $r(x)$ can be written uniquely as

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

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 set of real valued functions over $\mathcal{D}$.

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

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

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

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

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

The polynomials g and r in (2) are fundamental in the applications of algebraic geometry to finite spaces. In multivariate polynomial division if $f_1, \dots, f_t \in \mathbb{R}[x]$ form a *Gröbner basis* with respect to $\prec_\tau$ (see [2, Cap. 2]) and generate the ideal of polynomial functions vanishing over $\mathcal{D}$, then

$$g(x) = \sum_{i=1}^{t} h_i(x) f_i(x), \quad h_i(x) \in \mathbb{R}[x] \textit{ not } \text{unique}$$

and the unique r has its largest term in $\prec_\tau$ not divisible by the largest term of g_i, $i = 1, \dots, t$. Moreover, monomials not divisible by the largest terms of f_i, $i = 1, \dots, t$, form vector basis of monomial functions for the vector space $\mathcal{L}(\mathcal{D})$ of real functions on $\mathcal{D}$. Various general purpose softwares, including Maple, Mathematica, Matlab and computer algebra softwares, like CoCoA, Macaulay, Singular, allows manipulation with polynomial ideals, in particular can compute the reminder and the monomial basis.

For all $p \in \mathbb{R}[x]$, the polynomial r above is referred to as *reminder* or *normal form.* It is often indicated with the symbol $\mathrm{NF}_\tau(p, \{f_1, \dots, f_t\})$, or the shorter version $\mathrm{NF}(p)$, while $\langle f_1, \dots, f_t\rangle$ indicates the polynomial ideal generated by $f_1, \dots, f_t$.

In one dimension, a Gröbner basis reduces to a polynomial f vanishing over $\mathcal{D}$ and of degree $n = |\mathcal{D}|$ and 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 $g, f \in \langle f\rangle$, and
(3) $r(x) = p(x)$ if x is such that $f(x) = 0$.

In the next Section we are going to study 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 set of zeros of a polynomial which is orthogonal to the constant functions with respect to λ. We recall the basics on orthogonal polynomials we use next, see e.g. [4].

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

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

We consider only λ whose related inner product is definite positive, i.e. $||f|| > 0$ 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 $\pi_0, \pi_1, \dots$. Furthermore we have

$\pi_k \in \mathbb{R}[x]_k$, $\pi_0, \ldots, \pi_k$ form a real vector space basis of $R[x]_{\leq k}$, π_k is orthogonal to all polynomials of total degree smaller than k and 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*

$$\langle xp(x), q(x)\rangle_\lambda = \langle p(x), xq(x)\rangle_\lambda ,$$

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 [9].

Theorem 2.1 (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 $\pi_n(x)$, $n = 0, 1, \ldots$ 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, $\alpha_k = \dfrac{\langle x\pi_k, \pi_k\rangle}{\langle \pi_k, \pi_k\rangle}$ and $\beta_k = \dfrac{\langle \pi_k, \pi_k\rangle}{\langle \pi_{k-1}, \pi_{k-1}\rangle}$ hold true and the norm of π_n is computed from the β's as $||\pi_n||^2 = \beta_n\beta_{n-1}\ldots\beta_0$. For orthonormal polynomials $\tilde{\pi}_k = \pi_k/||\pi_k||$ the *Christoffel-Darboux formulae* hold

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

Non-example 2.2. Inner products of the Sobolev type, namely $\langle u, v\rangle_S = \langle u, v\rangle_{\lambda_0} + \langle u', v'\rangle_{\lambda_1} \cdots + \langle u^{(s)}, v^{(s)}\rangle_{\lambda_s}$ where λ_i are positive measures possibly having different support, do not satisfy the shift condition. Neither do the complex Hermitian inner products.

Theorem 2.3. *Let $\mathcal{D} = \{x \subset \mathbb{R} : \pi_n(x) = 0\}$ be the zero set of the n-th orthogonal polynomial with respect to the probability measure λ. If $p(x) = q(x)\pi_n(x) + r(x)$, then*

$$\mathbb{E}\left(p(X)\right) = \sum_{z\in\mathcal{D}} p(z)\lambda_z, \qquad \textit{if and only if } c_n(q) = 0.$$

Remark 2.4. *This theorem is a version of a well known result, see e.g.* [4, Sec. 1.4]. *We include the proof to underline a particular form of the error in the quadrature formula, to be used again in the next Theorem and in Section 6.*

Proof. The zero set of π_n, $\mathcal{D} = \{x : \pi_n(x) = 0\}$, 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 $l_z(x) = \prod_{w\in\mathcal{D}:w\neq z} \dfrac{x-w}{z-w}$ for $z \in \mathcal{D}$. Hence we have

$$\begin{aligned} \mathbb{E}\left(p(X)\right) &= \mathbb{E}\left(q(X)\pi_n(X)\right) + \sum_{z\in\mathcal{D}} p(z)\,\mathbb{E}\left(l_z(X)\right) \\ &= c_n(q)||\pi_n||^2_\lambda + \sum_{z\in\mathcal{D}} p(z)\lambda_z. \end{aligned}$$

□

A particular case of Theorem 2.3 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 [4, Chapter 1].

Example 2.5 (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 exactly.

In general for a polynomial of degree N possibly larger than $2n-1$, Theorem 2.3 gives the Fourier expansion of its reminder by π_n, indeed

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

Theorem 2.6 below generalises Theorem 2.3 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 $l_z(x)=\prod_{w\in\mathcal{D}:w\neq z}\dfrac{x-w}{z-w}$. 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 2.6. *With the above notation,* $\mathbb{E}\left(p(X)\right)=\sum_{z\in\mathcal{D}} p(z)\lambda_z$ *if and only if* $\sum_{k=0}^{+\infty} c_k(q)c_k(g)||\pi_k||_\lambda^2=0$.

Proof. From

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

we have

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

and this proves the theorem. □

The condition we find in Theorem 2.6 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

$$M\left[c_k(q)\right]_{k=0,\ldots,|\mathcal{D}|-1}=\left[q(z)\right]_{k=0,\ldots,|\mathcal{D}|-1} \tag{5}$$

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

Theorem 2.6 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 $\mathbb{E}\left(\prod_{w\in\mathcal{D}:w\neq z}\dfrac{X-w}{z-w}\right)$ for $z\in\mathcal{D}$. The Fourier coefficients of g can be computed analogously to those of q adapting Equation (5). If p is unknown and $p(x)=\sum_\alpha p_\alpha x^\alpha$ for a finite number of non-zero, unknown real numbers p_α, Theorem 2.6 characterizes all the polynomials for which the Gaussian quadrature rule is exact, namely $\mathbb{E}\left(p(X)\right)=\sum_{z\in\mathcal{D}}p(z)\lambda_z$. Furthermore, the characterization is a linear expression in the unknown p_α, where $p(x)=\sum_\alpha p_\alpha x^\alpha$. 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 6.1 we shall generalise Theorem 2.6 to higher dimension. To conclude this section, we observe that the remainder r admits an interpretation as a projection.

Proposition 2.7. *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 q is the unique polynomial such that $p-q\pi_n$ is orthogonal to all π_m with $m\geq n$.*

Proof. As $r=p-q\pi_n$ has degree at most $n-1$, it can be written as a linear combination of $\pi_0,\pi_1,\dots,\pi_{n-1}$ and is orthogonal to π_m for all $m\geq n$. Let q_1 and $q_2\in\mathbb{R}[x]$ such that both $p-q_1\pi_n$ and $p-q_2\pi_n$ are orthogonal to π_m for all $m\geq n$. Now $(q_1-q_2)\pi_n$ has degree not smaller than n, hence it is orthogonal to $\pi_0,\dots,\pi_{n-1}$. Necessarily it is $q_1-q_2=0$, because it is orthogonal to all π_k, $k=0,\dots$. □

Example 2.8. Substituting the Fourier expansions of q and p in the orthogonality relation, we find that the m-th coefficient in the Fourier expansion of p can be written as

$$\begin{aligned}\mathbb{E}\left(p(X)\pi_m(X)\right)&=\mathbb{E}\left(q(X)\pi_n(X)\pi_m(X)\right)\\ \sum_k c_k(p)\,\mathbb{E}\left(\pi_k(X)\pi_m(X)\right)&=\sum_j c_j(q)\,\mathbb{E}\left(\pi_j(X)\pi_n(X)\pi_m(X)\right)\\ c_m(p)||\pi_m||^2&=\sum_j c_j(q)\,\mathbb{E}\left(\pi_j(X)\pi_n(X)\pi_m(X)\right)\end{aligned}$$

For Hermite polynomials it can be simplified by e.g. using the product formula in Theorem 3.1 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, $d\lambda(x)=w(x)\,dx$, with $w(x)=\dfrac{e^{-x^2/2}}{\sqrt{2\pi}}$, $x\in\mathbb{R}$.

3.1. Stein-Markov operators for standard normal distribution.

For a real valued, differentiable function f, define

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

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

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

if ϕ,ψ are such that $\lim_{x\to\pm\infty}\phi(x)\psi(x)e^{-x^2/2}=0$ and are square integrable, see [6, 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.

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) $d\delta - \delta d = id$ from which the relationships $dH_n = nH_{n-1}$, $d^m H_n = \dfrac{n!}{m!} H_{n-m}$ for $m \leq n$ and the three-term recurrence relationship

$$H_{n+1} = xH_n - nH_{n-1} \tag{7}$$

are deduced.

(4) Hermite polynomials are orthogonal with respect to $d\lambda(x) = \dfrac{e^{-x^2/2}}{\sqrt{2\pi}} dx$. 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 $n = m$.

Already we 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 3.1.

Theorem 3.1. *The Fourier expansion of the product of $H_k H_n$ 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

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

□

Example 3.2 (Aliasing)**.** As an application of Theorem 3.1, 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 3.1 gives the formula to write h_j^{n+k} in terms of Fourier coefficients of lower order Hermite polynomials:

$$\mathrm{NF}(H_{n+k}) \equiv -\sum_{i=1}^{n\wedge k} \binom{n}{i}\binom{k}{i} i!\, \mathrm{NF}(H_{n+k-2i}) \equiv -\sum_{i=1}^{n\wedge k} \binom{n}{i}\binom{k}{i} i! \sum_{j=0}^{n-1} h_j^{n+k-2i} H_j$$

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}, $k = 1, \ldots, 6$ over $\mathcal{D} = \{H_n(x) = 0\}$

Example 3.3. Observe that if f has degree $n + 1$ equivalently $k = 1$ then

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

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

We give another proof of Theorem 2.3 for Hermite polynomials.

Corollary 3.4. *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*

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

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

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 2.3 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 3.5 gives two polynomial equations whose zeros are the design points and the weights. The proof is based on the *Christoffel-Darboux formulae*, Equations (4).

Theorem 3.5. *Let* $\mathcal{D}_n = \{x : H_n(x) = 0\}$.

(1) *There exists only one polynomial* λ *of degree* $n-1$ *such that* $\lambda(x_k) = \lambda_k$ *for all* $k = 1, \dots, n$,
(2) *furthermore* $\lambda_k = \frac{(n-1)!}{n} H_{n-1}^{-2}(x_k)$. *Equivalently*
(3) *the polynomial* λ *satisfies*

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

Proof. The univariate polynomial λ is the interpolatory polynomial of the values λ_k's at the n distinct points in $\mathcal{D}_n$ and hence it exists, unique of degree $n-1$. To prove item 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 formulae and evaluation at $\mathcal{D}_n = \{x_1, \dots, x_n\}$ give

$$\sum_{k=0}^{n-1} \tilde{H}_k(x_i)\tilde{H}_k(x_j) = 0 \text{ if } i \neq j \qquad \sum_{k=0}^{n-1} \tilde{H}_k(x_i)^2 = n\tilde{H}_{n-1}(x_i)^2 \tag{8}$$

In matrix form Equations (8) becomes

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

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

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

Now, let f be a polynomial of degree at most $n-1$, 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

$$\begin{aligned} \underline{c} &= \mathbb{H}_n^{-1}\underline{f} = \mathbb{H}_n^t n^{-1} \operatorname{diag}(\tilde{H}_{n-1}^{-2}(x_i) : i = 1, \dots, n)\underline{f} \\ &= \mathbb{H}_n^t n^{-1} \operatorname{diag}(\tilde{H}_{n-1}^{-2}(x_i) f(x_i) : i = 1, \dots, n) \end{aligned}$$

$$c_j = \frac{1}{n} \sum_{i=1}^{n} \tilde{H}_j(x_i) f(x_i) \tilde{H}_{n-1}^{-2}(x_i) \tag{9}$$

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 (9), obtain

$$c_{kj} = \frac{1}{n} \tilde{H}_j(x_k) \tilde{H}_{n-1}^{-2}(x_k) \tag{10}$$

The expected value of $l_k(Z)$ is

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

Substitution in Equation (10) 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 $k = 1, \dots, n$.

Item 3. 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. □

Item 2 in the theorem states that the weights are strictly positive. Furthermore, Theorem 2.3 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. Then Theorem 2.3 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 $\mathrm{P_n}(X = x_k) = \mathbb{E}(l_k(Z)) = \lambda_k$, $k = 1, \dots, n$

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

Example 3.6. For $n = 3$ the polynomial λ in Theorem 3.5 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 $x^3 = 3x$

$$2/3 = \lambda(x)H_2^2 = \theta_0 + \theta_1 4x + (\theta_0 + 4\theta_2)x^2 \tag{11}$$

Coefficients in Equation (11) 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 $H_3(x) = 0$ 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 $\lambda_0 = \lambda(0) = \frac{2}{3}$.

3.3. A code for the weighing polynomial.

The polynomial $\lambda(x)$ in Theorem 3.5 is called the *weighing polynomial*. Table 3.3 gives a code written in the specialised software for symbolic computation called CoCoA [1]. to compute the Fourier expansion of $\lambda(x)$ exploiting Theorem 3.5.

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, `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 than $H_N = H_1 H_{N-1} - (N-1)H_{N-2} = 0$, 'giving' the nodes of the quadrature. Line 7 is the polynomial in the second equation in the system in Item 3 of Theorem 3.5 and 'gives' the weights. In total there are N equations which are collected in an algebraic structure called an ideal whose Gröbner bases [2] 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. Explicitly w in such polynomial provides the weighing polynomial written in terms of the first $N-1$ Hermite polynomials.

Line 9 in Table 3.3 gives the polynomial obtained for $N = 4$, namely $\lambda(x) = \frac{5-h2}{12} = \frac{6-x^2}{12}$. 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

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

Table 2. Computation of the Fourier expansion of the weighing polynomial using Theorem 3.5

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 and gives a weighing polynomial of 22.349 characters. Observe that this computations can be done once for all and the results stored.

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 $\mathcal{D}_n = \{x : \pi_n(x) = 0\}$ with a number of points m, $0 < m < n$. 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, 8]. 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 2.3 we have

$$\mathbb{E}((p1_\mathcal{F})(X)) = \sum_{z\in\mathcal{F}} p(z)\lambda_z = \mathrm{E_n}\left(p(Y)1_\mathcal{F}(Y)\right) = \mathrm{E_n}\left(p(Y)|Y \in \mathcal{F}\right)\mathrm{P_n}(Y \in \mathcal{F})$$

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 $\mathrm{P_n}(Y = z) = \lambda_z$. The first equality follows from the fact that $f(x)1_\mathcal{F}(x)$ 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 4.1. *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*

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

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

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

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

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

$$\begin{aligned}\mathbb{E}\left(p(X)\right) &= \mathbb{E}\left(\sum_{j=0}^{N-m} b_j \pi_j(X) \sum_{i=0}^{m} c_i \pi_i(X)\right) + \mathbb{E}\left(r(X)\right)\\ &= b_0 c_0 ||\pi_0||_\lambda^2 + b_1 c_1 ||\pi_1||_\lambda^2 + \ldots + b_{(N-m)\wedge m} c_{(N-m)\wedge m} ||\pi_{(N-m)\wedge m}||_\lambda^2 + \sum_{z\in\mathcal{F}} p(z) \lambda_z^{\mathcal{F}}\end{aligned}$$

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

Note that the error of the Gaussian quadrature rule, $b_0 c_0 ||\pi_0||_\lambda^2 + b_1 c_1 ||\pi_1||_\lambda^2 + \ldots + b_{(N-m)\wedge m} c_{(N-m)\wedge m} ||\pi_{(N-m)\wedge m}||_\lambda^2$, is linear in the Fourier coefficients b_j, and also in the Fourier coefficients c_j relative to the node polynomial. This is generalised in Section 6.1. 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 $X_1, \ldots, X_d$ 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

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

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

The Lagrange polynomial of the point $y = (y_1, \ldots, y_d) \in \mathcal{D}_{n_1,\ldots,n_d}$ is defined as $l_y(x_1, \ldots, x_d) = \prod_{k=1}^d l_{y_k}^{n_k}(x_k)$, 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 $(l_y : y \in \mathcal{D}_{n_1,\ldots,n_d})$, is equal to the linear space generated by the monomials whose exponents lie on the integer grid $\{0, \ldots, n_1 - 1\} \times \ldots \times \{0, \ldots, n_d - 1\}$. Any polynomial $f \in \mathbb{R}[x]$ can be written as

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

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

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

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

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

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

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

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

Theorem 5.1. *It holds*

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

Proof. The proof is very similar to that of Theorem 2.3 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

$$f(x, y) = q_1(x, y)\pi_n(x) + q_2(x, y)\pi_n(y) + \sum_{(a,b)\in\mathcal{D}_{n,m}} f(a, b)\ l_a^n(x)l_b^m(y) =$$

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

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

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

□

Note in the proof above that a sufficient condition for $\mathbb{E}_\lambda\left(c_n(q_1)(Y)\right)$ being zero is that f has degree in x smaller then $2n - 1$, similarly for $\mathbb{E}_\lambda\left(c_m(q_2)(X)\right)$. 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

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

In the Gaussian set-up, by Theorem 3.5 applied to each variable, weights and nodes satisfy the polynomial system

$$\begin{cases} H_{n_1}(x_1) & = 0 \\ \lambda_1(x_1)H_{n_1-1}(x_1)^2 & = \dfrac{(n_1-1)!}{n_1} \\ & \vdots \\ H_{n_d}(x_d) & = 0 \\ \lambda_d(x_d)H_{n_d-1}(x_d)^2 & = \dfrac{(n_d-1)!}{n_d} \end{cases}$$

For the grid set-up of this section and for the Gaussian case, in analogy to Example 3.2 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 5.2 below.

Example 5.2. Consider a square grid of size n, $\mathcal{D}_{nn}$, 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

$$\begin{aligned}\mathbb{E}\left(f(Z_1,Z_2)H_k(X_1)H_h(X_2)\right) &= c_{hk}||H_k(X_1)||^2||H_h(X_2)||^2 \\ c_{kh} &= \frac{1}{k!h!}\sum_{(x,y)\in\mathcal{D}_{nn}} f(x,y)H_k(x)H_h(y)\lambda_x\lambda_y\end{aligned}$$

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

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

Example 5.3 deals with a general design and introduces the more general theory of Section 6.

Example 5.3. Let $\mathcal{F}$ be the zero set of

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

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 belongs to $\operatorname{Span}(H_0, H_1(x), H_1(y), H_1(x)H_1(y), H_2(y)) = \operatorname{Span}\left(1, x, y, xy, y^2\right)$. If, furthermore, f is such that

$$\begin{aligned} q_1(x,y) &= a_0 + a_1 H_1(x) + a_2 H_1(y) + a_3 H_1(x)H_1(y) \\ q_2 &= \theta_1(x) + \theta_2(x)H_1(y) + \theta_3(x)H_2(y) \\ q_3 &= a_4 + a_5 H_1(y) \end{aligned}$$

with $a_i, \theta_j \in \mathbb{R}$ for $i = 0,\dots,5$ and $j = 1,\dots,3$, then $\mathbb{E}(g_i(Z_1,Z_2)q_i(Z_1,Z_2)) = 0$ for $i = 1,2,3$ and 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 $\operatorname{Span}(H_0, H_1(x), H_1(y), H_1(x)H_1(y), H_2(y))$ and they are

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

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 5.3 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 3.2). Also we deduced a system of polynomial equations whose solution consists the weights of a quadrature formula. The mathematical tool that allowed this is Equation (6) and the particular structure it implies for Hermite polynomials on the recurrence relation for general, orthogonal polynomials

$$\pi_{k+1}(x) = (\gamma_k x - \alpha_k)\pi_k(x) - \beta_k \pi_{k-1}(x) \qquad x \in \mathbb{R} \tag{12}$$

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

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, and a generic set of orthogonal polynomials. We gain something and loose something. The essential computations are linear: such is the computation of a Gröbner basis for a finite set of distinct points [7]; the Buchberger Möller type of algorithm in Table 3 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 16)

Given a set of points and a term-ordering the algorithm in Table 3 returns the reduced Gröbner basis of the design ideal expressed as linear combination of orthogonal polynomial of low enough order. It does so directly; that is it computes the Gröbner basis by working only in the space of orthogonal polynomials.

We loose the equivalent of Theorem 3.1 for Hermite polynomials, in particular we do not know yet how to impose a ring structure on $\text{Span}(\pi_0,\ldots,\pi_n)$ 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 $i = 0,\ldots,n\wedge k, n+k$, which is fundamental for the aliasing structure discussed for Hermite polynomials.

For multivariate cubature formulae we refer e.g. to [10] which, together with [7], are basic references for this section. We are writing up in another manuscript our results on the degree of the cubature formula we obtain. 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 $\alpha = (\alpha_1,\ldots,\alpha_d) \in \mathbb{Z}^d_{\geq 0}$ associate the monomial $x_1^{\alpha_1}\cdots x_d^{\alpha_d}$, in short x^α, and the product $\pi_{\alpha_1}(x_1)\ldots\pi_{\alpha_d}(x_d)$, in short $\pi_\alpha(x)$. Note that $\{\pi_\alpha\}_\alpha$ are a system of orthogonal polynomials for the product measure λ^d. Theorem 6.1 describes the one-to-one correspondence between the x^αs and the $\pi_\alpha(x)$s.

Theorem 6.1. (1) *For $d = 1$ and $k \in \mathbb{Z}_{\geq 0}$, in the notation of Equation (12) 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 $k = 1, 2, \ldots,$

$$\begin{aligned} c_{-1}(x^k) &= c_{k+1}(x^k) = 0 \\ c_j(x^k) &= \frac{c_{j-1}(x^{k-1})}{\gamma_{j-1}} + \frac{c_j(x^{k-1})\alpha_j}{\gamma_j} + \frac{c_{j+1}(x^{k-1})\beta_{j+1}}{\gamma_{j+1}} \qquad j = 0, \ldots, k-1 \\ c_k(x^k) &= \frac{1}{\gamma_0 \cdots \gamma_{k-1}} \end{aligned}$$

(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*

$$\pi_\alpha = \sum_{\beta \leq \alpha} a_\beta x^\beta \qquad \text{and} \qquad x^\alpha = \sum_{\beta \leq \alpha} b_\beta \pi_\beta \tag{13}$$

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. See Appendix 7. ☐

Example 6.2. For π_j the j-th Hermite polynomial H_j, Item 1 of Theorem 6.1 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 6.1 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

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

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.

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 $LT(I(\mathcal{D})) = \langle LT_\sigma(f) : f \in I(\mathcal{D})\rangle$. If G is the σ-reduced Gröbner basis of $I(\mathcal{D})$, then $LT(I(\mathcal{D})) = \langle LT_\sigma(f) : f \in G\rangle$. Now we can define two interesting vector space bases of $\mathcal{L}(\mathcal{D})$. Let $L = \{\alpha \in \mathbb{Z}^d_{\geq 0} : x^\alpha \notin LT(I(\mathcal{D}))\}$, then we define

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

The sets L, $\mathcal{B}$ and $\mathcal{OB}$ 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 $\mathcal{OB}$. For example for $d = 2$, let $L = \{(0,0), (1,0), (0,1), (2,0)\}$, then $\mathcal{B} = \{1, x, y, x^2\}$ and $\mathcal{OB} = \{\pi_0(x)\pi_0(y), \pi_1(x)\pi_0(y), \pi_0(x)\pi_1(y), \pi_2(x)\pi_0(y)\} = \{1, \pi_1(x), \pi_1(y), \pi_2(x)\}$. 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, with abuse of notation, $\alpha <_\sigma \beta$ if and only if $x^\alpha <_\sigma x^\beta$ for each $\alpha, \beta \in \mathbb{Z}^d_{\geq 0}$. Further, given $\alpha \leq \beta$ componentwise, since x^α divides x^β and since $x^{\beta-\alpha} \geq_\sigma 1$, we have $x^\alpha \leq_\sigma x^\beta$, that is $\alpha \leq_\sigma \beta$.

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 $g = x^\alpha + \sum_{\beta \in L, \beta <_\alpha \alpha} a_\beta x^\beta$ with $a_\beta \in \mathbb{R}$.

Theorem 6.3 provides an alternative to the classical method of rewriting a polynomial f in terms of orthogonal polynomials by substituting each monomial in f by applying Theorem 6.1. It gives linear rules to rewrite 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 7.

Theorem 6.3.

(1) $\mathrm{Span}(\mathcal{B}) = \mathrm{Span}(\mathcal{OB})$;

(2) *Let G be the reduced σ-Gröbner basis of $I(\mathcal{D})$. The polynomial $g \in G$ 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*
$$p^* = \sum_{\beta \in L} a_\beta \pi_\beta \tag{14}$$
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 such that $p^*(z) = p(z)$ for all $d \in \mathcal{D}$ and $p^* \in \mathrm{Span}(\mathcal{OB})$.

Theorem 6.3 provides a pseudo-algorithm to compute a Gröbner basis for $I(\mathcal{D})$ and interpolating polynomials at $\mathcal{D}$ in terms of orthogonal polynomials of low order directly from $\mathcal{D}$ and $\mathcal{OB}$. In Table 3 we give a variation of the Buchberger-Möller algorithm [7] which starting from a finite set of distinct points $\mathcal{D}$ and a term-ordering σ 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 $I(\mathcal{D})$. 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 6.3. This permits to rewrite every polynomial $p \in \mathbb{R}[x]$ as a linear combination of 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 3 and
(2) determining the unique polynomial p^* such that $p^*(z) = f(z)$ for all $z \in \mathcal{D}$, by solving the linear system $[\pi_\beta(z)]_{z \in \mathcal{D}, \beta \in L} a = [f(z)]_{d \in \mathcal{D}}$. The polynomial p^* is expressed as linear combination of orthogonal polynomials.
(3) The coefficient a_0 of π_0 is the wanted approximation.

Recall that $p^*(x) = \sum_{z \in \mathcal{D}} f(z) l_d(x)$ is a linear combination of the indicator functions of the points in $\mathcal{D}$ (Lagrange polynomials) and hence $a_0 = \sum_{z \in \mathcal{D}} f(z)\, \mathbb{E}\left(l_d(X)\right)$. In particular, $\mathbb{E}_{\lambda^d}\left(l_z(X)\right) = \lambda_z$, $z \in \mathcal{D}$, can be computed by applying the above

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 $I(\mathcal{D})$ as linear combination of orthogonal polynomials and the set L.

Step 1: Let $L = \{0 \in Z^d_{\geq 0}\}$, $\mathcal{OB} = [1]$, $G = [\,]$ and $M = [x_1, \ldots, x_d]$.
Step 2: If $M = [\,]$ stop; else set $x^\alpha = \min_\sigma(M)$ and deleted 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 $\{x_1x^\alpha, \ldots, x_dx^\alpha\}$ 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 3. Buchberger-Möller algorithm using Orthogonal Polynomials

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 5.1. It would be interesting to generalise this section to non-product measures.

6.1. **Characterisation of polynomial functions with zero expectation.** In this section we characterise the set of polynomials with the same expected value.

As mentioned in Section 1.1 given $\mathcal{D}$, its vanishing ideal $I(\mathcal{D})$, a term-ordering σ and a Gröbner basis G of $I(\mathcal{D})$ with respect to σ, then any polynomial $p \in \mathbb{R}[x]$ can be written as

$$p(x) = \sum_{g\in G} q_g(x)g(x) + r(x)$$

where $r(x)$ is unique in Span($\mathcal{B}$) such that $r(z) = p(z)$ for all $z \in \mathcal{D}$ and can be written as $r(x) = \sum_{z\in\mathcal{D}} p(z)l_d(x)$, where l_z, $z \in \mathcal{D}$, are the product Lagrange polynomials in Section 5. Theorem 6.3 says how to write r over $\mathcal{OB}$.

If $p \in \mathbb{R}[x]$ is such that $\mathbb{E}\,(p(X)) = \mathbb{E}\,(r(X))$ then $\mathbb{E}\,(p(X) - r(X)) = 0$. Furthermore we have $p - r \in I(\mathcal{D})$. Hence instead of studying directly the set

$$\{p \in \mathbb{R}[x] : \mathbb{E}\,(p(X)) = \mathbb{E}\,(r(X))\}$$

we characterize the set

$$\mathcal{E}_0 = \{g \in I(\mathcal{D}) : \mathbb{E}\,(g(X)) = 0\}$$

in Theorem 6.4 whose proof can be found in Appendix 7. Hence if $p \in \mathbb{R}[x]$ is such that $p = g + r$ with $g \in \mathcal{E}_0$ and $r \in$ Span($\mathcal{B}$) then by linearity $\mathbb{E}(p) = \sum_{(z_1,\ldots,z_n)\in\mathcal{D}} p(z_1, \ldots, z_n)\lambda^{n_1}_{z_1} \cdots \lambda^{n_d}_{z_d}$.

Theorem 6.4. *Let λ^d be a product probability measure with product orthogonal polynomials $\pi_\alpha(x)$, $\alpha \in \mathbb{Z}^d_{\geq 0}$ and let X be a random vector following λ^d. Let $\mathcal{D} \subset \mathbb{R}^d$ be a set of distinct points, σ a term-ordering, G be the σ-reduced Gröbner basis of $I(\mathcal{D})$ written as linear combination of orthogonal polynomials, that is for*

$g \in G$ and $x^{\alpha_g} = LT(g)$ write

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

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

$$q_g = \sum_{\beta >_\sigma \alpha_g, \alpha_g >_\sigma \beta \notin L} c_\beta(q_g)\pi_\beta + c_0(q_g)\pi_{\alpha_g} + \sum_{\alpha_g >_\sigma \beta \in L} c_\beta(q_g)\pi_\beta \tag{15}$$

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

$$\sum_{g \in G} \|\pi_{\alpha_g}\|^2_\lambda c_0(q_g) - \sum_{g \in G} \sum_{\alpha_g >_\sigma \beta \in L} \|\pi_\beta\|^2_\lambda c_\beta(q_g) c_\beta(g) = 0 \tag{16}$$

Note that any polynomial can be written according to Equation (15) where the first sum includes terms higher in σ than $LT(g)$ and terms that do not appear in g, the last sum includes terms lower than $LT(g)$ in σ. Hence, the key observation in the proof is that $\mathbb{E}\left(\pi_m \; \pi_n\right) = 0$ is $n \neq m$ and then linearity of $\mathbb{E}$ is used. Importantly, only terms of low enough Fourier order (the second and third terms in Equation (15)) matter for the computation of the expectation.

Example 6.5. 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 $\sigma =$`DegLex` term-ordering over the monomials in $\mathbb{R}[x,y]$. The algorithm in Table 3 gives $\mathcal{OB} = \{1, H_1(y), H_1(x), H_1(x)H_1(y), H_2(x)\}$ and $G = \{g_1, g_2, g_3\}$ where

$$\begin{aligned}
g_1 &= H_2(y) - H_1(x) + 2H_1(y) - 4 \\
g_2 &= H_2(x)H_1(y) - 9H_2(x) + 47H_1(x)H_1(y) - 123H_1(x) + 271H_1(y) - 399 \\
g_3 &= H_3(x) - 47H_2(x) + 300H_1(x)H_1(y) - 845H_1(x) + 2040H_1(y) - 2987
\end{aligned}$$

By Theorem 6.4 for the purpose of computing its expectation a polynomial $p = q_1g_1 + q_2g_2 + q_3g_3 \in I(\mathcal{D})$ can be simplified to have the form

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

and furthermore by Equation (16)

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

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. For example the above equation is satisfied by

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

by adding $H_4(x)$ to q_1 and $H_4(y)$ to q_2 which are not influent in the computation of the expectation we get the following zero mean polynomial

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

7. Appendix: proof

Theorem 6.1:

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 recurrence formula we have

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

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, $j = 1, \ldots, d$. 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 $\beta = (\beta_1, \ldots, \beta_d)$ is such that $\beta \leq \alpha$ component wise, that is x^β divides x^α. □

Theorem 6.3:

Proof. Recall that $\mathcal{B}$ and $\mathcal{OB}$ 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 $(\alpha_1, \ldots, \alpha_d) \in L$ then $(\beta_1, \ldots, \beta_d) \in L$ if and only if $\beta_i \leq \alpha_i$ for all $i = 1, \ldots, d$.

(1) If $x^\alpha \in \mathcal{B}$ for some α, then by Theorem 6.1 $x^\alpha = \sum_{\beta \leq \alpha} b_\beta \pi_\beta$ Since $\beta \leq \alpha$ then $\beta \in L$ and so each $\pi_\beta \in \mathcal{OB}$: we have that x^α belongs to $\mathrm{Span}(\mathcal{OB})$. 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 such a matrix follows from the fact that each linear combination of its columns corresponds to a polynomial in $\mathrm{Span}(\mathcal{OB})$ which coincides with $\mathrm{Span}(\mathcal{B})$ whose elements do not vanish at $\mathcal{D}$.

Any polynomial $g \in G$ in the Göbner basis can be written as

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

where $x^\alpha = LT(g)$ is a multiple of an element of $\mathcal{B}$. By Theorem 6.1 we have

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

The polynomial π_α appears only in the first sum, for the other terms in the first sum observe that as $\beta < \alpha$ then $\beta \in L$ and also $\beta <_\sigma \alpha$. Analogously, for the second sum we 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,

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

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 $[\pi_\beta(z)]_{z \in \mathcal{D}, \alpha_g >_\sigma \beta \in L} b = [\pi_\alpha(z)]_{z \in \mathcal{D}}$. Furthermore, since $[\pi_\beta(z)]_{z \in \mathcal{D}, \alpha_g >_\sigma \beta \in L}$ 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 a polynomial whose coefficients are the elements of 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 $\mathcal{OB}$ at $\mathcal{D}$, it belongs to $\mathrm{Span}(\mathcal{OB})$. We conclude that p^* is the unique polynomial interpolating the values $p(d)$, $d \in \mathcal{D}$, w.r.t. $(\mathcal{D}, \mathrm{Span}(\mathcal{OB}))$.

□

Theorem 6.4:

Proof. As G is a Gröbner basis of $I(\mathcal{D})$, then for every $p \in I(\mathcal{D})$ 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

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

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

$$\mathbb{E}\left(q_g g\right) = \|\pi_{\alpha_g}\|^2_\lambda c_0(q_g) - \sum_{\alpha_g >_\sigma \beta \in L} c_\beta(q_g) c_\beta(g) \|\pi_\beta\|^2_\lambda$$

holds.

From Equation (15) we have

$$q_g g = \sum_{\beta >_\sigma \alpha_g, \alpha_g >_\sigma \beta \notin L} c_\beta(q_g)\pi_\beta g + c_0(q_g)\pi_{\alpha_g} g + \sum_{\alpha_g >_\sigma \beta \in L} c_\beta(q_g)\pi_\beta g$$

and substitute the Fourier expansion of g given in Theorem 6.3

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

In computing the expectation we use the fact that $\mathbb{E}\left(\pi_h \pi_k\right) = 0$ for different monomials h and k. Then the expectation of the first sum vanishes, the expectation of the middle term gives $\|\pi_{\alpha_g}\|^2_\lambda c_0(q_g)$ and the last sum gives $-\sum_{\alpha_g >_\sigma \beta \in L} c_\beta(q_g)c_\beta(g)\|\pi_\beta\|^2_\lambda$. $\square$

Acknowledgments. G. Pistone is supported by de Castro Statistics Initiative, Collegio Carlo Alberto, Moncalieri Italy. He thanks G. Monegato (Politecnico di Torino) for advise on orthogonal polynomials. 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.

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. An introduction to computational algebraic geometry and commutative algebra. MR2290010 (2007h:13036)

[3] Roberto Fontana, Giovanni Pistone, and Maria Piera Rogantin, *Classification of two-level factorial fractions*, J. Statist. Plann. Inference **87** (2000), no. 1, 149–172. MR1772046 (2001k:62094)

[4] Walter Gautschi, *Orthogonal polynomials: computation and approximation*, Numerical Mathematics and Scientific Computation, Oxford University Press, New York, 2004. Oxford Science Publications. MR2061539 (2005e:42001)

[5] Paolo Gibilisco, Eva Riccomagno, Maria Piera Rogantin, and Henry P. Wynn (eds.), *Algebraic and geometric methods in statistics*, Cambridge University Press, Cambridge, 2010. MR2640515 (2011a:62007)

[6] Paul Malliavin, *Integration and probability*, Graduate Texts in Mathematics, vol. 157, Springer-Verlag, New York, 1995. With the collaboration of Hélène Airault, Leslie Kay and Gérard Letac, Edited and translated from the French by Kay, With a foreword by Mark Pinsky. MR1335234 (97f:28001a)

[7] H. M. Möller and B. Buchberger, *The construction of multivariate polynomials with preassigned zeros*, Computer algebra (Marseille, 1982), 1982, pp. 24–31. MR680050 (84b:12003)

[8] Giovanni Pistone, Eva Riccomagno, and Henry P. Wynn, *Algebraic statistics*, Monographs on Statistics and Applied Probability, vol. 89, Chapman & Hall/CRC, Boca Raton, FL, 2001. Computational commutative algebra in statistics. MR2332740 (2008f:62098)

[9] Wim Schoutens, *Orthogonal polynomials in Stein's method*, J. Math. Anal. Appl. **253** (2001), no. 2, 515–531. MR1808151 (2001k:62021)

[10] Yuan Xu, *Polynomial interpolation in several variables, cubature formulae, and ideals*, Adv. Comput. Math. **12** (2000), no. 4, 363–376. Multivariate polynomial interpolation. MR1768956 (2001f:41008)

DIMA, Università di Genova, Italy. fassino@dima.unige.it

Collegio Carlo Alberto, Moncalieri, Italy. giovanni.pistone@carloalberto.org

DIMA, Università di Genova, Italy. riccomagno@dima.unige.it