



# Performance of cubature formulae in probabilistic model analysis and optimization



Fernando P. Bernardo\*

GEPSI-PSE Group, CIEPQPF, Department of Chemical Engineering, University of Coimbra, Portugal

## ARTICLE INFO

### Article history:

Received 20 May 2014

Received in revised form 25 November 2014

### Keywords:

Multidimensional integration

Cubature formulae

Uncertainty analysis

## ABSTRACT

In probabilistic model analysis and optimization, expected values of a model output  $f(x)$  in face of continuous random inputs  $x$  are estimated through  $n$ -dimensional integrals, where  $n = \dim(x)$ . Cubature formulae are approximations of these integrals by a weighted sum of function evaluations at carefully chosen points. When each function evaluation corresponds to a heavy computational simulation, and particularly in optimization problems, one needs very efficient formulae with few integration points, even though only having modest accuracy. In this paper, we evaluate the performance of several cubature formulae with few points, including Smolyak type formulae, also known as sparse grid integration, and recently proposed thinned cubatures, constructed using orthogonal arrays. Tests are made for a wide family of smooth and non-oscillatory functions  $f(x)$ , possibly with significant anisotropy, and covering both normal and uniform input probability distributions. Two practical case studies are also presented, one of analysis of a large scale mass transfer model with uncertain parameters and a second one of optimal production planning under uncertain market conditions. Results clearly indicate that cubatures with large negative weights, including Smolyak type formulae, are not reliable, contrary to positive thinned cubatures that produce very reasonable estimates up to dimension 24. These thinned cubatures may also surpass quasi-Monte Carlo methods also up to dimension 24.

© 2014 Elsevier B.V. All rights reserved.

## 1. Introduction

Large-scale models, used for design, diagnosis and/or optimization of engineering systems, including physical products, manufacturing processes and production/distribution systems [1–4], often include several uncertain parameters, such as physico-chemical properties, operating conditions and economic/market parameters. If these uncertainties (vector  $x \in X \subset \mathbb{R}^n$ ) are described through a continuous probability density function  $g(x)$ , the mean (or expected value) and variance of a model output  $f(x)$ , are given, respectively, by the  $n$ -dimensional integrals:

$$E[f] = \int_X f(x) g(x) dx \quad (1)$$

$$V[f] = \int_X (f(x) - E[f])^2 g(x) dx. \quad (2)$$

\* Correspondence to: Pólo II - Rua Sílvio Lima, 3030-790 Coimbra, Portugal.

E-mail address: [bernardo@eq.uc.pt](mailto:bernardo@eq.uc.pt).

The integrand  $f(x)$  is frequently the result of a heavy computational simulation and hence efficient numerical integration methods with few integration points are required, even though only having modest accuracy. More precisely, it is considered in this work that estimates of (1) and (2) with errors up to around 10% are useful, if in return only a few hundreds or thousands of points are required.

In optimization applications (e.g. product/systems design, production planning), the number of points is even more critical since an often used strategy is to discretize the original continuous probabilistic problem into a multi-scenario problem, where each integration point corresponds to a plausible scenario. The size and difficulty of this problem is thus determined by the number of integration points.

Also, in many engineering applications, the integrand function  $f(x)$ , usually a model for the cost or performance of the system under analysis, is typically smooth and non-oscillatory in the domain  $X$ , although sometimes having significant anisotropy (some of the  $x$  inputs have much higher effect on the output  $f$  than others). Regarding dimension  $n$ , we are interested in problems with up to around 20 uncertain parameters.

Given the above described problem domain and assumptions, cubature formulae, which are based on a polynomial approximation of  $f(x)$  and may have relatively few points, are a first reasonable choice to estimate integrals (1) and (2).

Referring to integral (1), a cubature formula (or rule) is an approximation of that integral by a weighted sum of  $N$  evaluations of function  $f$ :

$$Q[f] = \sum_{i=1}^N w_i f(x_i). \quad (3)$$

Integration points  $x_i$  and corresponding weights  $w_i$  depend on the integration region  $X$  and weight function  $g(x)$ , but not on  $f(x)$ . A degree  $d$  cubature is exact, i.e.  $Q[f] = I[f]$ , for polynomials  $f(x)$  of degree up to  $d$ . Note that degree  $d$  is the maximum sum of exponents in monomials. In the two-dimensional case, for instance, a degree  $d$  formula is exact for terms  $x_1^a x_2^b$ , with  $a + b \leq d$ . Given that, for any constant  $c$ ,  $E[c] = c$  and that  $Q[c] = E[c]$ , the sum of the weights  $w_i$  is equal to 1.

The same set of points of a given cubature formula may be used to estimate both integrals (1) and (2), as follows:

$$E[f] = \int_X f(x) g(x) dx \approx \sum_{i=1}^N w_i f(x_i) \quad (4)$$

$$V[f] = \int_X [f(x)]^2 g(x) dx - (E[f])^2 \approx \sum_{i=1}^N w_i [f(x_i)]^2 - \left[ \sum_{i=1}^N w_i f(x_i) \right]^2. \quad (5)$$

A cubature formula is said to be *interior* if all points  $x_i$  are in the interior of  $X$ ; and it is *positive* if all weights are positive. Interior formulae are preferred, since function evaluations outside  $X$  may correspond to unreasonable extrapolations. Also, positive formulae are in principle preferable since formulae with large negative weights are ill-conditioned. More precisely, the upper bound of the integration error is proportional to  $(1 + \sum |w_i|)$  [5, Theorem 6.2.4] and therefore negative formulae with  $\sum |w_i|$  much larger than 1 may have a poor performance.

By Tchakaloff's theorem [5, Theorem 6.2.2], there exists a degree  $d$  *positive and interior* cubature approximating integral (1) with at most  $\binom{n+d}{d}$  points, which is known as the Tchakaloff's upper bound. Unfortunately, this is only an existence theorem and there is no general method to construct such interior and positive formulae. The literature on cubatures is vast and there are several methods of construction and a large variety of formulae [6,7], but for general dimension  $n$  and degree  $d \geq 5$ , most known formulae either have large negative weights or a number of points increasing exponentially with  $n$ , well above Tchakaloff's bound.

A straightforward construction of a positive and interior cubature is a *product rule*, which is the tensor product of one-dimensional degree  $d$  rules (each one positive and interior), having a total number of points  $N$  that increases exponentially with  $n$  ( $N = M^n$ , if  $M$  abscissas are used in each dimension). This kind of product formula has more than degree  $d$  of accuracy, being exact for terms up to  $x_1^d x_2^d \dots x_n^d$ . An alternative construction, known as *Smolyak type rule* or *sparse grid integration* [8,9], avoids the exponential growth of  $N$  by combining (adding and subtracting) tensor products of one-dimensional rules with degrees from 1 to  $d$ , and excluding combinations of total degree greater than  $d$ . The result is a formula with exactly degree  $d$  of accuracy (relative to the sum of monomials exponents) and a number of points that only increase polynomially with  $n$  (from now on the expression *few points* will be used to refer to a number of points that only increase polynomially with  $n$ ). The drawback is that negative weights are produced during the combination of the several tensor products.

More recently, Victor [10] and Kuperberg [11], independently, proposed related methods to construct *positive and interior* cubature formulae with few points and in a number significantly below Tchakaloff's upper bound. Both methods are based on error-correcting codes or corresponding orthogonal arrays (OAs). The construction of Kuperberg [11], in particular, starts with an equal-weight degree  $d$  formula having a convolutional structure (such as a product formula), and then removes a large portion of its points using an orthogonal array that preserves degree  $d$  of accuracy. This procedure of points removal is designated by *thinning* and the resulting formula is thus called a *thinned formula*. This OA-based construction is the first general method able to generate positive cubatures with few points. The efficiency is particularly high for relatively low degree

$d$  and high dimension  $n$ . For instance, it is possible to obtain a positive cubature of degree 5 on the  $n$ -cube with a number of points  $N = O(n^2)$ .

In this article, we compare the performance of several cubature formulae, including Smolyak type formulae and thinned cubatures, in the context of probabilistic model analysis and optimization problems and aiming at highly efficient integration, using no more than a few thousands of points, so as to handle large-scale models. The battery of tests performed tries to be representative of a large domain of engineering models (normal and uniform random inputs  $x$ , with  $n = \dim(x)$  up to 24; smooth and non-oscillatory models  $f(x)$ , possibly with significant anisotropy), and includes simple test functions and two practical case studies.

The loss of accuracy due to large negative weights, an issue rarely discussed in the literature, is tested and results clearly show that cubatures with large negative weights, such as Smolyak type formulae, may result in unacceptably large errors. Positive thinned cubatures, on the other hand, which are here extensively tested for the first time, show a much better performance and may produce useful estimates up to dimension 24 with no more than  $2^{12}$  points. Moreover, we also compare how thinned cubatures perform against quasi-Monte Carlo (QMC) integration, often considered the best option for high dimension  $n$  (in particular, we use Sobol points [12,13]). Results are again encouraging, showing that thinned cubatures may beat QMC integration up to dimension 24.

Such a wide comparison study, as above described, has never been presented in the literature. Schürer [14], for instance, performed extensive tests of cubatures vs. QMC integration, but under high accuracy requisites and an upper limit of  $2^{25}$  function evaluations (which is well above the practical limit for large-scale models), and obviously did not include thinned cubatures that were developed later on. Moreover, the comparison tests here presented result in new practical insights into the selection of highly efficient integration methods to solve probabilistic problems, namely the low reliability of negative rules and a new standing for cubatures vs. QMC integration, being shown in this paper that thinned cubatures may be more competitive up to (at least) dimension 24.

The paper is organized as follows. In Section 2, we list the set of formulae to be tested, including positive and non-positive formulae of degrees 5 and 7. Then, in Section 3, the Kuperberg [11] method to construct thinned positive formulae is described in more detail, and from a practitioner's point of view, in an attempt to deliver this valuable material to a wider audience. Next, in Section 4, we present the results from several numerical tests, including two practical case studies. Finally, Section 5 states the most relevant conclusions and its practical value in the context of probabilistic analysis and optimization of large-scale models.

## 2. Cubature formulae

Table 1 lists the cubature formulae to be tested. Label NC refers to *Normal Cubatures* used to estimate expected values in face of normal random inputs  $x$ . The standard construction of these formulae is for integration over  $\mathbb{R}^n$  with weight function  $\exp(-u^T u)$ . The following transformation is needed to adapt the formula to a joint normal probability density function  $g(x)$ , with vector of mean values  $\mu$  and covariance matrix  $\Sigma$ :

$$x(u) = \mu + \sqrt{I} \Sigma^{1/2} u, \quad (6)$$

where  $I$  is the  $n \times n$  identity matrix. Similarly, label UC refers to *Uniform Cubatures*, originally constructed for  $u \in [-1; 1]^n$  and weight function  $g(x) = 1$ . The transformation:

$$x(u) = (x_L + x_U)/2 + u(x_U - x_L)/2 \quad (7)$$

maps the original points  $u$  into the actual points of interest  $x$ , uniformly distributed between  $x_L$  and  $x_U$  (vectorial operations in (7) are performed element by element).

The first two formulae, NC51 and NC52, are normal cubatures of degree 5 from Stroud's compilation [6, Appendix A]. The first one is positive (and thus  $\sum |w_i| = 1$ ) and has a number of points  $N$  increasing with  $2^n$ , which is below Tchakaloff's bound for  $n \leq 13$ . This formula is known to be competitive against QMC sampling up to dimension 8–10 [15]. The second one, NC52, is an example of a degree-5 normal formula with  $N$  only increasing with  $n^2$ , but having some negative weights, which may compromise its accuracy as will be seen later on. Three other formulae with similar efficiency and also having negative weights are reviewed and tested by Lu and Darmofal [16]. The number of points of these formulae is close to the theoretical Möller lower bound, which for degree 5 normal cubatures is  $N_{\min} = n^2 + n + 1$  [16].

Formula NC7 is a normal formula of degree 7, also from Stroud's compilation [6, Appendix A]. For dimension  $n \geq 9$ , a small fraction of its points have small negative weights and thus  $\sum |w_i|$  is only slightly larger than 1. We designate this kind of formula as quasi-positive.

Formulae ThNC51 and ThNC7 are thinned versions of the respective original formulae NC51 and NC7, obtained using adequate OAs as will be described in the next section. Both Victor [10] and Kuperberg [11] presented formula ThNC51 as an example of their thinning methodology. Formula ThNC7 is here presented for the first time using the same OA-based construction.

The Smolyak type formulae SmNC5 and SmNC7 are constructed based on Kronrod–Patterson univariate rules [9], resulting in a sparse grid with slightly less points than the one constructed using the classical univariate Gauss–Hermite rules. The number of points only increases with  $n^2$  for the case of degree 5 and  $n^3$  for degree 7, but in both cases large

**Table 1**  
Cubature formulae.

Label	Description	$N$	$\sum  w_i $
NC51	Normal cubature of degree 5, $n \geq 3$ , presented by Stroud [6].	$2^n + 2n$	1
NC52	A second normal cubature of degree 5, $n \geq 2$ , also presented by Stroud [6].	$2n^2 + 1$	1 if $2 \leq n \leq 4$ . Increases from 1.2 to 3, when $n \geq 5$ .
ThNC51	Formula NC51 with set of $2^n$ points (lying on the vertices of a $n$ -cube) thinned to $2^k$ points; $n \geq 6$ [10,11].	$2^k + 2n^a$ $k = k(n)^a$	1
NC7	Normal cubature of degree 7, $n \geq 2$ , presented by Stroud [6].	$2^{n+1} + 4n^2$	1 if $n \leq 4$ . Increases from 1 to 1.09, for $n$ between 8 and 19, and then slowly decreases till 1.
ThNC7	Formula NC7 with two sets of $2^n$ points thinned to $2^{k+1}$ points; $n \geq 8$ (new formula).	$2^{k+1} + 4n^2$ $k = k(n)^a$	Same behaviour as NC7
SmNC5	Degree-5 Smolyak rule based on Kronrod–Patterson univariate rules [9]; $n \geq 2$ .	$\approx 2n^2$	$O(n^2)$
SmNC7	Degree-7 Smolyak rule based on Kronrod–Patterson univariate rules [9]; $n \geq 2$ .	$\approx \frac{4}{3}n^3$	$O(n^3)$
UC5	Uniform cubature of degree 5, $n \geq 2$ , presented by Stroud [6].	$3n^2 + 3n + 1$	$O(n^2)$
ThUC5	Uniform cubature of degree 5, obtained by thinning a Chebyshev product formula with $4^n$ points to $2^k$ points; $n \geq 3$ [11].	$2^k$ $k = k(2n)^a$	1
ThUC7	Uniform cubature of degree 7, obtained by thinning a Chebyshev product formula with $8^n$ points to $2^k$ points; $n \geq 3$ [11].	$2^k$ $k = k(3n)^a$	1
SmUC5	Degree-5 Smolyak rule based on Kronrod–Patterson univariate rules [9]; $n \geq 2$ .	$\approx 2n^2$	$O(n^2)$
SmUC7	Degree-7 Smolyak rule based on Kronrod–Patterson univariate rules [9]; $n \geq 2$ .	$\approx \frac{4}{3}n^3$	$O(n^3)$

NC—normal cubature, UC—uniform cubature, Th—thinned cubature, Sm—Smolyak type cubature.

<sup>a</sup> Exponent  $k$  as a function of  $n$  is given in Table 2.**Table 2**  
Exponent  $k$  as a function of dimension  $n$ . The domain presented is for  $k \leq 12$ .

I. Degree-5 formulae: $p = n$ (formula ThNC51) or $p = 2n$ (formula ThUC5)							
$p$	6	7	8, 9	10–16	17–24	25–32	33–64
$k$	5	6	7	8	10	11	12
II. Degree-7 formulae: $p = n$ (formula ThNC7) or $p = 3n$ (formula ThUC7)							
$p$	8	9	10	11, 12	13–16		17–24
$k$	7	8	9	10	11		12

negative weights are produced, resulting in bad conditioned formulae ( $\sum |w_i|$  of the order of  $n^2$  and  $n^3$ , respectively for degrees 5 and 7).

Regarding uniform cubatures, UC5 is an old formula from Stroud's compilation (Appendix A), very efficient but bad conditioned. ThUC5 and ThUC7 are thinned equal-weight formulae constructed based on a Chebyshev product rule, as will be described in the next section. Finally, SmUC5 and SmUC7 are Smolyak type formulae [9], also based on Kronrod–Patterson univariate rules as their normal.

Besides cubature formulae, we also test the performance of QMC integration, more specifically using low-discrepancy samples of points based on the Sobol sequence [12,13]. There are some experimental evidences that Sobol points outperform other QMC schemes, namely for high dimensional financial problems [13], and they are often considered among the best general purpose QMC schemes [14,17]. In the particular context of integrals (1) and (2), however, Hammersley points, another type of low-discrepancy points, are also known to be very efficient [18]. In this work, we have tested both Sobol and Hammersley points and observed that, for the battery of tests of Section 4, these two sampling schemes have similar performance or, in some cases, Sobol points are slightly better. We have then adopted Sobol sampling as a fair benchmark for QMC integration (for the sake of clarity, results from Hammersley samples are not shown in this paper).

Sobol sequence of points uniformly fills the hypercube  $[0; 1]^n$  every  $2^m$  points,  $m = 1, 2, \dots$ , and thus, samples with minimal discrepancy should have size  $N = 2^m$ . These standard samples, after transformation (7), may be used to estimate integrals (4) and (5), when inputs  $x$  are uniformly distributed between  $x_L$  and  $x_U$ . In the case of normal distributions, one uses the inverse of the cumulative distribution function to transform standard points  $v_i$ ,  $i = 1, \dots, N$ , uniformly distributed in  $[0; 1]^n$ , into points  $u_i$  representative of a normal distribution with  $\mu = 0$  (zero vector) and  $\Sigma = I$ . This set is then transformed into the  $x$  points of interest following any normal distribution  $N(\mu, \Sigma)$ :  $x(u) = \mu + \Sigma^{1/2}u$ .

**Table 3**

Methods to construct orthogonal arrays  $OA(2^k, n, 2, 5)$ , with  $6 \leq n \leq 32$  [20]. Obtained OAs have minimal  $k$ , except for the case mentioned in footnote b.

Dimension	Method
$6 \leq n \leq 8$	Start with a full array $A$ having as rows the $2^n$ $n$ -tuples with elements $-1$ and $1$ . Define $A_{i,n} = \prod_{j=1}^{n-1} A_{i,j}$ and then delete duplicate rows from the new matrix $A$ . The result is an $OA(2^{n-1}, n, 2, 5)$ .
$n = 9$	Let $G$ be the $7 \times 9$ cyclic matrix $(111000000)$ (this is an abbreviation for the 7 row-vectors 111000000, 011100000, 001110000, ..., 000000111). Let this matrix be defined in $GF(2)$ (Galois field with elements 0 and 1). Then, calculate in $GF(2)$ the set of points $C = \{x \cdot G, x \in GF(2)^7\}$ . $C$ is called a code, in this case composed of $2^7$ codewords (each one of the points $x \cdot G$ ). Finally, perform the transformation from $GF(2)$ to $\{-1, 1\}$ : $0 \mapsto -1, 1 \mapsto 1$ . The result is an $OA(2^7, 9, 2, 5)$ .
$10 \leq n \leq 16$	Generate the code $C = \{x \cdot G, x \in \mathbb{Z}_4^4\}$ , where $G$ is the $4 \times 8$ generator matrix $G = 1(3121000)$ , and calculations done in $\mathbb{Z}_4$ (ring of integers modulo 4). Then perform the transformation from $\mathbb{Z}_4$ to $\{-1, 1\}^2$ : $0 \mapsto (-1, -1), 1 \mapsto (-1, 1), 2 \mapsto (1, 1), 3 \mapsto (1, -1) \mapsto (-1, 1)$ . <sup>a</sup> The result is an $OA(2^8, 16, 2, 5)$ . Simply take the first $n$ columns of this array to obtain an $OA(2^8, n, 2, 5)$ , $10 \leq n \leq 15$ .
$17 \leq n \leq 24^b$	Generate $C = \{x \cdot G, x \in GF(2)^{10}\}$ , with the $10 \times 24$ generator matrix $G$ of Table 5.11 in [20]. Perform the transformation from $GF(2)$ to $\{-1, 1\}$ : $0 \mapsto -1, 1 \mapsto 1$ . The result is an $OA(2^{10}, 24, 2, 5)$ . Take the first $n$ columns to obtain an $OA(2^{10}, n, 2, 5)$ , $17 \leq n \leq 23$ .
$25 \leq n \leq 32$	Generate $C = \{x \cdot G, x \in GF(2)^{12}\}$ , with $G = (11100111000101001100100000000000)1$ , $\dim(G) = 11 \times 32$ . Perform the transformation from $GF(2)$ to $\{-1, 1\}$ : $0 \mapsto -1, 1 \mapsto 1$ . The result is an $OA(2^{11}, 32, 2, 5)$ . Take the first $n$ columns to obtain an $OA(2^{11}, n, 2, 5)$ , $25 \leq n \leq 31$ .

<sup>a</sup> This is also known as the Nordstrom–Robinson code, a particular case of the Kerdock code.

<sup>b</sup> For  $17 \leq n \leq 20$ , the best known OA is an  $OA(2^9, n, 2, 5)$  obtained by a special construction [19] not used in this paper.

### 3. Construction of thinned cubatures

In this section, we illustrate the construction of positive thinned cubatures based on Kuperberg's work [11]. Before stating the main result supporting that construction, the following two definitions are needed.

First, the convolution of two cubature formulae. If  $A$  and  $B$  are two cubature formulae with respective points  $a$  and  $b$  (defined as position vectors), their convolution  $A * B$  is defined as the set of points  $\{c = a + b : a \in A \wedge b \in B\}$ , having respective weights  $w_c = w_a w_b$ . A product rule (above defined as the tensor product of one-dimensional rules) is a particular case of a convolutional formula, obtained through convolutions in perpendicular directions.

Second, the definition of orthogonal array. An  $N \times n$  array with entries from a set  $S$  with  $q$  elements is an orthogonal array with  $q$  levels and strength  $t$  ( $n \geq t$ ) if every  $N \times t$  subarray contains each of the  $q^t$   $t$ -tuples equally often as a row (say  $\lambda$  times). The notation here used for an orthogonal array is  $OA(N, n, q, t)$ .  $N$  must be a multiple of  $q^t$  and  $\lambda = N/q^t$  is the index of the array.

Now, we state Theorem 1.1 of Kuperberg [11]. Let  $t, n$  and  $l$  be positive integers, let  $q$  be a prime power, and let  $\mu$  be a measure in  $\mathbb{R}^n$ . For each  $1 \leq i \leq l$ , let  $F_i$  be an equal-weight formula with  $q$  points such that the convolution  $F = F_1 * F_2 * \dots * F_l$  is a degree  $t$  cubature for  $\mu$ . Then, an  $OA(q^k, l, q, t)$  yields a thinned cubature formula with  $q^k$  points.

This result applies to equal-weight cubature formulae approximating integral (1), with the measure  $\mu$  being such that  $d\mu = g(x) dx$ . It also applies to subsets of points of a formula having a convolutional structure and with all points having the same weight. We now describe particular constructions based on this general result, including those that generate the thinned formulae of Table 1.

The degree-5 formula NC51, from Stroud's compilation [6], has  $2^n + 2n$  points, with the set of  $2^n$  points having a product structure and the same weight. The product structure is the result of  $n$  convolutions of equal-weight one-dimensional formulae, each one with two coordinates,  $s$  and  $-s$ , with  $s$  given by  $s^2 = (n+2)/[2(n-2)]$ . We may then apply Kuperberg's theorem with  $q = 2$  and  $l = n$ , and thin the  $2^n$  points to the orthogonal array  $OA(2^k, n, 2, 5)$  with entries  $s$  and  $-s$ . The thinned formula thus obtained is labelled as ThNC51 and has  $2^k + 2n$  points, with  $k < n$ .

The construction of formula ThNC51 then relies on the existence of an orthogonal array  $OA(2^k, n, 2, 5)$ , and preferably one with minimal  $k$ . A comprehensive map of the best known OAs (those with minimal  $k$ ) and corresponding construction methods (often based on coding theory) are available online [19]. A basic reference is the book by Hedayat et al. [20]. In Table 3, we summarize construction methods for orthogonal arrays  $OA(2^k, n, 2, 5)$ , with  $6 \leq n \leq 32$ . The resulting arrays have entries  $\pm 1$  that should be substituted by adequate coordinates when constructing a given thinned cubature formula (coordinates  $\pm s$ , in the case of formula ThNC51). The resulting value for the power  $k$  is given in Table 2, for  $k \leq 12$ .

A non-specialist may have difficulties in understanding the material in Table 3. It should be noted that there is not a universal method to construct orthogonal arrays and for that reason Table 3 includes different methods. Further, most of the construction methods use non-conventional arithmetic. Fortunately, the textbook by Hedayat et al. [20] is clear enough for a non-specialist and covers all the necessary material. Let us here present in more detail one of the constructions, namely the one for  $10 \leq n \leq 16$ . It begins with a code  $C$  defined in  $\mathbb{Z}_4$ , the ring of integers modulo 4 with elements  $\{0, 1, 2, 3\}$ . This



code is generated as the following set of points (called codewords, in coding theory):

$$C = \{x \cdot G, x \in \mathbb{Z}_4^4\}.$$

Here,  $x$  runs over all possible 4-tuples in  $\mathbb{Z}_4$  and  $G$  is the generator matrix:

$$G = \begin{bmatrix} 1 & 3 & 1 & 2 & 1 & 0 & 0 & 0 \\ 1 & 0 & 3 & 1 & 2 & 1 & 0 & 0 \\ 1 & 0 & 0 & 3 & 1 & 2 & 1 & 0 \\ 1 & 0 & 0 & 0 & 3 & 1 & 2 & 1 \end{bmatrix}.$$

The set  $C$  has  $4^4$  points, each one with 8 coordinates in  $\mathbb{Z}_4$ . In other words, it is  $4^4 \times 8$  matrix with elements from  $\{0, 1, 2, 3\}$ . Now, performing the transformation from  $\{0, 1, 2, 3\}$  to  $\{-1, 1\}^2$  (the vertices of  $[-1, 1]^2$ ):

$$\phi : 0 \mapsto (-1, -1), 1 \mapsto (-1, 1), 2 \mapsto (1, 1), 3 \mapsto (1, -1) \mapsto (-1, 1),$$

one obtains an  $OA(2^8, 16, 2, 5)$  with elements  $\pm 1$  (the rows of this OA are the codewords of a code known as Nordstrom–Robinson code). This OA, substituting  $\pm 1$  by  $\pm s$ , gives the thinned set of  $2^8$  points of formula ThNC51 for  $n = 16$ . The reduction in the number of integration points is thus remarkable, with the original formula NC51 having  $2^{16} + 32 = 65\,568$  points, while its thinned version ThNC51 only has  $2^8 + 32 = 288$  points. For  $10 \leq n \leq 15$ , the shortest known OAs are obtained by simply taking the first  $n$  columns of  $OA(2^8, 16, 2, 5)$ , thus resulting in thinned formulae with  $2^8 + 2n$  points. For  $n = 12$ , for instance, the reduction is from  $2^{12} + 24 = 4120$  points to  $2^8 + 24 = 280$  points.

Table 3 covers construction of strength-5 OAs up to dimension  $n = 32$ . For higher dimension  $n$ , BCH and Kerdock codes are a good choice to construct infinite families of strength-5 OAs [20]: (i) for any even  $m \geq 4$ , there is an  $OA(2^{2m}, 2^m, 2, 5)$ , obtained from a Kerdock code; (ii) for both even and odd values of  $m \geq 5$ , the dual of a BCH code results in an  $OA(2^{2m+1}, 2^m, 2, 5)$ . For  $m$  even, the Kerdock OA is shorter than the corresponding OA based on the BCH code. These are the best presently known codes, only when  $n$  is a power of 2. When  $n$  is not a power of 2, we may still use Kerdock and BCH codes, although they are not the shortest ones [20]. For this, we simply generate an OA with  $2^m$  columns, such that  $2^{m-1} < n < 2^m$ , and then project it down to  $n$  dimensions (take the first  $n$  columns). For dimension  $n = 50$ , for instance, power  $m$  is equal to 6 and thus a Kerdock OA is a good choice. It is an  $OA(2^{12}, 64, 2, 5)$ , which may then be projected to 50 dimensions. This array is the basis to construct a thinned cubature formula of the type ThNC51, having a total of  $2^{12} + 2 \times 50 = 4196$  points.

The uniform cubature ThUC5 is also constructed using strength-5 OAs, but now by thinning an entire primal formula. The construction starts with a degree-5 Chebyshev quadrature in one dimension, which is an equal-weight Gaussian quadrature with 4 points in the interval  $[-1, 1]$  (details are given in Appendix B). The corresponding  $n$ -dimensional product formula is thus a convolutional equal-weight cubature with  $4^n$  points, which may be thinned using an  $OA(4^k, n, 4, 5)$ . A more clever construction may be devised by choosing a quadrature in  $[-1, 1]$  that is *itself* a convolution of more basic formulae [11]. This is the case of the Chebyshev quadrature, whose points, for general degree  $(2r + 1)$ , has the structure  $\pm z_1 \pm z_2 \pm \dots \pm z_r$ . This is clearly the convolution of  $r$  formulae each one with a pair of points  $\pm z_i$ . Therefore, the corresponding  $n$ -dimensional product formula is the convolution of  $rn$  pairs of points and may be thinned using an  $OA(2^k, rn, 2, d)$ . For degree  $d = 5$ , one has  $r = 2$  and hence one needs an  $OA(2^k, 2n, 2, 5)$ . This OA is of the same type of the OAs used to construct normal formulae ThNC51, with the main difference that now it has  $2n$  columns. For dimension  $n = 12$ , for instance, one needs an OA with 24 columns, which may be constructed using the method in row 4 of Table 3, resulting in an  $OA(2^{10}, 24, 2, 5)$ . A thinned cubature with  $2^{10}$  points may then be constructed based on this OA, as follows.

Let  $b_j$ ,  $j = 1, \dots, 2n$ , with each  $b_j \in \{-1, 1\}$ , be the row  $i$  of the  $OA(2^k, 2n, 2, 5)$ . The coordinates of point  $i$  of formula ThUC5 are given by

$$\{x_m : x_m = b_j z_1 + b_{j+1} z_2; j = 1, 3, 5, \dots, 2n - 1; m = (j + 1)/2\}. \quad (8)$$

For  $2n = 10$ , for instance, row 9 of the  $OA(2^8, 10, 2, 5)$  is  $(1, -1, -1, -1, -1, -1, -1, 1, -1, 1)$ . The corresponding point of formula ThUC5, in  $[-1, 1]^5$ , is  $(z_1 - z_2, -z_1 - z_2, -z_1 - z_2, -z_1 + z_2, -z_1 + z_2)$ .

Finally, cubatures ThNC7 and ThUC7 are degree-7 formulae, constructed using strength-7 OAs. The first one, ThNC7, is a new construction that we here propose based on formula NC7 of Stroud's compilation [6]. This one has two sets of  $2^n$  points with a product structure and the same weight. Each one of these sets may be thinned to  $2^k$  points using an  $OA(2^k, n, 2, 7)$ . The methods to construct OAs of this type are not here presented and may be found in references [19,20]. The corresponding values of the power  $k$  are given in Table 2, for dimension  $n$  up to 24. For dimension  $n = 12$ , for instance, there is an  $OA(2^{10}, 12, 2, 7)$  corresponding to a code defined by a parity check matrix [20, p. 105]. Using this OA to thin formula NC7, one then obtains a formula with  $2^{k+1} + 4n^2 = 2^{11} + 4 \times 12^2 = 2624$  points, while the original NC7 formula has  $2^{13} + 4 \times 12^2 = 8768$  points.

The second degree-7 cubature, ThUC7, is constructed likewise formula ThUC5. First, a degree-7 Chebyshev quadrature ( $r = 3$ ) is considered in each dimension, with 8 points  $\pm z_1 \pm z_2 \pm z_3$ . The  $n$ -fold product of this formula is the convolution of  $3n$  basic formulae, each one with a pair of points  $(\pm z_1, \pm z_2$  and  $\pm z_3)$ . This convolution may be thinned using an  $OA(2^k, 3n, 2, 7)$ , resulting in cubature ThUC7 with  $2^k$  points. Table 2 gives values of  $k$  up to 12, which in the case of formula

**Table 4**

Estimation of expected values of  $f(x)$  in face of normal inputs, and for increasing dimension  $n$ . Error  $E$  is absolute deviation in % relative to the analytical solution.  $N$  is the number of integration points.

$n$	NC52		ThNC51 <sup>a</sup>		ThNC7 <sup>b</sup>		SmNC5		SmNC7		Sobol <sup>c</sup>
	$N$	$E$	$N$	$E$	$N$	$E$	$N$	$E$	$N$	$E$	
3	19	5.8	14	4.0	52	1.5	19	4.4	39	1.1	0.0
5	51	8.6	42	1.9	164	1.4	51	13.1	151	7.2	0.1
10	201	12.0	276	0.1	1424	0.8	201	121.1	1201	85.1	0.1
15	451	13.2	286	0.7	4996	0.5	451	434.0	4151	441.8	0.5
20	801	13.6	1064	0.8	9792	0.3	801	1029.0	10001	1428.8	0.7

<sup>a</sup> For  $n \leq 5$ , formula NC51 is used.

<sup>b</sup> For  $n \leq 7$ , formula NC7 is used.

<sup>c</sup> Estimate with 512 Sobol points.

ThUC7 corresponds to a maximum dimension  $n = 8$ . For higher dimension  $n$ , it is still possible to construct cubatures of the type ThUC7, but the number of points may become prohibitive. One of the useful codes is the Delsarte–Goethals code that generates an infinite family of strength-7 OAs:  $OA(2^{3m-1}, 2^m, 2, 7)$ , for any even  $m \geq 6$ . For  $m = 6$ , for instance, this code generates an  $OA(2^{17}, 64, 2, 7)$ , which allows the construction of a cubature ThUC7 with  $n = 21$  and having  $2^{17} \sim 1.3 \times 10^5$  points.

#### 4. Numerical tests

This section presents results from several numerical tests, using different integration techniques. We recall that our accuracy requisites are modest and that the main focus is on highly efficient integration with few points. More precisely, we seek for a high efficiency that may be defined as  $Ef = 1/(E \cdot N)$ , with an error  $E$  that may be as high as 10% and an upper limit for  $N$  of about  $2^{12}$ .

All calculations are made in *Mathematica* [21], including construction of OAs and generation of Sobol samples.

##### 4.1. First illustrative integrand function

We first consider a simple integrand function whose basic form is dimension independent. This integrand was also used by Lu and Darmofal [16] to test the performance of several normal cubatures:

$$f(x) = \frac{1}{\sqrt{1 + x^T x}}. \quad (9)$$

Expected values of this function, for increasing dimension  $n$ , are estimated using different cubatures and also Sobol sampling. All inputs  $x_i$  are normally distributed with mean  $\mu = 0$  and standard deviation  $\sigma = \sqrt{2}/2$  (independent individual distributions). Results are shown in Table 4.

As dimension  $n$  increases, the loss of accuracy of negative formulae (NC52, SmNC5 and SmNC7) is evident, with the observed errors being well correlated with  $\delta = \sum |w_i|$  (see Table 1). For high dimension  $n$ , Smolyak type formulae, for which  $\delta$  increases with  $n^2$  or  $n^3$ , produce meaningless results. On the other hand, positive (ThNC51) or quasi-positive (ThNC7) cubatures have small errors independently of the dimension. Formula ThNC51 has a very high efficiency, in particular for high dimension  $n$ , producing estimates with an error less than 1% with a few hundreds of points (286 points for  $n = 15$  and 1064 for  $n = 20$ ). Sobol sampling, in this case, presents an even higher efficiency, yielding very good estimates with only 512 points.

Function (9) has a bell-shape around the origin that becomes sharper as  $n$  increases. Its integration is thus more difficult as  $n$  increases and is facilitated as  $\sigma$  decreases. For  $\sigma = \sqrt{2}/6$ , the same tendencies of Table 4 are obtained, but now with negative formulae having much lower errors (the highest error is 16% and occurs for formula SmNC5 and  $n = 20$ ). Thinned positive formulae are still much better, with errors below 1%.

In the case of uniform inputs, similar results are obtained, again with Smolyak type formula (and also formula UC5 for which  $\delta$  increases with  $n^2$ ) yielding unacceptably large errors. For uniform inputs  $x \in [-1; 1]^n$  and  $n \geq 10$ , all negative formulae (UC5 and Smolyak type) produce meaningless results (errors above 100%), while the error of ThUC5 stays below 1%. Similarly to the case above, integration difficulty decreases as the input domain shrinks, with negative formulae producing better results, but still worse than the ones of thinned cubatures.

The results of this section clearly illustrate that negative formulae may produce meaningless results and in situations in which positive formulae function very well. This bad performance is *a priori* indicated by the high value of  $\sum |w_i|$  and corresponding bad conditioning of the formulae. As will be seen in the next sections, there are cases in which negative formulae may perform reasonably, and thus results in this section should then be viewed as a pathological case that however illustrates the low reliability of negative formulae, including Smolyak type formulae.

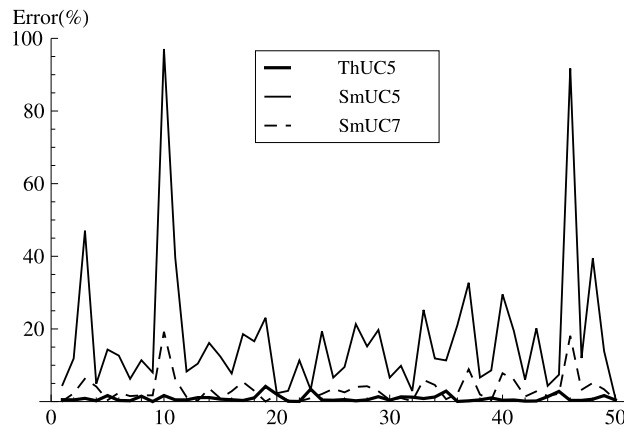


Fig. 1. Error in the variance estimate for 50 elements of family (10) with  $n = 15$ .

#### 4.2. Extensive numerical tests on a representative family of functions

In this section, a more extensive battery of tests is made, based on a single family of functions (adapted from one of the test functions of Genz [22]):

$$f(x) = \prod_{i=1}^n (c + \exp[-a_i(x_i - b_i)^2]). \quad (10)$$

This is the product of  $n$  Gaussian functions  $f_i(x_i)$ , each one with a peak at  $x_i = b_i$  that becomes sharper as  $a_i$  increases. This function will be integrated in  $[0; 1]^n$  with uniform weight or in  $\mathbb{R}^n$  with normal weight (independent random  $x_i$ 's with mean 0.5 and standard deviation  $0.5/3$ , and thus 99.7% of the distribution falls in the interval  $[0; 1]$ ). Parameters  $b_i$  are chosen randomly in  $[-0.2; 1.2]$  leading to both peak and monotonous behaviours (there is a peak along direction  $i$  if  $b \in (0; 1)$  and a monotonous path otherwise). Parameters  $a_i$  are also chosen randomly and in such a way that a subset of them is of high impact (between  $\sim 1$  and  $\sim 3$ ) while the other ones are low impact parameters ( $\sim 0.5$  or lower). In this way, one generates anisotropic functions having much sharper variations along some key directions. The level of anisotropy may be quantified by the ratio  $\max(a)/\min(a)$ , which is around 10 or more. Finally, parameter  $c$  is a constant, here taken as 0.1 ( $c = 0$  leads to functions with long tails having a close to zero value and then artificial numerical errors are obtained). The function value  $f(x)$  is always in  $[0; 1.1^n]$ .

This family of functions is thus a reasonable representation of engineering models where an output index  $y$  is a function of several inputs  $x$ , with much more pronounced variations (that may be monotonous or peak-like) for a subset of more important inputs.

Based on this family of functions, the following battery of integration tests is made. For each input distribution type (normal and uniform), and for dimension  $n$  between 3 and 24, one generates 50 elements of family (10) through random choice of the parameters  $a_i$  and  $u_i$ . In each instance, the performance of the several integration techniques is tested against the analytical solution, computing relative errors (absolute values in %). Both the mean and variance of  $f(x)$  are calculated (integrals (4) and (5)). Analytical solutions are easily obtained since integrals of (10) can be separated as a product of one-dimensional integrals. In rigour, the range of  $n$  values is not exactly 3–24: for low values of  $n$ , when formulae ThNC51 and ThNC7 are not defined, their non-thinned congeners are used; for degree-7 formulae, the highest  $n$  is limited by the maximum imposed on the number of points of about  $2^{12} = 4096$  (maximum  $n$  of 16 for formula ThNC7, 8 for formula ThUC7 and 15 for formulae SmNC7 and SmUC7). Sobol integration is tested with two sample sizes: 256 and 2048 points.

Fig. 1 shows the error in the variance estimate for 50 random elements of family (10) with dimension  $n = 15$  and for inputs with a uniform distribution. It is evident the much better performance of the thinned cubature ThUC5 in comparison with Smolyak type formulae, both of degrees 5 and 7. It is also manifest the high variability of Smolyak formulae results, with errors varying from practically 0 to about 100%, again illustrating the low reliability of this kind of formula. In contrast, the positive cubature ThNC51 consistently produces estimates with errors below 5% and with an average value of only 1%. It should be noted that these numbers do vary significantly from sample to sample of 50 integrands of family (10), but qualitative behaviour remains the same: Smolyak type formulae are not reliable, while thinned cubature ThUC5 consistently produces good estimates.

Figs. 2 and 3 show condensed results for the complete set of tests. The ordinate in the figures is an error index  $E(\%)$  calculated as the mean value plus the standard deviation of the relative error in the sample of 50 integrations, performed for each dimension  $n$ . Each figure has a pair of graphics, one showing the error  $E(\%)$  for the mean estimation (integral (4)) and the other the error for the variance estimation (integral (5)). As expected, variance estimates have larger errors than mean estimates, since the first ones require the integration of the square of  $f(x)$ .



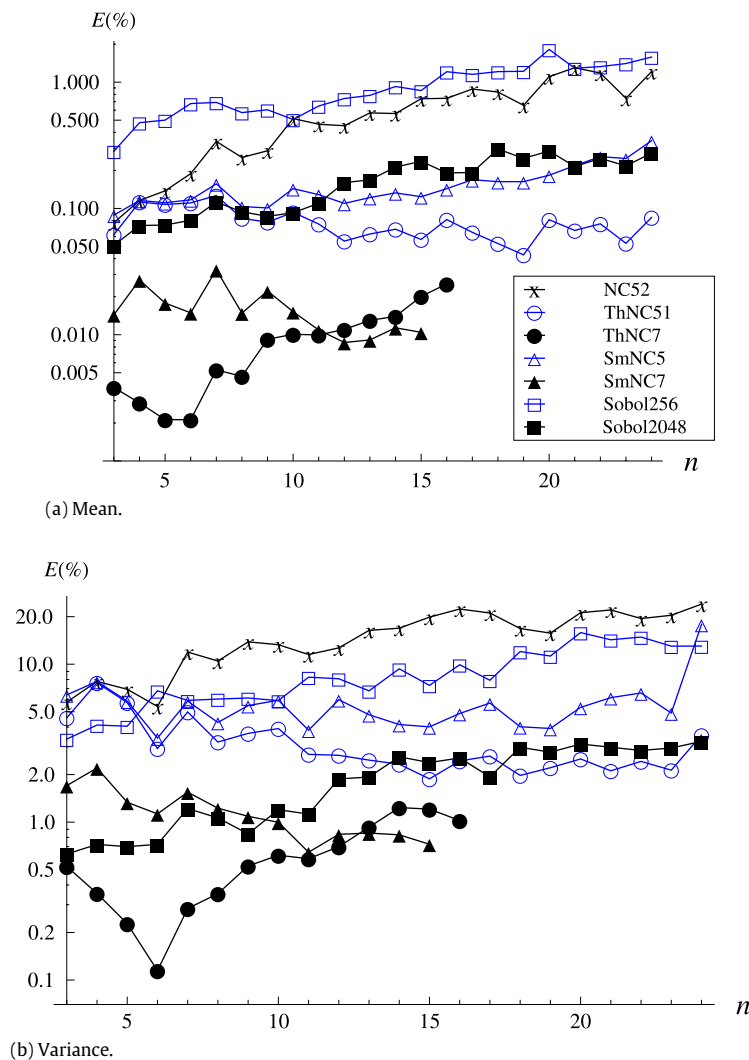
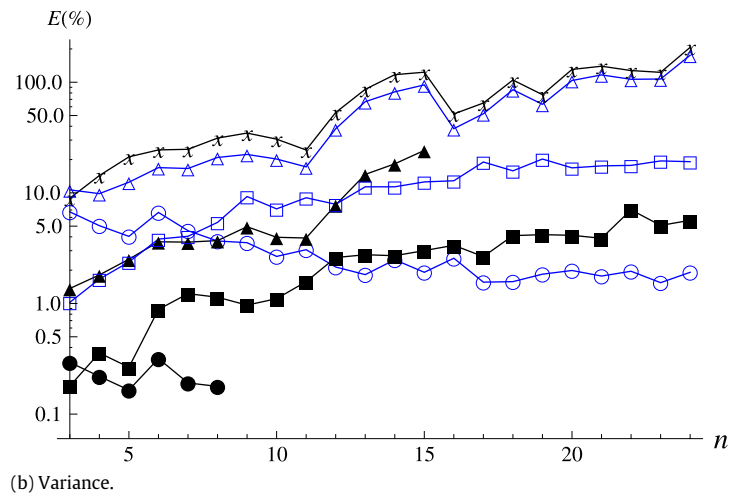
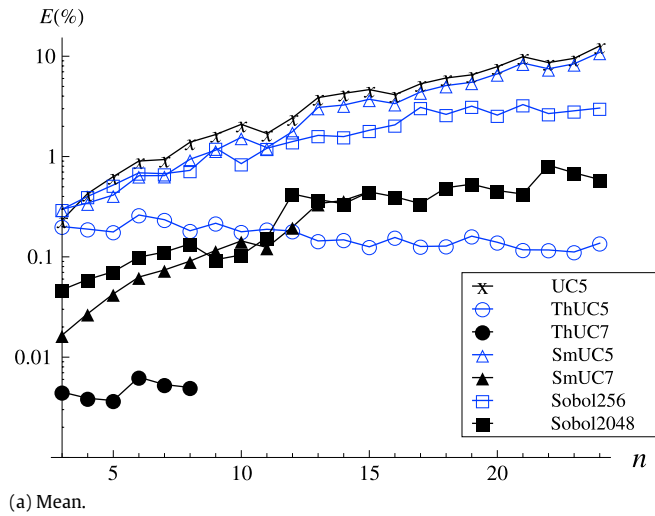


Fig. 2. Mean and variance estimate of family (10) functions with *normal* random inputs.

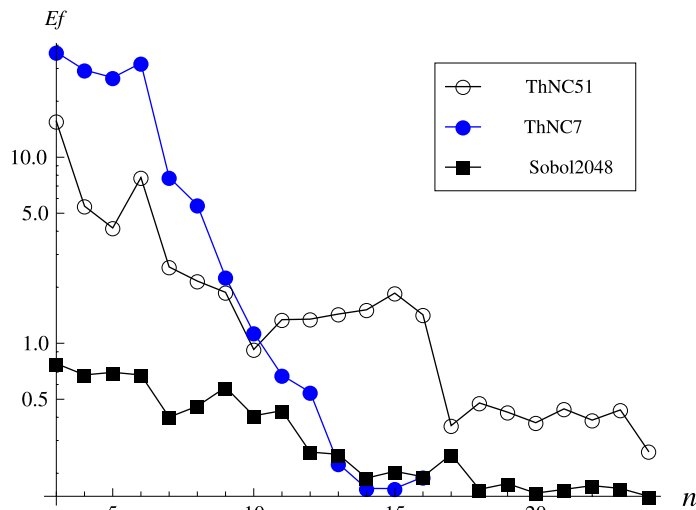
In this broader test domain, the low reliability of negative cubatures is again observed, and particularly for high dimension  $n$ . Negative formulae have a reasonable performance for normal inputs (Fig. 2) but fail in the case of uniform inputs (Fig. 3), with errors systematically increasing with dimension  $n$ , and up to values as high as 100% in the case of degree-5 estimates of the variance (Fig. 3(b), formulae UC5 and SmUC5). Positive and quasi-positive thinned cubatures, in contrast, remain reliable even for uniform inputs (Fig. 3(b)) and high dimension, with errors well below 10%. Formula ThUC5, in particular, has a remarkable performance, even beating the degree-7 formula SmUC7 for high dimension  $n$ .

Regarding thinned cubatures vs. Sobol sampling, since both techniques produce very reasonable estimates, a close comparison is needed, and for that purpose one evaluates the efficiency index  $Ef = 1/(E \cdot N)$  for all cases represented in Figs. 2 and 3. Thinned cubatures have most of the times a higher efficiency  $Ef$ , with the advantage being more expressive for the case of Fig. 2(a) (normal inputs and mean estimation) and then decreasing in the order Figs. 2(b), 3(a) and (b), such that in this last case cubatures and Sobol sampling have similar efficiencies. The intermediate case of Fig. 2(b) (normal inputs and variance estimation) is represented in Fig. 4, where it can be observed that formula ThNC51 is 2–10 times more efficient than a Sobol sample of 2048, with this value stabilizing around 2 for higher dimensions. The last case, corresponding to Fig. 3(b) (uniform inputs and variance estimation), is represented in Fig. 5. All these results agree with the fact that the distinctive advantage of cubatures, in comparison with QMC integration, is their polynomial exactness, being thus expected higher efficiencies when integrating very smooth functions, and a continuous decrease in performance when integrands become less smooth (variance estimation vs. mean estimation) and when the cubature approximation itself requires more points (uniform formulae vs. normal formula).

It is therefore difficult to state absolute conclusions regarding the performance of cubatures against QMC integration, since the second ones will always be superior in the case of very sharp integrands, while the advantage of the first ones is



**Fig. 3.** Mean and variance estimate of family (10) functions with *uniform* random inputs.



**Fig. 4.** Efficiency of thinned cubatures vs. Sobol sampling in estimating the variance of family (10) functions, with *normal* random inputs.

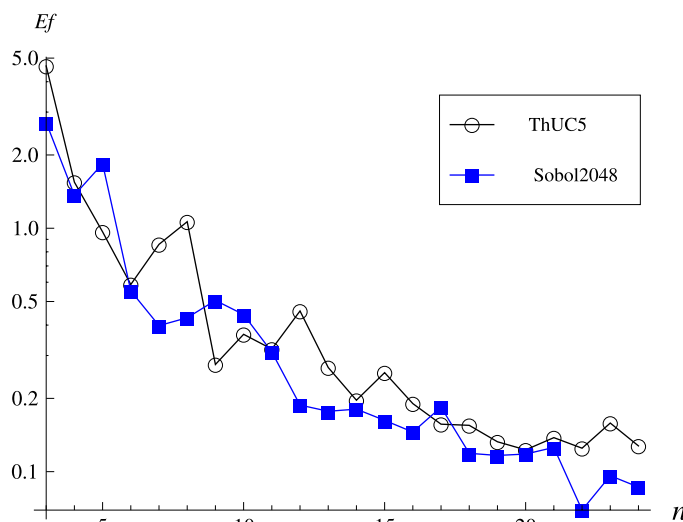


Fig. 5. Efficiency of thinned cubatures vs. Sobol sampling in estimating the variance of family (10) functions, with *uniform* random inputs.

their polynomial exactness. Nevertheless, the tests here performed are in fact representative of the degree of smoothness observed in many engineering models, used for instance in systems design and optimization. It is therefore fair to conclude that positive thinned cubatures *may* likely outperform QMC integration, and up to dimension 24. This represents a new and improved standing for cubatures vs. QMC integration, since the current heuristic limit in the integral dimension, up to which cubatures are considered advantageous, is around 8–10 [15]. Above dimension 24, the advantage of cubatures is likely to decrease since even the most economic formulae of degree 5 have 2048 points or more (see Table 2), and with this number of points it is probable that the best QMC schemes, namely Sobol sampling, will produce an equivalent or better result.

#### 4.3. Example of model analysis: transdermal drug delivery model

This section presents predictions of a mass transfer model with uncertain input parameters. The model equates the rate at which a drug is released from a pharmaceutical ointment and its subsequent transfer across several skin layers until it reaches the blood circulation system. Some of the model parameters are estimated as a function of ointment composition. The main output is the temporal profile of drug concentration in plasma, which is desired to be as constant as possible around an optimal level. A performance index  $y$  is then calculated as the mean deviation of the plasma concentration profile from the target value [23].

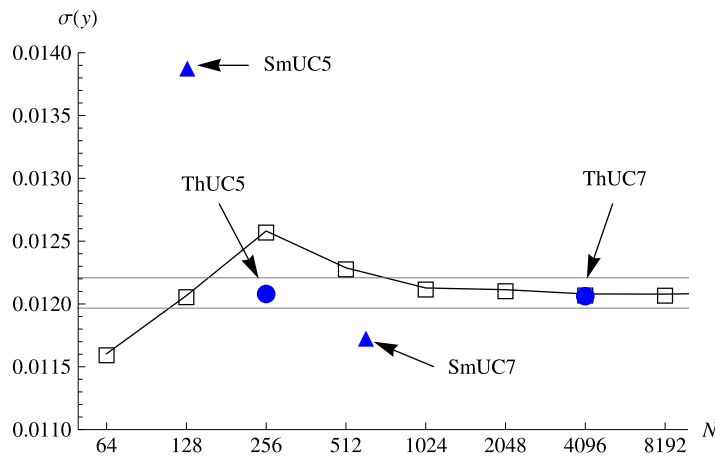
The model consists of a system of partial and ordinary differential equations. A numerical solution is obtained through spatial discretization (along a single drug transfer direction) that results in a large linear system of ordinary differential equations.

Eight input parameters are uncertain, all with uniform distributions between estimated lower and upper limits: a liquid–liquid equilibrium parameter, two kinetic parameters related to the drug molecule diffusion, a kinetic parameter related to drug clearance from blood, two parameters of the ointment production process (that determine ointment microstructure and thus the rate of drug release) and two more parameters related to final ointment application (volume applied and area of application). Model simulation with these 8 uncertain inputs requires the calculation of multidimensional integrals like (4) and (5), where the integrand function  $f(x)$  is the deterministic prediction of the performance index  $y$  for a particular instance of the 8 inputs  $x$ . Function  $y = f(x)$  is known to vary smoothly within the domain of the 8 inputs, and with values between around 0.03 and 0.07. It is also known that some of the inputs have a much higher impact on  $y$  than others.

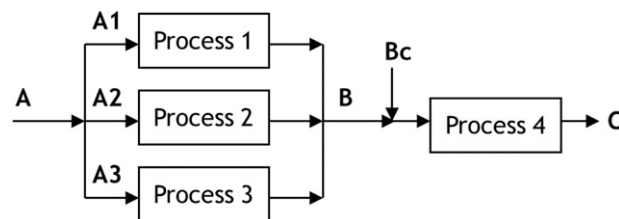
Fig. 6 shows estimated values for the standard deviation of  $y$  using different integration techniques. The horizontal lines are  $\pm 1\%$  margins around the Sobol solution with  $2^{14} = 16\,384$  points, taken as a quasi-exact solution. Thinned cubatures show the best performance. The 5th degree estimate is within the error band with only 256 points, while Sobol sampling needs 1024 points to attain this level of accuracy and the Smolyak type formulae do not attain it even with a 7th degree estimate.

#### 4.4. Example of model optimization: a chemical process planning problem

Let us consider the production network of Fig. 7 [15]. The intermediate product B may be produced via processes 1, 2 and/or 3 or purchased outside (quantity  $B_p$ ). The final product C is obtained through process 4. The production planning problem of this network is to find which of the processes 1, 2 and/or 3 should be operated and what should be the



**Fig. 6.** Standard deviation of the model output  $y$ , calculated using different integration techniques: thinned cubatures (circles), Smolyak type formulae (full triangles) and Sobol sampling (open squares).



**Fig. 7.** Production network.

amounts  $A1, A2, A3$  and  $Bp$  in order to maximize an objective function, equal to the revenues from selling product  $C$  minus raw-materials and processing costs (fixed and variable). Economic input parameters are given, including availability and cost of raw-materials, price and demand of the final product, and processing costs. Also, simple non-linear models for process conversions are known. There are 12 uncertain input parameters, including expected errors in the predicted process conversions. All uncertainties are considered to follow a normal distribution with known mean and variance values.

Ignoring uncertainties in the inputs, the optimal planning problem is a small MINLP (mixed-integer non-linear programming) problem, with 4 binary variables, 10 continuous variables and 18 restrictions. The 4 binaries correspond to the existence of processes 1, 2 and 3 ( $y1, y2$  and  $y3$ ) and whether or not  $B$  is purchase to the exterior ( $yBp$ ). There are several alternatives to formulate this problem handling uncertain parameters. Here, we consider binary variables and the feed to process 3 ( $A3$ ) as rigid decisions that are not adjustable to the different possible concretizations of the uncertain parameters, while the remaining variables are considered recourse decisions adjustable to different uncertainty scenarios. The objective is to maximize the expected profit (revenues minus costs) given the uncertainty in the input parameters. This problem may be formulated and solved using an integration formula with  $N$  points to estimate the objective function and considering all restrictions involving recourse variables indexed over the integration points  $i$ ,  $i = 1, \dots, N$ . In order to ensure feasibility in the entire uncertainty space, one further needs to write restrictions over index  $j$ , with  $j = 1, \dots, M$ , including here the worst case scenarios (a straightforward approach is to include all the  $2^n$  vertices of the uncertainty space; in case of infinite spaces, as is the case for normal distributions, one needs to consider a truncated space). The resulting problem is thus much larger than the original deterministic problem. In this planning example, one has  $5 + 9(N + M)$  decision variables and  $18(N + M)$  restrictions. A subset of  $M = 2^7$  vertices is chosen as extreme limiting scenarios. Normal distributions are truncated such that vertices are located at  $\mu \pm 3.09\sigma$ .

Table 5 shows optimization results, using as integration techniques the thinned cubature ThNC51 or Sobol sampling. Problems were formulated and solved in GAMS, using solvers SBB/CONOPT3 [24].

Problem size (number of variables and restrictions) increases linearly with  $N$ , while CPU time, as expected, increases exponentially with  $N$ . The use of a very efficient integration technique is thus critical for routine problem solving. In this case, the economic cubature ThNC51 yields a very reasonable approximation with only 280 points and 15 s of CPU time. Sobol samples also have good performances and with a similar number of points. Smolyak formula SmNC5, with 289 points, leads to a meaningless solution (impossibly high expected profit of 19722 USD/h, resulting from negative weights).

**Table 5**Planning problem optimal solutions. In all cases:  $y_1 = 0$ ,  $y_2 = y_3 = y_{Bp} = 1$  and  $A_3 = 3.207$  ton/h.

Integration technique	$N$	CPU (s) <sup>a</sup>	Profit (USD/h)	
			Expected value	Standard deviation
ThNC51	280	15	1402	325.1
Sobol	128	9	1403	309.3
Sobol	256	19	1406	321.4
Sobol	512	39	1403	319.4
Sobol	1024	108	1403	319.1
Sobol	2048	436	1403	319.5
Sobol	4096	1508	1403	319.1

<sup>a</sup> Intel Core i5 @ 2.5 GHz.

## 5. Conclusions

We have tested the performance of several cubature formulae (of degrees 5 and 7) in estimating multidimensional integrals relevant in probabilistic analysis problems, with normal or uniform random inputs. Tests were made for a wide family of integrand functions  $f(x)$  representative of typical engineering models (dimension up to 24, smooth and non-oscillatory variations, possibly with significant anisotropy). Two practical case studies were also presented. Results show that cubature formulae with large negative weights, including Smolyak type formulae (also known as sparse grid formulae), are not reliable, producing unacceptably large errors or even meaningless results. On the other hand, thinned positive and quasi-positive formulae, constructed based on orthogonal arrays, have a much better performance and may produce useful estimates up to dimension 24 with no more than  $2^{12}$  points. Further, these thinned cubatures, which were here extensively tested for the first time, may perform better than quasi-Monte Carlo integration schemes, namely Sobol sampling, also up to dimension 24. We believe that these results may help other researchers in choosing highly efficient integration techniques and, in particular, that may contribute to disseminate thinned positive cubatures as a powerful alternative.

## Appendix A. Cubature formulae

The following three formulae are normal cubatures from Stroud's compilation [6], for integration over the entire space  $\mathbb{R}^n$  with weight function  $\exp(-u^T u)$ . The index *FS* designates a set of fully symmetric points, generated by permutation of coordinates and their sign. For instance,  $(r, 0, \dots, 0)_{FS}$  represents the six points  $(\pm r, 0, 0)$ ,  $(0, \pm r, 0)$  and  $(0, 0, \pm r)$ .

### Formula NC51

Domain:  $n \geq 3$ .Points  $u_i$  and respective weights  $w_i$ :

$$\begin{aligned} (r, 0, \dots, 0)_{FS} & w_0 \\ (s, s, \dots, s)_{FS} & w_1. \end{aligned}$$

Parameters:  $r^2 = \frac{n+2}{4}$ ;  $s^2 = \frac{n+2}{2(n-2)}$ ;  $w_0 = \frac{4}{(n+2)^2}$ ;  $w_1 = \frac{(n-2)^2}{2^n(n+2)^2}$ .Number of points:  $N = 2^n + 2n$ .Positive formula:  $\sum |w_i| = 1$ .

### Formula NC52

Domain:  $n \geq 2$ .Points  $u_i$  and respective weights  $w_i$ :

$$\begin{aligned} (0, 0, 0, \dots, 0)_{FS} & w_0 \\ (r, 0, 0, \dots, 0)_{FS} & w_1 \\ (s, s, 0, \dots, 0)_{FS} & w_2. \end{aligned}$$

Parameters:  $r^2 = \frac{n+2}{2}$ ;  $s^2 = \frac{n+2}{4}$ ;  $w_0 = \frac{2}{n+2}$ ;  $w_1 = \frac{4-n}{2(n+2)^2}$ .Negative formula for  $n \geq 5$ , with  $\sum |w_i| = \frac{4-4n+3n^2}{(2+n)^2}$ .Number of points:  $N = 2n^2 + 1$ .

### Formula NC7

Domain:  $n \geq 2$ .

Points  $u_i$  and respective weights  $w_i$ :

$$(r \cdot r_1, 0, \dots, 0)_{FS} \quad B \cdot A_1$$

$$(r \cdot r_2, 0, \dots, 0)_{FS} \quad B \cdot A_2$$

$$(s \cdot r_1, s \cdot r_1, \dots, s \cdot r_1)_{FS} \quad C \cdot A_1$$

$$(s \cdot r_2, s \cdot r_2, \dots, s \cdot r_2)_{FS} \quad C \cdot A_2$$

$$(t \cdot r_1, t \cdot r_1, 0, \dots, 0)_{FS} \quad D \cdot A_1$$

$$(t \cdot r_2, t \cdot r_2, 0, \dots, 0)_{FS} \quad D \cdot A_2$$

Parameters:  $r = 1$ ;  $s^2 = \frac{1}{n}$ ;  $t^2 = \frac{1}{2}$ ;  $r_1^2, r_2^2 = \frac{n+2 \pm \sqrt{2(n+2)}}{2}$ ;

$$B = \frac{2(8-n)}{n(n+2)(n+4)}; \quad C = \frac{2^{-n+1}n^3}{n(n+2)(n+4)}; \quad D = \frac{8}{n(n+2)(n+4)};$$

$$A_1, A_2 = \frac{n+2 \pm \sqrt{2(n-2)}}{4(n+2)}.$$

Negative formula for  $n \geq 9$ , with  $\sum |w_i| = \frac{-24+10n+n^2}{(2+n)(4+n)}$ .

Number of points:  $N = 2^{n+1} + 4n^2$ .

The following formula is a uniform cubature from Stroud's compilation [6], for integration over the  $n$ -cube  $[-1; 1]^n$  with unit weight function. The index  $S$  designates a set of symmetric points, generated by permutation of coordinates without change of sign.

*Formula UC5*

Domain:  $n \geq 2$ .

Points  $u_i$  and respective weights  $w_i$ :

$$(0, 0, 0, \dots, 0) \quad w_0$$

$$(r, r, 0, \dots, 0)_S \quad w_1$$

$$(-r, -r, 0, \dots, 0)_S \quad w_1$$

$$(r, 0, 0, \dots, 0)_{FS} \quad w_2$$

$$(s, -t, 0, \dots, 0)_S \quad w_3$$

$$(-s, t, 0, \dots, 0)_S \quad w_3$$

$$(s, 0, 0, \dots, 0)_{FS} \quad w_4$$

$$(t, 0, 0, \dots, 0)_{FS} \quad w_4.$$

Parameters:  $r = \sqrt{\frac{7}{15}}$ ;  $s = \sqrt{\frac{7+\sqrt{24}}{15}}$ ;  $t = \sqrt{\frac{7-\sqrt{24}}{15}}$ ;

$$w_0 = \frac{5n^2 - 15n + 14}{14}; \quad w_1 = \frac{25}{168}; \quad w_2 = \frac{-25(n-2)}{168}; \quad w_3 = \frac{5}{48}; \quad w_4 = \frac{-5(n-2)}{48}.$$

Negative formulae for  $n \geq 3$ , with  $\sum |w_i| = \frac{10n^2-20n+7}{7}$ .

Number of points:  $N = 3n^2 + 3n + 1$ .

## Appendix B. Chebyshev quadrature formula

The Chebyshev quadrature of degree 5 has 4 points  $\pm z_1 \pm z_2$ , all with the same weight. The parameters  $z_i$  are the square root of the roots (all real positive) of the polynomial:  $x^2 - \frac{x}{3} + \frac{1}{45}$ .

In general, the Chebyshev quadrature of degree  $(2s+1)$  has  $2^s$  points  $\pm z_1 \pm z_2 \pm \dots \pm z_s$ , all with the same weight. The parameters  $z_i$  are the square root of the roots (all real positive) of the polynomial [11]:

$$x^s - \frac{x^{s-1}}{3} + \frac{x^{s-2}}{45} - \dots + \frac{(-1)^s}{1 \cdot 3 \cdot 15 \dots (4^s - 1)}.$$



## References

- [1] U.M. Diwekar, E.S. Rubin, Parameter design methodology for chemical processes using a simulator, *Ind. Eng. Chem. Res.* 33 (1994) 292–298.
- [2] S. Ahmed, N.V. Sahinidis, Robust process planning under uncertainty, *Ind. Eng. Chem. Res.* 37 (1998) 1883–1892.
- [3] E.N. Pistikopoulos, Uncertainty in process design and operations, *Comput. Chem. Eng.* 19 (1995) S553–S563.
- [4] F.P. Bernardo, E.N. Pistikopoulos, P.M. Saraiva, Quality costs and robustness criteria in chemical process design optimization, *Comput. Chem. Eng.* 25 (2001) 27–40.
- [5] A.R. Krommer, C.W. Ueberhuber, *Computational Integration*, SIAM, Philadelphia, PA, 1998.
- [6] A.H. Stroud, *Approximate Calculation of Multiple Integrals*, Prentice Hall, London, 1971.
- [7] R. Cools, P. Rabinowitz, Monomial cubature rules since Stroud: a compilation, *J. Comput. Appl. Math.* 48 (1993) 309–326.
- [8] S.A. Smolyak, Quadrature and interpolation formulas for tensor products of certain classes of functions, *Sov. Math. Dokl.* 4 (1963) 240–243.
- [9] F. Heiss, V. Winschel, Likelihood approximation by numerical integration on sparse grids, *J. Econometrics* 144 (2008) 62–80.
- [10] N. Victor, Asymmetric cubature formulae with few points in high dimension for symmetric measures, *SIAM J. Numer. Anal.* 42 (2004) 209–227.
- [11] G. Kuperberg, Numerical cubature using error-correcting codes, *SIAM J. Numer. Anal.* 44 (2006) 897–907.
- [12] H. Niederreiter, *Random Number Generation and Quasi-Monte Carlo Methods*, SIAM, Philadelphia, 1992.
- [13] P. Glasserman, *Monte Carlo Methods in Financial Engineering*, Springer, 2003.
- [14] R. Schürer, A comparison between (quasi-)Monte Carlo and cubature rule based methods for solving high-dimensional integration problems, *Math. Comput. Simul.* 62 (2003) 509–517.
- [15] F.P. Bernardo, E.N. Pistikopoulos, P.M. Saraiva, Integration and computational issues in stochastic design and planning optimization problems, *Ind. Eng. Chem. Res.* 38 (1999) 3056–3068.
- [16] J. Lu, D.L. Darmofal, Higher-dimensional integration with Gaussian weight for applications in probabilistic design, *SIAM J. Sci. Comput.* 26 (2004) 613–624.
- [17] A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, S. Tarantola, Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index, *Comput. Phys. Comm.* 181 (2010) 259–270.
- [18] U.M. Diwekar, J.R. Kalagnanam, An efficient sampling technique for off-line quality control, *Technometrics* 39 (1997) 308–319.
- [19] W.C. Schmid, R. Schürer, MinT, the online database for optimal parameters of  $(t, m, s)$ -nets,  $(t, s)$ -sequences, orthogonal arrays, linear codes, and OOA's, in.
- [20] A.S. Hedayat, N.J.A. Sloane, J. Stufken, *Orthogonal Arrays. Theory and Applications*, Springer-Verlag, New York, 1999.
- [21] Wolfram MathWorld. <http://mathworld.wolfram.com> (Accessed Jan 2014).
- [22] A.C. Genz, Testing multidimensional integration routines, in: B. Ford, J.C. Rault, F. Thomasset (Eds.), *Tools, Methods, and Languages for Scientific and Engineering Computation*, North-Holland, Amsterdam, 1984, pp. 81–94.
- [23] F.P. Bernardo, P.M. Saraiva, A theoretical model for transdermal drug delivery from emulsions and its dependence upon formulation, *J. Pharm. Sci.* 97 (2008) 3781–3809.
- [24] GAMS—On-line Documentation. <http://www.gams.com/docs/document.htm> (Accessed May 2014).