Tchakaloff-like compression of QMC volume and surface integration on the union of balls

G. Elefante[1,2] · A. Sommariva[2] · M. Vianello[2]

Abstract

We present an algorithm for Tchakaloff-like compression of quasi-Monte Carlo volume and surface integration on an arbitrary union of balls, via non-negative least squares. We also provide the corresponding Matlab codes together with several numerical tests.

1 Introduction

Numerical modelling by finite collections of disks, balls and spheres is relevant within different application fields. Problems involving intersection, union and difference of such geometrical objects arise for example in molecular modelling, computational geometry, computational optics, wireless network analysis; cf., e.g., [1–4, 16, 23, 26, 29, 32] with the references therein. A basic problem is the computation of areas and volumes of such sets, followed by the more difficult task of computing volume and surface integrals there by suitable quadrature formulas.

The numerical quadrature problem on intersection and union of planar disks has been recently treated in [35, 37], providing low-cardinality algebraic formulas with positive weights and interior nodes. On the other hand, numerical volume and surface integration on the union of balls has been studied in the molecular modelling field, where one of the main difficulties concerns efficient and accurate tracking of the complicated resulting surface geometry; cf. e.g. [2, 20, 32] with the references therein.
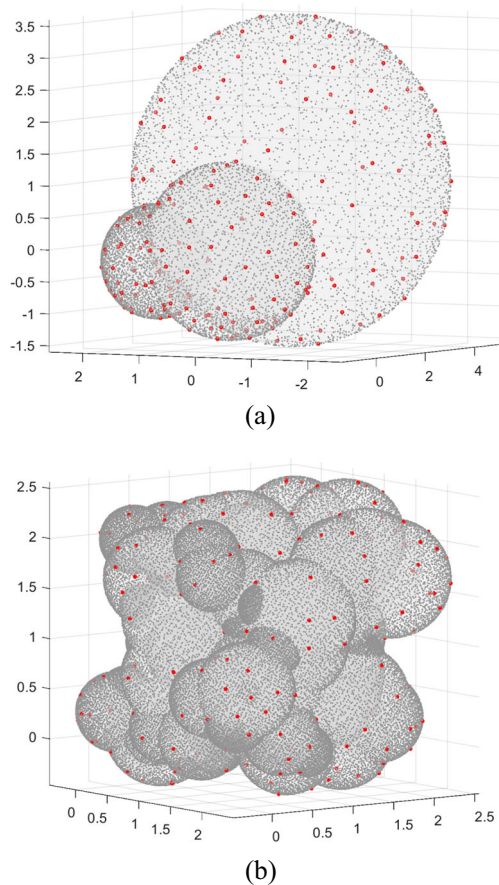
✉ M. Vianello
marcov@math.unipd.it

[1] Department of Mathematics, University of Turin, Turin, Italy

[2] Department of Mathematics, University of Padova, Padua, Italy

(a)

(b)

In this paper, we contribute to this field by providing compressed quasi-Monte Carlo (QMC) formulas for volume and surface integration on the union of balls, along the lines of [17]. Such formulas *preserve the approximation power of QMC* up to the best uniform polynomial approximation error of a given degree to the integrand, but using a *much lower number of sampling points*; see Fig. 1 for two examples with QMC sampling compression. The key tools are a dated but overlooked result by Davis and Wilhelmsen on the so-called "Tchakaloff sets" for positive linear functionals and the Lawson–Hanson algorithm for NNLS (Non-Negative Least Squares), which allows to extract a set of "equivalent" re-weighted nodes from a huge low-discrepancy sequence. We term the method "TDW" (Tchakaloff–Davis–Wilhelmsen), due to its theoretical background.

We stress that differently from [5, 17], the present approach is able to compress not only QMC volume integration, but also QMC integration on compact subsets of algebraic surfaces (in particular, the surface of a union of balls which is a subset of a union of spheres). Notice that one of the main difficulties in surface instances consists in adapting the compression algorithm to work on spaces of polynomials

restricted to an algebraic variety, and finding an appropriate polynomial basis. Indeed, to our knowledge the present work is the first attempt in this direction within the QMC framework. It is also worth stressing that the method could be easily adapted to models involving arbitrary unions of ellipsoids instead of balls, or unions of polyhedra, or even a combination of such objects, in all cases with the advantage of avoiding difficult trackings of the complicated resulting surface geometry.

The paper is organized as follows. In Sect. 2, we discuss theoretical and computational issues of QMC compression for volume and surface integration in $\mathbb{R}^3$. In Sect. 3 we describe our implementation, presenting several numerical tests, including an example illustrating the potential applicability to biomolecular modelling (computation of Generalized Born Radii). The open-source codes are freely available at [18].

2 Tchakaloff-like QMC compression

Compression of QMC formulas is nothing but a special instance of *discrete measure compression*, a topic which has received an increasing attention in the literature of the last decade, in both the probabilistic and the deterministic setting. Indeed, several papers and some software have been devoted to the extraction of a smaller set of re-weighted mass points from the support of a high-cardinality discrete measure, with the constraint of preserving its moments up to a given polynomial degree; cf., e.g., [21, 22, 27, 30, 34, 40] with the references therein.

From the quadrature point of view, this topic has a strong connection with the famous Tchakaloff's theorem [39] on the existence of low-cardinality formulas with positive weights. On the other hand, Tchakaloff's theorem itself is contained in a somewhat deeper but overlooked result by Wilhelmsen [41] on the discrete representation of positive linear functionals on finite-dimensional function spaces (which generalizes a previous result by Davis [8]). Indeed, only quite recently this theorem has been rediscovered as a basic tool for positive cubature via adaptive NNLS moment-matching, cf. [17, 25, 36, 38].

Theorem 1 (Davis [8]–Wilhelmsen [41]) *Let Ψ be the linear span of continuous, real-valued, linearly independent functions $\{\phi_j\}_{j=1,\dots,N}$ defined on a compact set $\Omega \subset \mathbb{R}^d$. Assume that Ψ satisfies the Krein condition (i.e. there is at least one $f \in \Psi$ which does not vanish on Ω) and that L is a positive linear functional on Ψ, i.e. $L(f) > 0$ for every $f \in \Psi$, $f \geq 0$ not vanishing everywhere in Ω.*

If $\{P_i\}_{i=1}^{\infty}$ is an everywhere dense subset of Ω, then for sufficiently large s, the set $X_s = \{P_i\}_{i=1,\dots,s}$ is a Tchakaloff set, i.e. there exist weights $w_k > 0$, $k = 1, \dots, \nu$, and nodes $\{\mathcal{T}_k\}_{k=1,\dots,\nu} \subset X_s \subset \Omega$, with $\nu = card(\{\mathcal{T}_k\}) \leq N$, such that

$$L(f) = \sum_{k=1}^{\nu} w_k f(\mathcal{T}_k), \quad \forall f \in \Psi. \tag{1}$$

As an immediate consequence, we may state the following

Corollary 1 *Let λ be a positive measure on Ω, such that* supp(λ) *is determining for* $\mathbb{P}_n^d(\Omega)$, *the space of total-degree polynomials of degree not exceeding n, restricted to Ω (i.e., a polynomial in $\mathbb{P}_n^d(\Omega)$ vanishing there vanishes everywhere on Ω). Then the thesis of Theorem 1 holds for $L(f) = \int_\Omega f \, d\lambda$.*

Indeed, the integral of a nonnegative and not everywhere vanishing polynomial $f \in \mathbb{P}_n^d(\Omega)$ must be positive (otherwise f would vanish on supp(λ)). Observe that the classical version of Tchakaloff's theorem corresponds to

$$L(f) = I(f) = \int_\Omega f(P) \, dP,$$

with $\Psi = \mathbb{P}_n^d(\Omega)$ and

$$N = N_n^d = dim(\mathbb{P}_n^d(\Omega)). \tag{2}$$

From now on we shall concentrate on the 3-dimensional case ($d = 3$), though most considerations could be extended to general dimension. Notice that the formulation of the Davis–Wilhelmsen theorem is sufficiently general to include volume integrals, i.e. Ω is the closure of a bounded open set and $N = dim(\mathbb{P}_n^3(\mathbb{R}^3)) = \binom{n+3}{3} = (n+1)(n+2)(n+3)/6$, as well as surface integrals on compact subsets of an algebraic variety (in this case $dP = d\sigma$ for the surface measure). In the latter case the dimension of the polynomial space could collapse, for example with $\Omega = S^2 \subset \mathbb{R}^3$ we have $N = (n+1)^2 < \binom{n+3}{3} = (n+1)(n+2)(n+3)/6$.

On the other hand, the Davis–Wilhelmsen theorem can also be applied to a discrete functional like a QMC formula applied to $f \in C(\Omega)$

$$L(f) = Q(f) = \frac{\mu(\Omega)}{M} \sum_{i=1}^M f(P_i) \approx I(f), \tag{3}$$

where

$$X_M = \{P_i\}_{i=1,\dots,M}, \quad M > N,$$

is a *low-discrepancy sequence* on Ω, and $\mu(\Omega)$ can be either a volume or a surface area. Typically one generates a low-discrepancy sequence of cardinality say M_0 on a bounding box or bounding surface $\mathcal{B} \supseteq \Omega$, from which the low-discrepancy sequence on Ω is extracted by a suitable in-domain algorithm. We observe that if $\mu(\Omega)$ is unknown or difficult to compute (as in the case of the union of balls), it can be approximated as $\mu(\Omega) \approx \mu(\mathcal{B})M/M_0$.

Positivity of the functional for $f \in \Psi = \mathbb{P}_n^3(\Omega)$ is ensured whenever the set X_M is $\mathbb{P}_n^3(\Omega)$-determining, i.e. a polynomial vanishing there vanishes everywhere on Ω, or equivalently $dim(\mathbb{P}_n^3(X_M)) = N = dim(\mathbb{P}_n^3(\Omega))$, or even

$$rank(V_M) = N, \tag{4}$$

where

$$V_M = V^{(n)}(X_M) = [\phi_j(P_i)] \in \mathbb{R}^{M \times N} \tag{5}$$

is the corresponding rectangular Vandermonde-like matrix. Notice that, X_M being a sequence, for every $s \leq M$ we have that

$$V_s = V^{(n)}(X_s) = [(V_M)_{ij}], \quad 1 \leq i \leq s, \ 1 \leq j \leq N. \tag{6}$$

The full rank requirement for V_M is not restrictive, in practice. In the case of volume integrals, i.e. when Ω is a three-dimensional domain (a union of balls in the present context), the probability that the $N \times N$ submatrix determinant $det(V_N)$ vanishes dealing with uniformly distributed points is zero, as recently proved in [9]. The same holds true for integration on smooth surfaces admitting an analytic parametrization, cf. [19]. Though the present surface context corresponds to a more complicated instance, since the surface of a union of balls has singularities, we have still numerical evidence that the full rank requirement is always satisfied, working with uniformly distributed points with respect to the surface measure (see Sect. 2.2 about the sampling procedure).

2.1 The TDW algorithm

By Theorem 1, when $M \gg N$ we can then try to find a Tchakaloff set X_s, with $N \leq s < M$, such that a sparse nonnegative solution vector u exists to the underdetermined *moment-matching system*

$$V_s^t u = p = V_M^t e, \quad e = \frac{\mu(\Omega)}{M} (1, \ldots, 1)^t. \tag{7}$$

In practice, we solve (7) via Lawson–Hanson active-set method [24] applied to the NNLS problem

$$\min_{u \geq 0} \|V_s^t u - p\|_2, \tag{8}$$

accepting the solution when the residual size is small, say

$$\|V_s^t u - p\|_2 < \varepsilon \tag{9}$$

where ε is a given tolerance. The nonzero components of u then determine the nodes and weights of a compressed QMC formula extracted from X_s, that is $\{w_k\} = \{u_i : u_i > 0\}$ and $\{\mathcal{T}_k\} = \{P_i : u_i > 0\}$, giving

$$Q_n(f) = \sum_{k=1}^{\nu} w_k f(\mathcal{T}_k), \quad \nu \leq N \ll M, \tag{10}$$

where $Q_n(f) = Q(f)$ for every $f \in \mathbb{P}_n^3(\Omega)$.

Notice that existence of a representation like (10) for $s = M$ is ensured by *Caratheodory's theorem on finite-dimensional conic combinations*, applied to the columns of V_M^t (cf. [30] for a full discussion on this point in the general framework of discrete measure compression). In such a way, however, we would have to work with a much larger matrix, that is we would have to solve directly

$$\min_{u \geq 0} \| V_M^t u - p \|_2, \tag{11}$$

a method developed in the last decade, that we term *Caratheodory-Tchakaloff subsampling*, following [30].

On the contrary, solving (8) on an increasing sequence of smaller problems $s := s_1, s_2, s_3, \ldots$ with $s_1 < s_2 < s_3 < \cdots \leq M$,

$$\min_{u \geq 0} \| V_{s_j}^t u - p \|_2, \quad j = 1, 2, 3, \ldots, \ s_1 \geq N, \tag{12}$$

corresponding to increasingly dense subsets $X_{s_1} \subset X_{s_2} \subset \cdots \subseteq X_M$ (say, "bottom-up"), until the residual becomes sufficiently small, could substantially lower the computational cost. Indeed, as shown in [17] for volume integrals, with a suitable choice of the sequence $\{s_j\}$ the residual becomes extremely small in few iterations, with a final extraction cardinality much lower than M. In the sequel, we shall term *TDW (Tchakaloff–Davis–Wilhelmsen)* the bottom-up approach (12), due to its theoretical background.

Concerning the approximation power of QMC compression, following [17] it is easy to derive the following error estimate

$$
\begin{aligned}
|Q_n(f) - I(f)| &\leq E^{qmc}(f) + 2\,\mu(\Omega)\,E_n(f; X) \\
&\leq E^{qmc}(f) + 2\,\mu(\Omega)\,E_n(f; \Omega),
\end{aligned} \tag{13}
$$

valid for every $f \in C(\Omega)$, where $E^{qmc}(f) = |Q(f) - I(f)|$ and $E_n(f; K) = \inf_{\phi \in \mathbb{P}_n^3(K)} \| f - \phi \|_K$, with $\| g \|_K$ denoting the sup-norm of a bounded function on the discrete or continuous compact set K.

The meaning of (13) is that the compressed QMC functional $Q_n(f)$ retains the approximation power of the original QMC formula, up to a quantity proportional to the best polynomial approximation error to f in the uniform norm on X (and hence by inclusion in the uniform norm on Ω). We recall that the latter can be estimated depending on the regularity of f by multivariate Jackson-like theorems, cf. e.g. [31] for volume integrals where Ω is the closure of a bounded open set, and [33] for the case of the sphere.

On the other hand, we do not deepen here the vast and well-studied topic of QMC convergence and error estimates, recalling only that (roughly) the QMC error $E^{qmc}(f)$ is close to $\mathcal{O}(1/M)$ for smooth functions, to be compared with the $\mathcal{O}(1/\sqrt{M})$ error of MC. For basic concepts and results of QMC theory like discrepancy, star-discrepancy, Hardy–Krause variation, Erdös–Turán–Koksma and Koksma–Hlawka inequalities, on domains and manifolds, we refer the reader to the relevant literature, like e.g. [6, 15] with the references therein.

2.2 Non-overlapping unions

The QMC compression algorithm can be easily extended to the case where Ω (either a volume or a surface) is the finite union of non-overlapping subsets, say $\Omega = \cup_{\ell=1}^{L} \Omega_\ell$, such that sequences of low-discrepancy points are known on bounding sets $\mathcal{B}_\ell \supset \Omega_\ell$, and the indicator function of each Ω_ℓ is computable. Non-overlapping means here that the possible nonempty intersection is at the boundary and its measure is zero.

In this case the overall QMC points are $X = \cup_{\ell=1}^{L} Y_\ell$, with $Y_\ell = \{P_{\ell,i}\}_{i=1}^{M_\ell}$ and $M = card(X) = \sum_{\ell=1}^{L} M_\ell$, where Y_ℓ are the low-discrepancy points of $\mathcal{B}_\ell$ lying in Ω_ℓ. We stress that the low-discrepancy points have to be chosen alternately in order to produce an evenly distributed sequence X_M on the whole Ω, picking the first point in each Ω_ℓ, then the second point in each Ω_ℓ and so on, i.e. constructing the sequence $\{P_{1,1}, P_{2,1}, \ldots, P_{L,1}, P_{1,2}, P_{2,2}, \ldots, P_{L,2}, \ldots\}$.

Moreover, by additivity of the integral the QMC functional becomes

$$Q(f) = \sum_{\ell=1}^{L} Q_\ell(f) \approx \sum_{\ell=1}^{L} I_\ell(f) = I(f),$$

$$Q_\ell(f) = \frac{\mu(\Omega_\ell)}{M_\ell} \sum_{i=1}^{M_\ell} f(P_{\ell,i}) \approx I_\ell(f) = \int_{\Omega_\ell} f(P)\,dP, \quad 1 \le \ell \le L. \quad (14)$$

3 Implementation and numerical tests

In order to show the effectiveness of the TDW compression procedure described in the previous section, we briefly sketch a possible implementation and we present some numerical tests for both, volume and surface integration on an arbitrary union of balls.

Indeed, we compare "Caratheodory-Tchakaloff" compression of multivariate discrete measures as implemented in the general-purpose package dCATCH [14], with the TDW approach. All the tests have been performed with a CPU AMD Ryzen 5 3600 with 48 GB of RAM, running Matlab R2022a. The Matlab codes and demos, collected in a package named *Qbubble*, are freely available at [18]. Below, we first give some highlights on the main features of the implemented algorithm. These are essentially:

- for volume integrals we simply take Halton points of the smaller Cartesian bounding box for the union of balls, say $[a_1, b_1] \times [a_2, b_2] \times [a_3, b_3]$, and select those belonging to the union of balls; for surface integrals we follow the procedure sketched in Sect. 2.2, taking on each sphere low-discrepancy mapped Halton points by an area preserving transformation (see (19) in Sect. 3.2 below), and then selecting those belonging to the surface of the union of balls;
- in view of extreme ill-conditioning of the standard monomial basis, we start from the product Chebyshev total-degree basis of the smaller bounding box for Ω (for either volumes or surfaces), namely

$$p_j(x, y, z) = T_{\alpha_1(j)}(\sigma_1(x)) \, T_{\alpha_2(j)}(\sigma_2(y)) \, T_{\alpha_3(j)}(\sigma_3(z)), \quad j = 1, \ldots, J,$$

where $J = (n+1)(n+2)(n+3)/6$, $\sigma_i(t) = \frac{2t - b_i - a_i}{b_i - a_i}$, $i = 1, 2, 3$, and $j \mapsto \alpha(j)$ corresponds to the graded lexicographical ordering of the triples $\alpha = (\alpha_1, \alpha_2, \alpha_3)$, $0 \leq \alpha_1 + \alpha_2 + \alpha_3 \leq n$;

- for surface integrals we determine a suitable polynomial basis by computing the rank and then possibly performing a column selection by QR factorization with column pivoting of the trivariate Chebyshev–Vandermonde matrix;

- in order to cope with ill-conditioning of the Vandermonde-like matrices V_{s_j} (that increases with the degree), we perform a single QR factorization with column pivoting $V_{s_j} = Q_{s_j} R_{s_j}$ to construct an orthogonal polynomial basis w.r.t. the discrete scalar product $\langle f, g \rangle_{X_{s_j}} = \sum_{i=1}^{s_j} f(P_i)g(P_i)$ and substitute V_{s_j} by Q_{s_j} in (12); consequently the QMC moments p in (7) have to be modified to $(R_{s_j}^{-1})^t p$ (via Gaussian elimination);

- the (modified) TDW-NNLS problems (12) are solved by the recent implementation of the Lawson–Hanson active-set method named LHDM, based on the concept of "Deviation Maximization" instead of "column pivoting" for the underlying QR factorizations, since it gives experimentally a speed-up of at least 2 and up to 4 times with respect to the standard Matlab function `lsqnonneg` (cf. [10, 12–14]).

In the next subsections we present several numerical tests, to show the effectiveness of the TDW approach for volume and surface QMC compression on the union of balls.

3.1 Volume integration on the union of balls

In this subsection we consider volume integration on the union of balls, namely

$$\Omega = \bigcup_{j=1}^{\mathcal{N}} B(C_j, r_j) \tag{15}$$

where $B(C_j, r_j) \subset \mathbb{R}^3$ is the closed 3-dimensional ball with center C_j and radius r_j. Here we generate a sequence of Halton points in the smallest Cartesian bounding box for Ω and, then, we select those belonging to the union, say $X = \{P_i\}$, simply by checking that $\|P_i - C_j\|_2 \leq r_j$ for some j.

More precisely, we consider the following (see Fig. 1)

- first example: union of the 3 balls with centers $C_1 = (0, 0, 0)$, $C_2 = (0, 1.3, -0.2)$, $C_3 = (2.5, 0, 1)$ and radii $r_1 = 1.4$, $r_2 = 0.9$, $r_3 = 1$, respectively;
- second example: union of 100 balls with randomly chosen and then fixed centers in $[0, 2]^3$ and radii in $[0.2, 0.6]$.

The results concerning application of the TDW approach are collected in Table 1, where we compress QMC volume integration by more than one million of Halton points, preserving polynomial moments up to degree 3, 6, 9, 12, 15 (the moments correspond to the product Chebyshev basis of the minimal Cartesian bounding box for the ball union).

We start from 2,400,000 Halton points in the bounding box and we set $s_1 = 2N$ and $s_{j+1} = 2s_j$, $j \geq 1$. The residual tolerance is $\varepsilon = 10^{-10}$. The comparisons

Table 1 Compression of QMC volume integration on the union of 3 balls (top) and 100 balls (bottom), in a bounding box with 2,400,000 Halton points; M is the starting QMC cardinality, i.e. the number of Halton points in the union of balls

Degree n	3	6	9	12	15
Card. QMC			$M = 1{,}128{,}709$		
Card. Q_n^{dcatch}	20	84	220	452	806
Card. Q_n^{tdw}	20	84	220	455	816
Compr. ratio	5.6e+04	1.3e+04	5.1e+03	2.5e+03	1.4e+03
cpu Halton seq			0.9s		
cpu Q_n^{dcatch}	3.4s	19s	49s	140s	310s
cpu Q_n^{tdw}	0.22s	0.9s	2.4s	5.7s	26s
Speed-up	15.4	21.1	20.5	24.4	11.9
Mom. resid. Q_n^{dcatch}	8.9e−12	8.9e−12	8.9e−12	⋆5.1e−06	⋆1.1e−05
Mom. resid. Q_n^{tdw}					
Iter. 1	4.6e−16	1.5e−02	1.6e−01	3.8e−01	7.1e−01
Iter. 2		1.1e−15	1.9e−15	3.6e−15	8.1e−15
Card. QMC			$M = 1{,}195{,}806$		
Card. Q_n^{dcatch}	20	83	220	450	795
Card. Q_n^{tdw}	20	84	220	455	816
Compr. ratio	5.6e+04	1.3e+04	5.1e+03	2.8e+03	1.5e+03
cpu Halton seq			1.3s		
cpu Q_n^{dcatch}	3.4s	23s	65s	150s	370s
cpu Q_n^{tdw}	0.25s	0.87s	26s	95s	67s
Speed-up	13.8	26.6	25.0	15.7	5.6
Mom. resid. Q_n^{dcatch}	1.1e−11	⋆1.2e−05	1.1e−11	⋆5.6e−05	⋆7.3e−05
Mom. resid. Q_n^{tdw}					
Iter. 1	2.2e−16	9.4e−02	5.0e−01	1.5e+00	1.8e+00
Iter. 2		1.3e−15	2.2e−15	4.7e−15	8.3e−02
Iter. 3					7.3e−15

of the present TDW compression algorithm, for short Q_n^{tdw}, are made with a global compression algorithm that works on the full Halton sequence X_M, namely the general purpose discrete measure compressor dCATCH developed in [14], which essentially solves directly (11) by Caratheodory–Tchakaloff subsampling as proposed in [30, 34]. In particular, we display the cardinalities and compression ratios, the cpu-times for the construction of the low-discrepancy sequence (cpu Halton seq.) and those for the computation of the compressed rules, where the new algorithm shows speed-ups from about 6 to more than 24 in the present degree range, ensuring moment residuals always below the required tolerance in at most 3 iterations. It is worth stressing a phenomenon

Table 2 Geometric mean of the volume integration errors over 100 trials of the random polynomial (16) for the dCATCH and TDW algorithms on the union of 3 balls and 100 balls

Degree n	3	6	9	12	15
3 balls: dcatch	1e−11	1e−11	1e−11	1e−07	1e−07
3 balls: tdw	2e−11	5e−11	1e−10	2e−10	6e−10
100 balls: dcatch	1e−11	6e−06	1e−11	9e−06	1e−07
100 balls: tdw	1e−11	2e−11	3e−11	5e−11	7e−11

already observed in [17], that is, possible failure of Q_n^{dcatch} which in some cases gives much larger residuals than Q_n^{tdw}.

In order to check polynomial exactness of the QMC compressed rules, in Table 2 we show the relative QMC compression errors and their geometric mean over 100 trials of the polynomial

$$g(P) = (ax + by + cz + d)^n, \quad P = (x, y, z) \tag{16}$$

where a, b, c, d are uniform random variables in $[0, 1]$.

Moreover, in Table 3 we show the integration relative errors on three test functions with different regularity, namely

$$f_1(P) = |P - P_0|^5, \quad f_2(P) = \cos(x + y + z), \quad f_3(P) = \exp(-|P - P_0|^2) \tag{17}$$

where $P_0 = (0, 0, 0) \in \Omega$, the first being of class C^4 with discontinuous fifth derivatives whereas the second and the third are analytic. The reference values of the integrals have been computed by a QMC formula starting from 10^8 Halton points in the bounding box.

We see that the compressed formulas on more than one million points show errors of comparable order of magnitude, that as expected from estimate (13) decrease while increasing the polynomial compression degree until they reach a size close to the underlying QMC error.

3.2 Surface integration on the union of balls

We turn now to surface integration, on a domain Ω that is the boundary of an arbitrary union of balls, namely

$$\Omega = \partial \bigcup_{j=1}^{\mathcal{N}} B(C_j, r_j) = \bigcup_{j=1}^{\mathcal{N}} \partial B(C_j, r_j) \setminus \bigcup_{j=1}^{\mathcal{N}} \overset{\circ}{B}(C_j, r_j), \tag{18}$$

i.e. the set of all points lying on some sphere $\partial B(C_j, r_j)$, $j = 1, \ldots, \mathcal{N}$, but not internally to any of the balls $B(C_k, r_k)$, $k \neq j$.

We present two examples, corresponding to the same centers and radii considered above for volume integration, i.e. the surface of the union of 3 balls and of 100

Table 3 Errors of compressed QMC volume integration on the union of 3 balls (top) and 100 balls (bottom); the reference values of the integrals are computed via QMC starting from 10^8 Halton points in the bounding box

Degree n	3	6	9	12	15
$E^{qmc}(f_1)$			3.5e−04		
$E^{dcatch}(f_1)$	1.3e−01	3.4e−04	3.5e−04	3.5e−04	3.5e−04
$E^{tdw}(f_1)$	2.3e−03	3.2e−04	3.5e−04	3.5e−04	3.5e−04
$E^{qmc}(f_2)$			7.3e−04		
$E^{dcatch}(f_2)$	2.4e+00	7.0e−02	4.3e−03	7.3e−04	7.3e−04
$E^{tdw}(f_2)$	7.5e−01	3.7e−03	4.8e−04	7.4e−04	7.3e−04
$E^{qmc}(f_3)$			8.7e−05		
$E^{dcatch}(f_3)$	7.1e−01	1.4e−01	9.4e−03	2.2e−03	1.1e−04
$E^{tdw}(f_3)$	5.8e−01	2.8e−02	1.5e−02	9.5e−04	2.5e−05
$E^{qmc}(f_1)$			1.1e−04		
$E^{dcatch}(f_1)$	8.3e−02	8.8e−05	1.1e−04	1.1e−04	1.1e−04
$E^{tdw}(f_1)$	1.7e−03	9.8e−05	1.1e−04	1.1e−04	1.1e−04
$E^{qmc}(f_2)$			1.7e−04		
$E^{dcatch}(f_2)$	2.9e−01	8.7e−04	1.6e−04	1.7e−04	1.7e−04
$E^{tdw}(f_2)$	5.6e−02	1.5e−04	1.7e−04	1.7e−04	1.7e−04
$E^{qmc}(f_3)$			2.2e−04		
$E^{dcatch}(f_3)$	2.3e−01	2.3e−03	8.4e−04	2.3e−04	2.2e−04
$E^{tdw}(f_3)$	6.1e−03	3.6e−03	1.2e−04	2.3e−04	2.2e−04

balls in Sect. 3.1. Notice that Ω is a subset of an algebraic surface, the union of the corresponding active spheres. Though the polynomial spaces dimension could be determined theoretically by algebraic geometry methods (cf., e.g., [7]), we do not enter this delicate matter here, since the algorithm computes numerically such a dimension by a rank revealing approach on a Vandermonde-like matrix.

In this case we have applied the extension discussed in Sect. 2.2, constructing an evenly distributed sequence X_M on the whole Ω by taking a large number of low discrepancy points on each sphere $\partial B(C_j, r_j)$, and then selecting those belonging to the portions of the sphere that contribute to the surface of the union, that are those not internal to any other ball. Namely, we have taken on each sphere the mapped Halton points from the rectangle $[-1, 1] \times [0, 2\pi]$ by the *area preserving* transformation

$$(t, \phi) \mapsto C_j + r_j(\sqrt{1 - t^2} \cos(\phi), \sqrt{1 - t^2} \sin(\phi), t), \tag{19}$$

which preserves also the low-discrepancy property (cf. e.g. [11]). The points are finally ordered by picking alternatively one point per active portion of the surface of the

Table 4 Compression of QMC surface integration on the union of 3 balls (top, starting from 500,000 low-discrepancy points on each sphere) and 100 balls (bottom, starting from 60,000 low-discrepancy points on each sphere); M is the starting QMC cardinality, i.e. the number of Halton points on the surface of the union of balls

Degree n	3	6	9	12	15
Card. QMC			$M = 1{,}024{,}179$		
Card. Q_n^{dcatch}	20	83	200	371	572
Card. Q_n^{tdw}	20	83	200	371	596
Compr. ratio	5.1e+04	1.2e+04	5.1e+03	2.8e+03	1.7e+03
cpu Halton seq			0.88s		
cpu Q_n^{dcatch}	2.8s	17s	50s	140s	320s
cpu Q_n^{tdw}	0.31s	1.1s	2.7s	5.8s	65s
Speed-up	9.0	15.5	18.5	24.1	4.9
Mom. resid. Q_n^{dcatch}	8.6e−12	8.9e−12	8.9e−12	8.9e−12	⋆ 1.6e−06
Mom. resid. Q_n^{tdw}					
Iter. 1	7.2e−01	1.4e−15	2.8e−15	4.2e−15	2.6e−01
Iter. 2	3.7e−16				1.3e−01
Iter. 3					3.3e−12
Card. QMC			$M = 1{,}032{,}718$		
Card. Q_n^{dcatch}	20	84	219	455	807
Card. Q_n^{tdw}	20	84	220	455	816
Compr. ratio	5.2e+04	1.2e+04	4.7e+03	2.3e+03	1.3e+03
cpu Halton seq			1.5s		
cpu Q_n^{dcatch}	2.8s	16s	43s	110s	240s
cpu Q_n^{tdw}	0.32s	1.1s	3.0s	6.8s	24s
Speed-up	8.8	14.5	14.3	16.2	10
Mom. resid. Q_n^{dcatch}	9.0e−13	9.1e−13	⋆ 3.2e−06	9.3e−13	⋆ 1.8e−05
Mom. resid. Q_n^{tdw}					
Iter. 1	2.2e+00	1.2e+00	5.5e−01	6.2e−01	1.4e−01
Iter. 2	7.5e−16	1.3e−15	2.5e−15	5.2e−15	1.2e−14

union, with a local weight attached to each point. An illustration of compressed points extracted starting from 4000 mapped Halton points on each sphere is given in Fig. 1.

In Table 4 we report for these surface integration examples the same quantities appearing in Table 1 for the volume integration, where we use again the dCATCH code in [13] to compress the QMC formula on the whole X_M, since also that algorithm was conceived to work with polynomial spaces possibly restricted to algebraic surfaces. Here we start from 500,000 mapped Halton points on each sphere in the 3 balls example, and from 60,000 on each sphere in the 100 balls instance, obtaining a sequence of about one million low-discrepancy points on the corresponding ball union surfaces. As before we set $s_{j+1} = 2s_j$, $j \geq 1$ with $s_1 = 2N$ and $\varepsilon = 10^{-10}$.

Table 5 Geometric mean of the surface integration errors over 100 trials of the random polynomial (16) for the dCATCH and TDW algorithms on the union of 3 balls and 100 balls

Degree n	3	6	9	12	15
3 balls: dcatch	4e−12	4e−12	4e−12	4e−12	1e−09
3 balls: tdw	7e−12	4e−12	1e−11	3e−12	1e−12
100 balls: dcatch	4e−13	3e−13	1e−07	3e−13	3e−07
100 balls: tdw	1e−13	1e−13	2e−13	2e−13	3e−13

Again we get impressive compression ratios, and speed-ups varying from about 5 to more than 16. Moreover, the TDW algorithm gives always a residual below the given tolerance, whereas dCATCH turns out to be more prone to failure (see the residuals for degree $n = 15$ in the example with 3 balls and degrees $n = 9, 15$ in the example with 100 balls).

The compression errors and their geometric mean concerning surface integration of the random polynomial (16), restricted to the boundary of the union, are stated in Table 5. In Table 6 we show the surface integration errors for the three test functions in (17), where P_0 is a suitably chosen point on the surface of the ball union. We see again that the compressed formulas on more than one million points show errors of comparable order of magnitude, that as expected from estimate (13) decrease while increasing the polynomial compression degree, until they reach a size close to the underlying QMC error.

3.2.1 Compressed QMC surface integration of vector fields

We consider here surface integration of a vector field $\vec{F} = (F_1, F_2, F_3)$ on $\Omega = \partial \bigcup_{j=1}^{N} B(C_j, r_j) = \bigcup_{\ell=1}^{L} \Omega_\ell$, that is

$$\int_\Omega \langle \vec{F}(P), \vec{n}(P) \rangle \, d\sigma = \sum_{\ell=1}^{L} \int_{\Omega_\ell} \langle \vec{F}(P), \vec{n}(P) \rangle \, d\sigma, \tag{20}$$

where $\vec{n}(P)$ is the exterior normal vector to the surface at P and $\langle \cdot \cdot \rangle$ denotes the scalar product in $\mathbb{R}^3$. In this case even if the vector field is smooth, the integrand naturally presents singularities of the normal vector crossing the curves corresponding to intersections of different spheres (cf. Sect. 2.2 for the general splitting into nonoverlapping pieces Ω_ℓ). For this reason, it is here convenient to apply QMC compression locally, i.e. on each Ω_ℓ separately, where polynomial approximation can work. In such a way of course we get lower compression ratios, namely if M is the overall number of sampling points we have a compression ratio $M/(L \cdot dim(\mathbb{P}_n^3(\Omega_\ell))) = M/(L(n + 1)^2)$ instead of $M/dim(\mathbb{P}_n^3(\Omega))$.

In order to give a specific application, we consider the computation of a relevant quantity in (bio)molecular solvation models, the so-called Born Radius of an atom. This quantity, that measures in some sense its degree of shielding from solvent by the surrounding molecule's atoms, is a key ingredient for the computation of the solvation free energy. In particular, we consider the Generalized Born (GB) model, where the

Table 6 Errors of compressed QMC surface integration on the union of 3 balls (top) and 100 balls (bottom); the reference values of the integrals are computed via QMC starting from 10^6 points on each sphere

Degree n	3	6	9	12	15
$E^{qmc}(f_1)$			3.9e−06		
$E^{dcatch}(f_1)$	2.3e−04	8.3e−06	4.0e−06	3.9e−06	3.9e−06
$E^{tdw}(f_1)$	3.7e−04	3.6e−06	4.0e−06	3.9e−06	3.9e−06
$E^{qmc}(f_2)$			8.6e−05		
$E^{dcatch}(f_2)$	3.5e−01	3.2e−02	8.3e−04	8.4e−05	8.6e−05
$E^{tdw}(f_2)$	1.3e+00	2.2e−02	8.3e−06	8.5e−05	8.6e−05
$E^{qmc}(f_3)$			5.8e−06		
$E^{dcatch}(f_3)$	3.9e−01	5.8e−03	6.9e−04	5.8e−05	8.9e−06
$E^{tdw}(f_3)$	2.5e−02	3.8e−03	4.7e−06	9.1e−05	6.0e−06
$E^{qmc}(f_1)$			4.0e−05		
$E^{dcatch}(f_1)$	2.9e−03	2.8e−05	3.9e−05	4.0e−05	4.0e−05
$E^{tdw}(f_1)$	3.2e−02	3.7e−05	3.9e−05	4.0e−05	4.0e−05
$E^{qmc}(f_2)$			2.0e−04		
$E^{dcatch}(f_2)$	1.4e−01	6.4e−04	1.7e−04	2.0e−04	2.0e−04
$E^{tdw}(f_2)$	1.4e−01	6.2e−05	1.9e−04	2.0e−04	2.0e−04
$E^{qmc}(f_3)$			1.6e−04		
$E^{dcatch}(f_3)$	2.9e−02	1.7e−02	4.3e−04	1.5e−04	1.6e−04
$E^{tdw}(f_3)$	1.6e−02	1.7e−03	2.6e−04	1.5e−04	1.6e−04

k-th Born Radius of the j-th atom can be defined via the surface integrals

$$I_{j,k} = \int_\Omega \langle \vec{F}_{j,k}(P), \vec{n}(P) \rangle \, d\sigma, \quad \vec{F}_{j,k}(P) = \frac{P - C_j}{|P - C_j|^k}, \quad 4 \leq k \leq 7, \quad (21)$$

for $1 \leq j \leq \mathcal{N}$, where $\Omega = \partial \bigcup_{j=1}^{\mathcal{N}} B(C_j, r_j)$ is for example the Van der Waals surface of the solute molecule (the surface of a union of balls modelling the single atoms). For example, in the popular R_6 GB model, the Born radius of the j-th atom is defined by $R_6(j) = \left(\frac{1}{4\pi} I_{j,6}\right)^{-1/3}$, $1 \leq j \leq N$; cf., e.g., [2, 20, 28] with the references therein, for an overview on Generalized Born Radii in molecular modelling.

For the only purpose of illustration, we have computed the integrals $I_{j,k}$ in (21) for the union of balls in Fig. 1b, considered as a virtual "molecule" with $\mathcal{N} = 100$ "atoms". In particular, we can avoid the discontinuity curves by excluding in the sampling on each given sphere not only the interior, but also the boundary of the other balls. This has no effect on surface integration since there is a finite number of discontinuity curves, that have null surface measure.

The numerical results are collected in Table 7, where we report the RMSRE (Root Mean Square Relative Error) in the QMC computation of the array of surface integrals

Table 7 Cardinalities and root mean square relative errors (22) of QMC surface integration for the Born-like integrals (21) on a virtual molecule with 100 atoms (piecewise versus global compression); the reference values of the integrals are computed via QMC starting from 10^6 points on each sphere

Degree n	3	6	9	12	15
Card. QMC			1,032,718		
Card. dcatch pcw	1719	6519	21,699	41,216	68,707
Compr. ratio	601	158	48	25	15
$\mathcal{E}_4$ QMC			4.3e−04		
$\mathcal{E}_4$ dcatch glob	2.3e+00	7.3e−01	4.4e−01	4.1e−01	1.6e−01
$\mathcal{E}_4$ dcatch pcw	2.9e−02	5.0e−03	9.7e−04	7.0e−04	4.5e−04
$\mathcal{E}_5$ QMC			7.6e−04		
$\mathcal{E}_5$ dcatch glob	5.8e+00	1.7e+00	7.4e−01	6.2e−01	3.2e−01
$\mathcal{E}_5$ dcatch pcw	5.5e−02	1.0e−02	2.0e−03	1.4e−03	8.1e−04
$\mathcal{E}_6$ QMC			1.2e−03		
$\mathcal{E}_6$ dcatch glob	1.1e+01	2.9e+00	9.9e−01	6.8e−01	4.7e−01
$\mathcal{E}_6$ dcatch pcw	8.2e−02	1.6e−02	3.0e−03	2.2e−03	1.3e−03
$\mathcal{E}_7$ QMC			1.6e−03		
$\mathcal{E}_7$ dcatch glob	1.9e+01	3.4e+00	1.2e+00	7.1e−01	5.6e−01
$\mathcal{E}_7$ dcatch pcw	1.0e−01	2.1e−02	3.8e−03	3.0e−03	1.8e−03

$(I_{1,k}, \ldots, I_{\mathcal{N},k})$, that is

$$\mathcal{E}_k = \sqrt{\frac{1}{\mathcal{N}} \sum_{j=1}^{\mathcal{N}} \frac{(\mathcal{Q}_{j,k} - I_{j,k})^2}{I_{j,k}^2}}, \quad 4 \le k \le 7, \tag{22}$$

where either

$$\mathcal{Q}_{j,k} = Q(\langle \vec{F}_{j,k}, \vec{n} \rangle) = \sum_{\ell=1}^{L} Q_\ell(\langle \vec{F}_{j,k}, \vec{n} \rangle), \quad 1 \le j \le \mathcal{N},$$

is the uncompressed piecewise QMC formula, or

$$\mathcal{Q}_{j,k} = Q_n(\langle \vec{F}_{j,k}, \vec{n} \rangle) = \sum_{\ell=1}^{L} Q_{\ell,n}(\langle \vec{F}_{j,k}, \vec{n} \rangle), \quad 1 \le j \le \mathcal{N},$$

is the piecewise compressed one (named "pcw"). As above we start from 60,000 mapped Halton points per sphere and take as reference values of the integrals those computed with 10^6 mapped Halton points per sphere. We apply directly piecewise Caratheodory–Tchakaloff subsampling by the package dCATCH, since experimentally TDW does not give a clear advantage, essentially because the local number of

points is not huge. On the other hand, here global compression does not work satisfactorily due to the discontinuities of the normal vector, as we can clearly see in Table 7, where the errors of "dcatch glob" decrease slowly with the degree, staying at least two orders of magnitude above those of piecewise compression. The resulting compression ratios are much smaller than those obtainable by global compression in smooth instances, but still meaningful. As expected the errors of piecewise compressed QMC integration decrease with the degree, until they reach a size close to the underlying QMC error.

References

1. Avis, D., Bhattacharya, B.K., Imai, H.: Computing the volume of the union of spheres. Vis. Comput. **3**, 323–328 (1988)
2. Bardhan, J.P., et al.: Numerical integration techniques for curved-element discretizations of molecule-solvent interfaces. J. Chem. Phys. **127**, 014701 (2007)
3. Bauman, B., Sommariva, A., Vianello, M.: Compressed cubature over polygons with applications to optical design. J. Comput. Appl. Math. **370**, 112658 (2020)
4. Bauman, B., Xiao, H.: "Gaussian quadrature for optical design with noncircular pupils and fields, and broad wavelength range", Proc. SPIE 7652, International Optical Design Conference 2010, 76522S (9 September 2010); https://doi.org/10.1117/12.872773
5. Bittante, L., De Marchi, S., Elefante, G.: A new quasi-Monte Carlo technique based on nonnegative least-squares and approximate Fekete points. Numer. Math. Theory Methods Appl. **9**, 640–663 (2016)
6. Brandolini, L., Choirat, C., Colzani, L., Gigante, G., Seri, R., Travaglini, L.: Quadrature rules and distribution of points on manifolds. Ann. Sc. Norm. Super. Pisa Cl. Sci. **5**(XIII), 889–923 (2014)
7. Cox, D.A., Little, J., O'Shea, D.: Ideals, Varieties, and Algorithms, 4th edn. Springer, Berlin (2015)
8. Davis, P.J.: A construction of nonnegative approximate quadratures. Math. Comput. **21**, 578–582 (1967)
9. Dell'Accio, F., Sommariva, A., Vianello, M.: Random sampling and unisolvent interpolation by almost everywhere analytic functions. Appl. Math. Lett. **145**, 108734 (2023)
10. Dell'Orto, M., Dessole, M., Marcuzzi, F.: The Lawson–Hanson algorithm with deviation maximization: finite convergence and sparse recovery. Numer. Linear Algebra Appl. **30**, e2490 (2023)
11. De Marchi, S., Elefante, G.: Quasi-Monte Carlo integration on manifolds with mapped low-discrepancy points and greedy minimal Riesz s-energy points. Appl. Numer. Math. **127**, 110–124 (2018)
12. Dessole, M., Marcuzzi, F.: Deviation maximization for rank-revealing QR factorizations. Numer. Algorithms **91**, 1047–1079 (2022)
13. Dessole, M., Marcuzzi, F., Vianello, M.: Accelerating the Lawson–Hanson NNLS solver for large-scale Tchakaloff regression designs. Dolomites Res. Notes Approx. DRNA **13**, 20–29 (2020)
14. Dessole, M., Marcuzzi, F., Vianello, M.: dCATCH: a numerical package for d-variate near G-optimal Tchakaloff regression via fast NNLS. MDPI-Mathematics **8**(7), 1122 (2020). Special Issue "Numerical Methods"
15. Dick, J., Pillichshammer, F.: Digital Nets and Sequences. Discrepancy Theory and Quasi-Monte Carlo Integration. Cambridge University Press, Cambridge (2010)
16. Duan, X., Quan, C., Stamm, B.: A boundary-partition-based Voronoi diagram of d-dimensional balls: definition, properties, and applications. Adv. Comput. Math. **46**, 44 (2020)
17. Elefante, G., Sommariva, A., Vianello, M.: CQMC: an improved code for low-dimensional compressed quasi-MonteCarlo cubature. Dolomites Res. Notes Approx. DRNA **15**, 92–100 (2022)
18. Elefante, G., Sommariva, A., Vianello, M.: Qbubble: compressed QMC volume and surface integration on union of balls (in Matlab). https://github.com/alvisesommariva/qBubble

19. Elefante, G., Sommariva, A., Vianello, M.: Qsurf: compressed QMC integration on parametric surfaces. J. Approx. Softw. (to appear). https://ojs.unito.it/index.php/JAS
20. Fogolari, F., et al.: Bluues: a program for the analysis of the electrostatic properties of proteins based on generalized Born radii. BMC Bioinform. **13**(Suppl 4), S18 (2012)
21. Hayakawa, S.: Monte Carlo cubature construction. Jpn. J. Ind. Appl. Math. **38**, 561–577 (2021)
22. Hayakawa, S., Tanaka, K.: Monte Carlo construction of cubature on Wiener space. Jpn. J. Ind. Appl. Math. **39**, 543–571 (2022)
23. Kim, I.J., Na, H.: An efficient algorithm calculating common solvent accessible volume. PLoS ONE **17**(3), e0265614 (2022)
24. Lawson, C.L., Hanson, R.J.: Solving Least Squares Problems. Classics in Applied Mathematics, vol. 15. SIAM, Philadelphia (1995)
25. Legrain, G.: Non-negative moment fitting quadrature rules for fictitious domain methods. Comput. Math. Appl. **99**, 270–291 (2021)
26. Librino, F., Levorato, M., Zorzi, M.: An algorithmic solution for computing circle intersection areas and its applications to wireless communications. Wirel. Commun. Mob. Comput. **14**, 1672–1690 (2014)
27. Litterer, C., Lyons, T.: High order recombination and an application to cubature on Wiener space. Ann. Appl. Probab. **22**, 1301–1327 (2012)
28. Onufriev, A.V., Case, A.V.: Generalized born implicit solvent models for biomolecules. Annu. Rev. Biophys. **48**, 275–296 (2019)
29. Petitjean, M.: Spheres unions and intersections and some of their applications in molecular modeling. In: Mucherino, A., et al. (eds.) Distance Geometry: Theory, Methods, and Applications, pp. 61–83. Springer, New York (1976)
30. Piazzon, F., Sommariva, A., Vianello, M.: Caratheodory–Tchakaloff subsampling. Dolomites Res. Notes Approx. DRNA **10**, 5–14 (2017)
31. Plésniak, W.: Multivariate Jackson inequality. J. Comput. Appl. Math. **233**, 815–820 (2009)
32. Quan, C., Stamm, B.: Mathematical analysis and calculation of molecular surfaces. J. Comput. Phys. **322**, 760–782 (2016)
33. Ragozin, D.L.: Constructive polynomial approximation on spheres and projective spaces. Trans. Am. Math. Soc. **162**, 157–170 (1971)
34. Sommariva, A., Vianello, M.: Compression of multivariate discrete measures and applications. Numer. Funct. Anal. Optim. **36**, 1198–1223 (2015)
35. Sommariva, A., Vianello, M.: Numerical quadrature on the intersection of planar disks. FILOMAT **31**, 4105–4115 (2017)
36. Sommariva, A., Vianello, M.: Computing Tchakaloff-like cubature rules on spline curvilinear polygons. Dolomites Res. Notes Approx. DRNA **14**, 1–11 (2021)
37. Sommariva, A., Vianello, M.: Cubature rules with positive weights on union of disks. Dolomites Res. Notes Approx. DRNA **15**, 73–81 (2022). (Special Issue for the 60th of S. De Marchi)
38. Sommariva, A., Vianello, M.: Low-cardinality positive interior cubature on NURBS-shaped domains. BIT Numer. Math. **63**, 22 (2023)
39. Tchakaloff, V.: Formules de cubatures mécaniques à coefficients non négatifs, (French). Bull. Sci. Math. **81**, 123–134 (1957)
40. Tchernychova, M.: Caratheodory cubature measures. Ph.D. dissertation in Mathematics (supervisor: T. Lyons), University of Oxford (2015)
41. Wilhelmsen, D.R.: A nearest point algorithm for convex polyhedral cones and applications to positive linear approximation. Math. Comput. **30**, 48–57 (1976)