Compressed QMC volume and surface integration on union of balls

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

University of Padova, Italy

Abstract

We discuss an algorithm for Tchakaloff-like compression of Quasi-MonteCarlo (QMC) volume/surface integration on arbitrary union of balls (multibubbles), that can be of interest in the field of biomolecular modelling. The key tools are Davis-Wilhelmsen theorem on "Tchakaloff sets" for positive linear functionals on polynomial spaces, and Lawson-Hanson algorithm for NNLS. We also provide the corresponding Matlab package.

Keywords: union of balls, biomolecule surface, volume integrals, surface integrals, Quasi-MonteCarlo formulas, quadrature compression, NonNegative Least Squares.

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., [3, 4, 15, 21, 24, 26, 29] 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 [33, 35], providing low-cardinality algebraic formulas with positive weights and interior nodes. Though relevant literature and software are devoted to the computation of volumes and surface areas of arbitrary union of balls, mainly in the molecular modelling field, there are apparently few numerical integration codes on such domains. On the other hand, numerical integration on molecular surfaces, modelled as the surface of union of balls
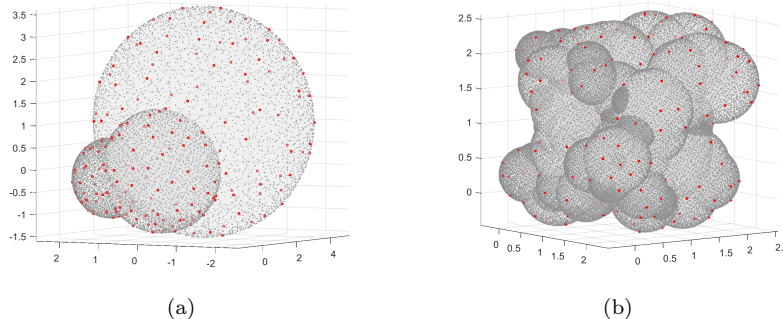
Figure 1: Compressed QMC points (red) extracted from low-discrepancy points (grey) on the surface of ball union at degree $n = 9$. Left: 200 points extracted from about 8200 (3 balls), compression ratio 43; Right: 220 points extracted from about 69000 (100 balls), compression ratio over 300.

such as van der Waals and Solvent Accessible Surfaces, is important in several biomolecular physics problems; cf. e.g. [1, 2, 26, 29] with the references therein.

In this paper, we contribute to fill the gap by providing compressed Quasi-Montecarlo (QMC) formulas for volume and surface integration on union of balls, along the lines of [16]. 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 Figure 1 for two examples with QMC sampling compression. The key tools are Davis-Wilhelmsen theorem on the so-called "Tchakaloff sets" for positive linear functionals and Lawson-Hanson algorithm for NNLS, which allows to extract a set of "equivalent" re-weighted nodes from a huge low-discrepancy sequence.

We stress that differently from [16], 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, 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, with the advantage of avoiding difficult trackings of the complicated resulting surface geometry.

The paper is organized as follows. In Section 2, we discuss theoretical and computational issues of QMC compression for volume and surface integration in $\mathbb{R}^3$. In Section 3 we describe our implementation, presenting several numerical tests. The open-source codes are freely available at [17].

2

2 Compressed QMC formulas

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., [19, 20, 25, 27, 32, 38] with the references therein.

From the quadrature point of view, this topic has a strong connection with the famous Tchakaloff theorem [37] on the existence of low-cardinality formulas with positive weights. On the other hand, Tchakaloff theorem itself is contained in a somewhat deeper but somehow overlooked result by Wilhelmsen [39] on the the discrete representation of positive linear functionals on finite-dimensional function spaces (which generalizes a previous result by Davis [7]). Indeed, only quite recently this theorem has been rediscovered as a basic tool for positive cubature via adaptive NNLS moment-matching, cf. [16, 23, 34, 36].

Theorem 1. *(Davis, 1967 - Wilhelmsen, 1976) 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 m, the set $X_m = \{P_i\}_{i=1,\dots,m}$ is a Tchakaloff set, i.e. there exist weights $w_k > 0$, $k = 1, \dots, \nu$, and nodes $\{\mathcal{T}_k\}_{k=1,\dots,\nu} \subset X_m \subset \Omega$, 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 $\mathrm{supp}(\lambda)$ 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 $supp(\lambda)$). Observe that the classical version of Tchakaloff theorem corresponds to

$$L(f) = L_{\mathrm{INT}}(f) = \int_\Omega f(P) \, dP ,$$

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 in general dimension. Notice that the formulation of Davis-Wilhelmsen theorem is sufficiently general to include volume integrals, i.e. Ω is the closure of a bounded open set and $N = dim(\mathbb{P}_n^d(\mathbb{R}^3)) = \binom{n+3}{3} = (n+1)(n+2)(n+3)/6$, 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, Wilhelmsen theorem can also be applied to a discrete functional like a QMC formula applied to $f \in C(\Omega)$

$$L(f) = L_{\text{QMC}}(f) = \frac{\mu(\Omega)}{M} \sum_{i=1}^{M} f(P_i) \approx \int_{\Omega} f(P) \, dP \,, \tag{3}$$

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 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. 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 $k \leq M$ we have that

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

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 $det(V_N) = 0$ dealing with uniformly distributed points is null, as recently proved in [8]. The same holds true for integration on smooth surfaces admitting an analytic parametrization, cf. [18]. 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 Remark 1 below).

4

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

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

In practice, we solve (7) via Lawson-Hanson active-set method [22] applied to the NNLS problem

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

accepting the solution when the residual size is small, say

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

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_m, that is $\{w_k\} = \{u_i : u_i > 0\}$ and $\{\mathcal{T}_k\} = \{P_i : u_i > 0\}$, giving

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

where $L_{\mathrm{QMC}}^*(f) = L_{\mathrm{QMC}}(f)$ for every $f \in \mathbb{P}_n^3(\Omega)$.

Notice that existence of a representation like (10) for $m = M$ is ensured by Caratheodory theorem on finite-dimensional conic combinations, applied to the columns of V_M^t (cf. [27] 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

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

On the contrary, solving (8) on an increasing sequence of smaller problems $m := m_1, m_2, m_3, \dots$ with $m_1 < m_2 < m_3 < \dots \leq M$,

$$\min_{u \geq 0} \|V_{m_j}^t u - p\|_2 , \quad j = 1, 2, 3, \dots , \; m_1 \geq N , \tag{12}$$

corresponding to increasingly dense subsets $X_{m_1} \subset X_{m_2} \subset \dots \subseteq X_M$ (say, "bottom-up"), until the residual becomes sufficiently small, could substantially lower the computational cost. Indeed, as shown in [16] for volume integrals, with a suitable choice of the sequence $\{m_j\}$ the residual becomes extremely small in few iterations, with a final extraction cardinality much lower than M.

Concerning the approximation power of QMC compression, following [16] it is easy to derive the following error estimate

$$|L_{\mathrm{QMC}}^*(f) - L_{\mathrm{INT}}(f)| \leq \mathcal{E}_{\mathrm{QMC}}(f) + 2\,\mu(\Omega)\,E_n(f; X)$$

$$\leq \mathcal{E}_{\mathrm{QMC}}(f) + 2\,\mu(\Omega)\,E_n(f; \Omega) , \tag{13}$$

valid for every $f \in C(\Omega)$, where $\mathcal{E}_{\mathrm{QMC}}(f) = |L_{\mathrm{QMC}}(f) - L_{\mathrm{INT}}(f)|$ 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 L_{QMC}^* 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. [28] for volume integrals where Ω is the closure of a bounded open set.

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 $\mathcal{E}_{\mathrm{QMC}}(f)$ 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, we refer the reader to devoted books like e.g. [14].

Remark 1. The QMC compression algorithm can be easily extended to the case where Ω (either a volume or a surface) is the finite union of nonoverlapping subsets, say $\Omega = \cup_{\ell=1}^L \Omega_\ell$, such that sequences of low-discrepancy points are known on bounding sets $\mathcal{B}_\ell \supset \Omega_\ell$. 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 alternatively 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

$$L_{QMC}(f) = \sum_{\ell=1}^L \sum_{i=1}^{M_\ell} w_{\ell,i} f(P_{\ell,i}) \approx \sum_{\ell=1}^L \int_{\Omega_\ell} f(P)\, dP = \int_\Omega f(P)\, dP \,, \qquad (14)$$

where $w_{\ell i} = \mu(\Omega_\ell)/M_\ell$, $i = 1, \ldots, M_\ell$, and hence the QMC moments in (7) have to be computed with such weights.

3   Implementation and numerical tests

In order to show the effectiveness of the bottom-up 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 arbitrary union of balls.

Indeed, we compare "Caratheodory-Tchakaloff" compression of multivariate discrete measures as implemented in the general-purpose package *dCATCH* [13], with the bottom-up 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 [17].

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 bounding box
$$[a_1, b_1] \times [a_2, b_2] \times [a_3, b_3] \supset \Omega$$
and select those belonging to Ω; for surface integrals we follow the procedure sketched in Remark 1, taking on each sphere $\mathcal{B}_\ell$ low-discrepancy mapped Halton points by an area preserving transformation (see (19) in Section 3.2 below), and then selecting those belonging to the surface;

- 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
$$p_j(x, y, z) = T_{\alpha_1(j)}\left(\sigma_1(x)\right) T_{\alpha_2(j)}\left(\sigma_2(y)\right) T_{\alpha_3(j)}\left(\sigma_3(z)\right) \ , \quad j = 1, \ldots, J \ ,$$
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 $\alpha = (\alpha_1, \alpha_2, \alpha_3)$, $0 \le \alpha_1 + \alpha_2 + \alpha_3 \le n$;

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

- the (modified) bottom-up NNLS problems (12) are solved by the recent implementation of 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 with respect to the standard Matlab function `lsqnonneg` (cf. [9, 11, 12]).

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

3.1 Volume integration on union of balls

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

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

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 Figure 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 bottom-up 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 $m_1 = 2N$ and $m_{j+1} = 2m_j$, $j \geq 1$. The residual tolerance is $\varepsilon = 10^{-10}$. The comparisons of the present bottom-up compression algorithm, for short Q_c^{bu}, 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 [13], which essentially solves directly (11) by Caratheodory-Tchakaloff subsampling as proposed in [32, 27].

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 already observed in [16], that is possible failure of Q_c^{dCATCH} which in some cases give much larger residuals than Q_c^{bu}.

In order to check polynomial exactness of the QMC compressed rules, in Figure 2 we show the relative QMC compression errors and their logarithmic averages (i.e. the sum of the log of the errors divided by the number of trials) 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 2 we show the integration relative errors on three test functions with different regularity, namely

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

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 QMC error.

deg	3	6	9	12	15
card. QMC	$M = 1,128,709$				
card. Q_c^{dCATCH}	20	84	220	452	806
card. Q_c^{bu}	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.	9.0e-01s				
cpu Q_c^{dCATCH}	3.4e+00s	1.9e+01s	4.9e+01s	1.4e+02s	3.1e+02s
cpu Q_c^{bu}	2.2e-01s	9.0e-01s	2.4e+00s	5.7e+00s	2.6e+01s
speed-up	15.4	21.1	20.5	24.4	11.9
mom. resid. Q_c^{dCATCH}	8.9e-12	8.9e-12	8.9e-12	⋆ 5.1e-06	⋆ 1.1e-05
mom. resid. Q_c^{bu}					
iter. 1	4.55e-16	1.51e-02	1.63e-01	3.81e-01	7.12e-01
iter. 2		1.12e-15	1.85e-15	3.62e-15	8.06e-15

deg	3	6	9	12	15
card. QMC	$M = 1,195,806$				
card. Q_c^{dCATCH}	20	83	220	450	795
card. Q_c^{bu}	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.	1.3e+00s				
cpu Q_c^{dCATCH}	3.4e+00s	2.3e+01s	6.5e+01s	1.5e+02s	3.7e+02s
cpu Q_c^{bu}	2.5e-01s	8.7e-01s	2.6e+00s	9.5e+00s	6.7e+01s
speed-up	13.8	26.6	25.0	15.7	5.6
mom. resid. Q_c^{dCATCH}	1.1e-11	⋆ 1.2e-05	1.1e-11	⋆ 5.6e-05	⋆ 7.3e-05
mom. resid. Q_c^{bu}					
iter. 1	2.08e-16	9.41e-02	4.99e-01	1.51e+00	1.78e+00
iter. 2		1.32e-15	2.20e-15	4.72e-15	8.30e-02
iter. 3					7.32e-15

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.
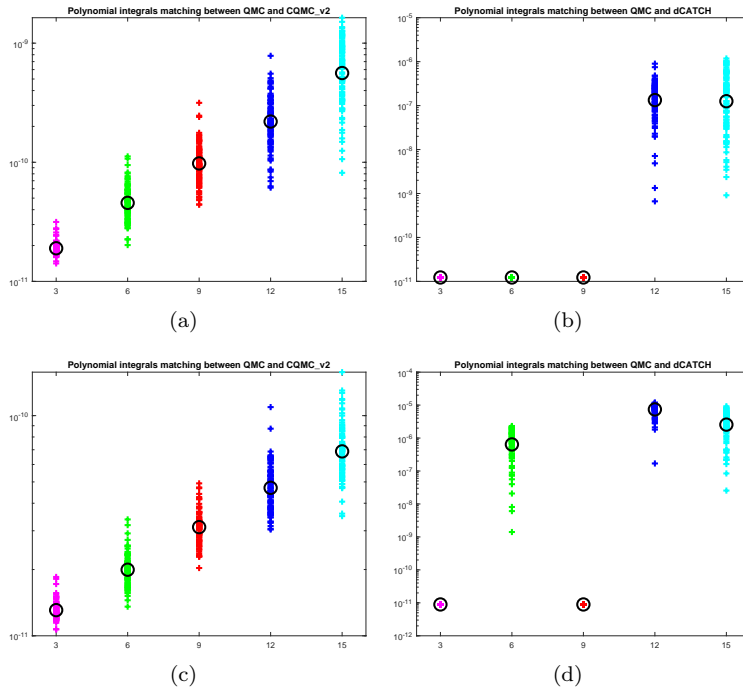
Figure 2: QMC compression relative errors and their logarithmic average (circles) over 100 trials of random polynomials for the bottom-up algorithm (left) and $dCATCH$ (right) on the union of 3 balls (top) and 100 balls (bottom). Note that the scales of the left and right figure are different.

10

deg	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
$E^{bu}(f_1)$	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
$E^{bu}(f_2)$	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	2.1e-03	1.1e-04
$E^{bu}(f_3)$	5.8e-01	2.8e-02	1.5e-02	9.5e-04	2.5e-05

deg	3	6	9	12	15
$E^{QMC}(f_1)$			1.1e-04		
$E^{dCATCH}(f_1)$	8.3e-02	8.8e-05	1.1e-04	1.1e-04	1.1e-04
$E^{bu}(f_1)$	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
$E^{bu}(f_2)$	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
$E^{bu}(f_3)$	6.1e-03	3.6e-03	1.2e-04	2.3e-04	2.2e-04

Table 2: Errors of compressed QMC volume integration on the union of 3 balls (top) and 100 balls (bottom); the reference values are computed via QMC starting from 10^8 Halton points in the bounding box.

3.2 Surface integration on union of balls

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

$$\Omega = \partial \bigcup_{j=1}^{s} B(C_j, r_j) = \bigcup_{j=1}^{s} \partial B(C_j, r_j) \backslash \bigcup_{j=1}^{s} \overset{\circ}{B}(C_j, r_j) , \qquad (18)$$

i.e. the set of all points lying on some sphere $\partial B(C_j, r_j)$, $j = 1, \ldots, s$, 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 balls in Section 3.1. Notice that Ω is a subset of an algebraic surface, i.e. the union of the corresponding spheres. Though the polynomial spaces dimension could be computed theoretically by algebraic geometry methods (cf., e.g., [6]), 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 Remark 1, 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

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

which preserves also the low-discrepancy property. The points are finally ordered by picking alternatively one point per active portion of the surface of the 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 Figure 1.

In Table 3 we report for this surface integration examples the same quantities appearing in Table 1 for the volume integration, where we use again the $dCATCH$ code in [12] 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 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 $m_{j+1} = 2m_j$, $j \geq 1$ with $m_1 = 2N$ and $\varepsilon = 10^{-10}$.

Again we get impressive compression ratios, and speed-ups varying from about 5 to more than 16. Moreover, the bottom-up 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 logarithmic average errors concerning surface integration of the random polynomial (16), restricted to the boundary of the union, are plotted in Figure 3. In Table 4 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 QMC error.

References

[1] D. Avis, B.K. Bhattacharya, H. Imai, Computing the volume of the union of spheres, The Visual Computer 3 (1988) 323–328.

[2] J.P. Bardhan, M.D. Altman, D.J. Willis, S.M. Lippow, B.Tidor, J.K. White, Numerical integration techniques for curved-element discretizations of molecule-solvent interfaces, J. Chem. Phys. 127 (2007), 014701.

[3] B. Bauman, A. Sommariva, M. Vianello, Compressed cubature over polygons with applications to optical design, J. Comput. Appl. Math. 370 (2020).

[4] B. Bauman, H. Xiao, Gaussian quadrature for optical design with noncircular pupils and fields, and broad wavelength range, Proc. SPIE 7652, 2010.

[5] L. Bittante, S. De Marchi, G. Elefante, A new quasi-Monte Carlo technique based on nonnegative least-squares and approximate Fekete points, Numer. Math. Theory Methods Appl. 9 (2016), 640–663.
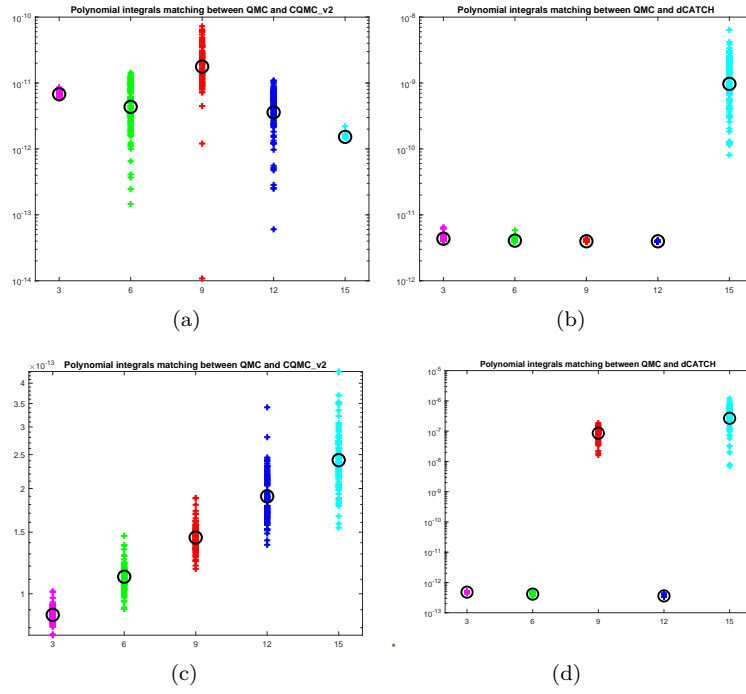
Figure 3: QMC compression relative errors and their logarithmic average (circles) over 100 trials for the bottom-up algorithm (left) and $dCATCH$ (right) of random polynomials on the surface of the union of 3 balls (top) and 100 balls (bottom). Note that the scales of the left and right figure are different.

13

deg	3	6	9	12	15
card. QMC	$M = 1{,}024{,}179$				
card. Q_c^{dCATCH}	20	83	200	371	572
card. Q_c^{bu}	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.	8.8e-01s				
cpu Q_c^{dCATCH}	2.8e+00s	1.7e+01s	5.0e+01s	1.4e+02s	3.2e+02s
cpu Q_c^{bu}	3.1e-01s	1.1e+00s	2.7e+00s	5.8e+00s	6.5e+01s
speed-up	9.0	15.9	18.2	24.0	4.9
mom. resid. Q_c^{dCATCH}	8.6e-12	8.9e-12	8.9e-12	8.9e-12	⋆ 1.6e-06
mom. resid. Q_c^{bu}					
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

deg	3	6	9	12	15
card. QMC	$M = 1{,}032{,}718$				
card. Q_c^{dCATCH}	20	84	219	455	807
card. Q_c^{bu}	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.	1.5e+01s				
cpu Q_c^{dCATCH}	2.8e+00s	1.6e+01s	4.3e+01s	1.1e+02s	2.4e+02s
cpu Q_c^{bu}	3.2e-01s	1.1e+00s	3.0e+00s	6.8e+00s	2.4e+01s
speed-up	8.7	14.5	14.3	16.2	9.9
mom. resid. Q_c^{dCATCH}	9.0e-13	9.1e-13	⋆ 3.2e-06	9.3e-13	⋆ 1.8e-05
mom. resid. Q_c^{bu}					
iter. 1	2.09e+00	1.22e+00	5.53e-01	6.18e-01	1.41e-01
iter. 2	7.49e-16	1.30e-15	2.52e-15	5.16e-15	1.18e-14

Table 3: 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).

[6] D.A. Cox, J. Little, D. O'Shea, Ideals, varieties, and algorithms, 4th edition, Springer, 2015.

[7] P.J. Davis, A construction of nonnegative approximate quadratures, Math. Comp. 21 (1967) 578–582.

[8] F. Dell'Accio, A. Sommariva, M. Vianello, Random sampling and unisolvent interpolation by almost everywhere analytic functions, Appl. Math. Lett. 145 (2023).

[9] M. Dell'Orto, M. Dessole, F. Marcuzzi, The Lawson-Hanson Algorithm with Deviation Maximization: Finite Convergence and Sparse Recovery, Numer. Linear Algebra Appl. 30 (2023), e2490.

[10] S. De Marchi, G. Elefante, Quasi-Monte Carlo integration on manifolds with mapped low-discrepancy points and greedy minimal Riesz s-energy points, Appl. Numer. Math. 127 (2018) 110–124.

[11] M. Dessole, F. Marcuzzi, Deviation maximization for rank-revealing QR factorizations, Numer. Algorithms 91 (2022) 1047-1079.

[12] M. Dessole, F. Marcuzzi, M. Vianello, Accelerating the Lawson-Hanson NNLS solver for large-scale Tchakaloff regression designs, Dolomites Res. Notes Approx. DRNA 13 (2020), 20–29.

deg	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
$E^{bu}(f_1)$	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
$E^{bu}(f_2)$	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
$E^{bu}(f_3)$	2.5e-02	3.8e-03	4.7e-06	9.1e-05	6.0e-06

deg	3	6	9	12	15
$E^{QMC}(f_1)$	4.0e-05				
$E^{dCATCH}(f_1)$	2.9e-03	2.8e-05	3.9e-05	4.0e-05	4.0e-05
$E^{bu}(f_1)$	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
$E^{bu}(f_2)$	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
$E^{bu}(f_3)$	1.6e-02	1.7e-03	2.6e-04	1.5e-04	1.6e-04

Table 4: Errors of compressed QMC surface integration on the union of 3 balls (top) and 100 balls (bottom); the reference values are computed via QMC starting from 10^6 points on each sphere.

[13] M. Dessole, F. Marcuzzi, M. Vianello, dCATCH: a numerical package for d-variate near G-optimal Tchakaloff regression via fast NNLS, MDPI-Mathematics 8(7) (2020) - Special Issue "Numerical Methods".

[14] J. Dick and F. Pillichshammer, Digital Nets and Sequences. Discrepancy Theory and Quasi-Monte Carlo Integration, Cambridge University Press, Cambridge, 2010.

[15] X. Duan, C. Quan, B. Stamm, A boundary-partition-based Voronoi diagram of d-dimensional balls: definition, properties, and applications, Adv. Comput. Math. 46 (2020).

[16] G. Elefante, A. Sommariva, M. Vianello, CQMC: an improved code for low-dimensional Compressed Quasi-MonteCarlo cubature, Dolomites Res. Notes Approx. DRNA 15 (2022) 92–100.

[17] G. Elefante, A. Sommariva, M. Vianello, Qbubble: compressed QMC volume and surface integration on union of balls (in Matlab) https://www.math.unipd.it/~alvise/software.html.

[18] G. Elefante, A. Sommariva, M. Vianello, Qsurf: compressed QMC integration on parametric surfaces, to appear in J. Approx. Software (https://ojs.unito.it/index.php/JAS).

[19] S. Hayakawa, Monte Carlo cubature construction, Jpn. J. Ind. Appl. Math. 38 (2021) 561–577.

[20] S. Hayakawa, K. Tanaka, Monte Carlo construction of cubature on Wiener space, Jpn. J. Ind. Appl. Math. (2022), 543–571.

[21] I.J. Kim, H. Na, An efficient algorithm calculating common solvent accessible volume, PLoS ONE 17(3): e0265614 (2022).

[22] C.L. Lawson, R.J. Hanson, Solving least squares problems. Classics in Applied Mathematics 15, SIAM, Philadelphia, 1995.

[23] G. Legrain, Non-Negative Moment Fitting Quadrature Rules for Fictitious Domain Methods, Comput. Math. Appl. 99 (2021) 270–291.

[24] F. Librino, M. Levorato, M. Zorzi, An algorithmic solution for computing circle intersection areas and its applications to wireless communications, Wireless Communications and Mobile Computing 14 (2014) 1672–1690.

[25] C. Litterer, T. Lyons, High order recombination and an application to cubature on Wiener space, Ann. Appl. Probab. 22 (2012) 1301–1327.

[26] M. Petitjean, Spheres Unions and Intersections and Some of their Applications in Molecular Modeling, in: A. Mucherino & al., Eds., Distance Geometry: Theory, Methods, and Applications, Springer New York, 2013, pp. 61–83.

[27] F. Piazzon, A. Sommariva, M. Vianello, Caratheodory-Tchakaloff Subsampling, Dolomites Res. Notes Approx. DRNA 10 (2017) 5–14.

[28] W. Plésniak, Multivariate Jackson Inequality, J. Comput. Appl. Math. 233 (2009) 815–820.

[29] C. Quan, B. Stamm, Mathematical analysis and calculation of molecular surfaces, J. Comput. Phys. 322 (2016), 760–782.

[30] D.L. Ragozin, Constructive Polynomial Approximation on Spheres and Projective Spaces, Trans. Amer. Math. Soc. 162 (1971) 157–170.

[31] M. Slawski, Non-negative least squares: comparison of algorithms, `https://sites.google.com/site/slawskimartin`.

[32] A. Sommariva, M. Vianello, Compression of multivariate discrete measures and applications, Numer. Funct. Anal. Optim. 36 (2015) 1198–1223.

[33] A. Sommariva, M. Vianello, Numerical quadrature on the intersection of planar disks, FILOMAT 31 (2017) 4105–4115.

[34] A. Sommariva, M. Vianello, Computing Tchakaloff-like cubature rules on spline curvilinear polygons, Dolomites Res. Notes Approx. DRNA 14 (2021) 1–11.

[35] A. Sommariva, M. Vianello, Cubature rules with positive weights on union of disks, Dolomites Res. Notes Approx. DRNA 15 (2022) 73–81 (Special Issue for the 60th of S. De Marchi).

[36] A. Sommariva, M. Vianello, Low-cardinality Positive Interior cubature on NURBS-shaped domains, BIT Numer. Math. 63 (2023).

[37] V. Tchakaloff, Formules de cubatures mécaniques à coefficients non négatifs, (French), Bull. Sci. Math. 81 (1957) 123–134.

[38] M. Tchernychova, Caratheodory cubature measures. Ph.D. dissertation in Mathematics (supervisor: T. Lyons), University of Oxford, 2015.

[39] D.R. Wilhelmsen, A Nearest Point Algorithm for Convex Polyhedral Cones and Applications to Positive Linear approximation, Math. Comp. 30 (1976) 48–57.