Sparse grid quadrature on products of spheres

Markus Hegland · Paul Leopardi

Abstract We examine sparse grid quadrature on weighted tensor products (WTP) of reproducing kernel Hilbert spaces on products of the unit sphere $\mathbb{S}^2$, in the case of worst case quadrature error for rules with arbitrary quadrature weights. We describe a dimension adaptive quadrature algorithm based on an algorithm of Hegland [12], and also formulate an adaptation of Wasilkowski and Woźniakowski's WTP algorithm [25], here called the WW algorithm. We prove that the dimension adaptive algorithm is optimal in the sense of Dantzig [5] and therefore no greater in cost than the WW algorithm. Both algorithms therefore have the optimal asymptotic rate of convergence of quadrature error given by Theorem 3 of Wasilkowski and Woźniakowski [25]. A numerical example shows that, even though the asymptotic convergence rate is optimal, if the dimension weights decay slowly enough, and the dimensionality of the problem is large enough, the initial convergence of the dimension adaptive algorithm can be slow.

Keywords reproducing kernel Hilbert spaces, quadrature, tractability, sparse grids, knapsack problems, spherical designs

1 Introduction

This paper examines sparse grid quadrature on weighted tensor products of reproducing kernel Hilbert spaces (RKHS) of real valued functions on the unit sphere $\mathbb{S}^2 \subset \mathbb{R}^3$, in the case of worst case quadrature error for rules with arbitrary quadrature weights. As per our previous paper on sparse grid quadrature on the torus [13], the rates of convergence of the quadrature rules constructed here are examined using the theory of Wasilkowski and Woźniakowski [25].

Mathematical Sciences Institute, Australian National University.
E-mail: paul.leopardi@gmail.com

The setting is the same as that used by Kuo and Sloan [19], and Hesse, Kuo and Sloan [15] to examine quasi-Monte Carlo (QMC) quadrature on products of the sphere $\mathbb{S}^2$, except that here we examine quadrature with arbitrary weights.

Quadrature on products of spheres is interesting, not just for purely theoretical reasons, but also for practical reasons. Integration over products of the unit sphere is equivalent to multiple integration over the unit sphere. Such multiple integrals can be approximated in a number of ways, including Monte Carlo methods. Applications of tensor product spaces on spheres and approximate integration over products of spheres include quantum mechanics [27], and transport and multiple scattering problems in various topic areas, including acoustics [21], optical scattering problems [1,18,23], and neutron transport problems [24]. One prototypical problem to be solved is scattering by a sequence of spheres. This can be modelled using a multiple integral of a function on the product of the spheres. The decay in the weights of successive spheres could model the decreasing influence of scattering on each successive sphere, as opposed to just cutting off the calculation after an arbitrary number of scatterings.

Quadrature with arbitrary weights, as opposed to equal weight quadrature, is interesting because, given the same set of quadrature points, optimal quadrature weights give a quadrature error at least as good as that of equal weight quadrature. This can result in the quadrature error converging more quickly to zero. This is illustrated by a numerical example presented at the conference on Monte Carlo and Quasi-Monte Carlo methods in Warsaw in 2010 [14].
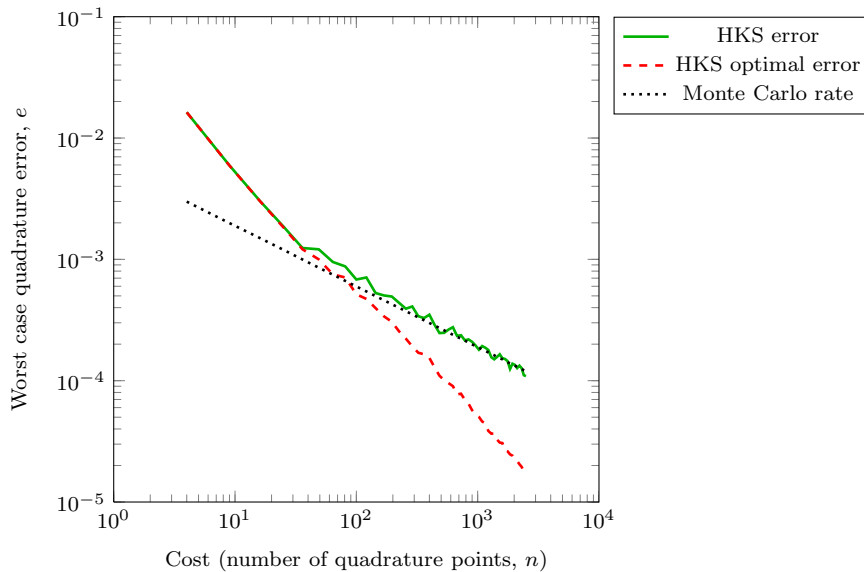

Fig. 1 Error of HKS rule vs HKS optimal rule for $(\mathbb{S}^2)^8$, $r = 3$, $\gamma_{8,k} = 0.1^k$.

Figure 1, based on slide 25 of the presentation, compares the performance of two quadrature rules, both based on the rule of Hesse, Kuo and Sloan [15], for a specific case of quadrature on $(\mathbb{S}^2)^8$. (The parameters r and γ mentioned in the caption are described in Section 2 below.) The two rules use the same sequence of point sets, differing only in the quadrature weights. The first rule, whose curve is labelled "HKS error", uses QMC weights. Its error initially converges to zero rapidly, then begins to converge at the Monte Carlo error rate. The second rule, labelled "HKS optimal error", uses optimal weights. Its error continues to converge to zero rapidly.

As noted in our previous paper [13], rates of convergence and criteria for strong tractability of quadrature with arbitrary weights are known in the case of weighted Korobov spaces on the unit torus [16,22]. As far as we know, this paper is the first to examine the analogous questions for quadrature with arbitrary weights on the corresponding spaces on products of spheres.

The paper by Kuo and Sloan [19] gives the criteria for tractability and strong tractability for QMC quadrature on products of spheres in the worst case setting. Since, for a given finite set of points, quadrature with optimal weights yields an error no larger than that of QMC quadrature, the quadrature problem for arbitrary weights in the worst case is tractable whenever the corresponding QMC quadrature problem is tractable. The relevant definitions and criteria for tractability and strong tractability for our setting are discussed in greater detail in Section 2.

The algorithms we examine are an adaptation of the algorithm used in our previous paper [12,13], and an adaptation of the WTP algorithm of Wasilkowski and Woźniakowski [25]. We examine these algorithms theoretically, giving bounds for the asymptotic convergence rate of quadrature error in the worst case. We also examine the performance of these algorithms in practice, via a small number of numerical examples.

The main results of this paper are:

1. (Theorem 2)
   Under the conditions described in the theorem, our dimension adaptive (DA) algorithm is optimal in the sense of Dantzig [5, Figure 3, p. 274] and therefore no greater in cost than the adapted Wasilkowski and Woźniakowski (WW) algorithm.
2. (Theorem 8, Corollary 9 and Theorem 10).
   The cost of the adapted WW algorithm in our setting is bounded such that, whenever the problem is strongly tractable, the asymptotic rate of convergence of the worst case quadrature error is essentially the same as that for the one dimensional problem.
3. (Numerical examples).
   Our numerical examples use exponentially decreasing sequences of dimension weights, and a finite sequence of spherical designs as listed in Table 1. This sequence of spherical designs yields a sequence of quadrature points on a single sphere having all of the properties needed by Theorem 8.

Since the problem is strongly tractable in this case, if the sequence of spherical designs could be extended indefinitely, then the adapted WW algorithm would have an asymptotic rate of convergence of the worst case quadrature error essentially the same as that for the one dimensional problem. As a result of Theorem 11 the DA algorithm would also have this same asymptotic convergence rate in this case.

4. (Figure 6).
   If the dimension weights decay slowly enough (e.g. $\gamma_{d,k} = 0.9^d$), and the dimensionality is high enough (e.g. $d \geqslant 16$) the initial rate of convergence of the DA algorithm can be slow. Specifically, for $\gamma_{d,k} = 0.9^d$ and $d = 16$, the DA algorithm needs more than $100\,000$ function evaluations to reduce the worst case quadrature error from 1 to 0.1.

The remainder of this paper is organized as follows. Section 2 describes the setting in detail, and includes a discussion on the tractability of the problem. Section 3 describes the optimization problem involved in dimension adaptive sparse grid quadrature. Section 4 introduces the DA algorithm and shows that it is optimal in the sense of Dantzig [5]. Section 5 analyses a version of the WTP algorithm of Wasilkowski and Woźniakowski, derives bounds for the asymptotic rate of convergence of its worst case quadrature error to zero, and applies these bounds to the DA algorithm. Section 6 contains numerical results, comparing implementations of the two algorithms, and showing how the DA algorithm performs as the dimension is increased. Appendix A contains a proof of Theorem 8.

2 Setting

The setting used here is a special case of a general setting that also applies to our previous paper [13].

Let $\mathcal{D} \subset \mathbb{R}^{s+1}$ be a compact manifold with probability measure μ. It follows that the constant function $\mathbf{1}$, with $\mathbf{1}(x) = 1$ for all $x \in \mathcal{D}$, is integrable and $\int_{\mathcal{D}} \mathbf{1}(x)\,d\mu(x) = 1$. Then let H be a Hilbert space of functions $f : \mathcal{D} \to \mathbb{R}$, with inner product $\langle \cdot, \cdot \rangle_H$, and kernel K, with the following properties.

1. For every $x \in \mathcal{D}$, the function $k_x \in H$, given by $k_x(y) := K(x, y)$, satisfies

$$f(x) = \langle k_x, f \rangle_H, \quad \text{for all } f \in H; \tag{1}$$

2. Every $f \in H$ is integrable, and the constant function $\mathbf{1}$ is in H, such that

$$\int_{\mathcal{D}} f(x)\,d\mu(x) = \langle \mathbf{1}, f \rangle_H. \tag{2}$$

We recognize H as a reproducing kernel Hilbert space (RKHS). In this framework, a quadrature rule Q, defined by

$$Q(f) := \sum_{i=1}^{n} w_i f(x_i) \tag{3}$$

is a continuous linear functional and $Q(f) = \langle q, f \rangle_H$ with representer

$$q = \sum_{i=1}^{n} w_i k_{x_i}, \tag{4}$$

such that $Q(f) = \langle q, f \rangle_H$. Here we have used (1) and the Riesz representation theorem for H. In the remainder of this paper, we also refer to a function $q \in H$ of the form (4) as a quadrature rule, with the understanding that q represents the linear functional Q of the form (3).

Given the quadrature points x_i, an optimal choice of weights w_i minimizes the worst case quadrature error $e(q)$, which is

$$e(q) := \sup_{f \in H, \|f\| \leqslant 1} |\langle \mathbf{1}, f \rangle_H - \langle q, f \rangle_H| = \|\mathbf{1} - q\|_H. \tag{5}$$

The optimal q^* is thus defined as

$$q^* := \operatorname{argmin}_q \{ \|\mathbf{1} - q\|_H \mid q \in \operatorname{span}\{k_{x_1}, \ldots, k_{x_n}\} \}.$$

The weights of an optimal quadrature rule are thus obtained by solving a linear system of equations with a matrix whose elements are the values of the reproducing kernel $K(x_i, x_j) = \langle k_{x_i}, k_{x_j} \rangle_H$. The right-hand side of these equations is a vector with elements all equal to one.

We now describe an auxiliary reproducing kernel Hilbert space $\mathcal{H}$ of functions on $\mathcal{D}$. The space $\mathcal{H}$ has a kernel $\mathcal{K}$ satisfying (1), but instead of (2), this auxiliary space satisfies

$$\int_{\mathcal{D}} f(x) \, d\mu(x) = 0, \quad \text{for all } f \in \mathcal{H}.$$

Thus the function $\mathbf{1}$ is not an element of this space.

We now extend $\mathcal{H}$ into the space $\mathcal{H}^\gamma$, which consists of all functions of the form $g = a\mathbf{1} + f$, where $a \in \mathbb{R}$, and $f \in \mathcal{H}$, with the norm $\|\cdot\|_{\mathcal{H}^\gamma}$ defined by

$$\|g\|_{\mathcal{H}^\gamma}^2 = |a|^2 + \frac{1}{\gamma} \|f\|_{\mathcal{H}}^2.$$

It is easily verified that $\mathcal{H}^\gamma$ is an RKHS with reproducing kernel

$$\mathcal{K}_\gamma(x, y) = 1 + \gamma \mathcal{K}(x, y),$$

where $\mathcal{K}$ is the reproducing kernel of $\mathcal{H}$. In particular, the space $\mathcal{H}^\gamma$ with kernel $\mathcal{K}_\gamma$ contains the function $\mathbf{1}$ and satisfies both properties (1) and (2).

For functions on the domain $\mathcal{D}^d$ we consider the tensor product space $\mathcal{H}_d := \bigotimes_{k=1}^{d} \mathcal{H}^{\gamma_k}$ where $1 \geqslant \gamma_1 \cdots \geqslant \gamma_d \geqslant 0$. This is an RKHS of functions on $\mathcal{D}^d$ with reproducing kernel $\mathcal{K}_d(x, y) := \prod_{k=1}^{d} (1 + \gamma_k \mathcal{K}(x_k, y_k))$ where $x_k, y_k \in \mathcal{D}$ are the components of $x, y \in \mathcal{D}^d$. Also

$$\int_{\mathcal{D}^d} f(x) \, d\mu_d(x) = \langle \mathbf{1}, f \rangle_{\mathcal{H}_d},$$

where μ_d is the product measure, $\langle \cdot, \cdot \rangle_{\mathcal{H}_d}$ is the scalar product on the tensor product space $\mathcal{H}_d$, and $\mathbf{1}$ is the constant function on $\mathcal{D}^d$ with value 1.

The specific setting for this paper is that of Kuo and Sloan [19], with $s := 2$, except that we allow quadrature rules with arbitrary weights. We now describe this setting. We take our domain $\mathcal{D}$ to be the unit sphere $\mathbb{S}^2 := \{x \in \mathbb{R}^3 \mid x_1^2 + x_2^2 + x_3^2 = 1\}$, and consider the real space $L_2(\mathbb{S}^2)$ with respect to the uniform probability measure μ on $\mathbb{S}^2$. We use the orthonormal basis of real spherical harmonics $Y_{\ell,m}(x)$, $\ell = 0, \ldots, \infty$, $m = 1, \ldots, 2\ell + 1$, as specified by Hesse, Kuo and Sloan [15, Section 3.1]. As per [15, (5)], the addition theorem for spherical harmonics with this normalization yields

$$\sum_{m=1}^{2\ell+1} Y_{\ell,m}(x)\, Y_{\ell,m}(y) = (2\ell + 1) P_\ell(x \cdot y)$$

for all $x, y \in \mathbb{S}^2$, where P_ℓ is the Legendre polynomial of degree ℓ.

For any function $f \in L_2(\mathbb{S}^2)$, we expand f in the Fourier series

$$f(x) = \hat{f}_{0,0} + \sum_{\ell=1}^{\infty} \sum_{m=1}^{2\ell+1} \hat{f}_{\ell,m} Y_{\ell,m}(x).$$

For a positive dimension weight γ, we define the RKHS

$$\mathbb{H}_{1,\gamma}^{(r)} := \{f : \mathbb{S}^2 \to \mathbb{R} \mid \|f\|_{\mathbb{H}_{1,\gamma}^{(r)}} < \infty\},$$

where

$$\langle f, g \rangle_{\mathbb{H}_{1,\gamma}^{(r)}} := \hat{f}_{0,0}\, \hat{g}_{0,0} + \gamma^{-1} \sum_{\ell=1}^{\infty} \sum_{m=1}^{2\ell+1} \left(\ell(\ell+1)\right)^r \hat{f}_{\ell,m}\, \hat{g}_{\ell,m}.$$

Kuo and Sloan [19] show that the reproducing kernel of $\mathbb{H}_{1,\gamma}^{(r)}$ is

$$K_{1,\gamma}^{(r)}(x, y) := 1 + \gamma A_r(x \cdot y), \quad \text{where for } z \in [-1, 1],$$

$$A_r(z) := \sum_{\ell=1}^{\infty} \frac{2\ell + 1}{\left(\ell(\ell+1)\right)^r} P_\ell(z).$$

The sum defining A_r converges when $r > 3/2$.

For $\gamma := (\gamma_{d,1}, \ldots, \gamma_{d,d})$, we now define the tensor product space

$$\mathbb{H}_{d,\gamma}^{(r)} := \bigotimes_{k=1}^{d} \mathbb{H}_{1,\gamma_{d,k}}^{(r)}.$$

This is a weighted RKHS on $(\mathbb{S}^2)^d$, with reproducing kernel

$$K_{d,\gamma}^{(r)}(x, y) := \prod_{k=1}^{d} K_{1,\gamma_{d,k}}^{(r)}(x_k, y_k).$$

Kuo and Sloan [19] studied equal weight (QMC) quadrature on the space $\mathbb{H}_{d,\gamma}^{(r)}$, and found that it is strongly tractable if and only if $\sum_{k=1}^{d} \gamma_{d,k} < \infty$ as $d \to \infty$. The definition of strong tractability used by Kuo and Sloan [19] is specific to QMC quadrature. Here we expand the definition to quadrature with arbitrary weights.

We suppose for given $n > 0$ we can find points $\{x_1, \ldots, x_n\} \subset (\mathbb{S}^2)^d$ and corresponding quadrature weights, defining the quadrature rule $Q_{n,d}$ on $\mathbb{H}_{d,\gamma}^{(r)}$, with worst case error $e_{n,d} := e(Q_{n,d})$. We also define the quadrature rule $Q_{0,d} := 0$ so that the corresponding worst case quadrature error is $e_{0,d} = \|\mathbf{1}\|_{\mathbb{H}_{d,\gamma}^{(r)}} = 1$.

For $\varepsilon \in (0,1)$ we want to find the smallest cost in terms of the number of function evaluations, that is the smallest $n = \text{cost}(d, \varepsilon)$ for which points $\{x_1, \ldots, x_n\} \subset (\mathbb{S}^2)^d$ exist such that $e_{n,d} \leqslant \varepsilon\, e_{0,d}$. The integration problem for arbitrary quadrature weights in the worst-case setting is said to be *strongly tractable* in the space $\mathbb{H}_{d,\gamma}^{(r)}$ if

$$\text{cost}(d, \varepsilon) \leqslant C\varepsilon^{-p}, \tag{6}$$

where C and p are non-negative constants independent of ε and d. If (6) holds then the infimum of p is called the ε-exponent of strong tractability.

Since, for a given finite set of points, quadrature with optimal weights yields an error no larger than that of QMC quadrature, strong tractability for arbitrary weight quadrature rules holds for $\mathbb{H}_{d,\gamma}^{(r)}$ whenever strong QMC tractability holds. In particular, as a consequence of a theorem of Kuo and Sloan [19, Theorem 4], strong tractability for arbitrary weight quadrature rules holds for positive dimension weights $\gamma_{d,k}$ when

$$\limsup_{d \to \infty} \sum_{k=1}^{d} \gamma_{d,k} < \infty.$$

Hesse, Kuo and Sloan [15] go on to construct sequences of QMC rules on the space $\mathbb{H}_{d,\gamma}^{(r)}$, and prove that their worst case error converges at least as quickly as the Monte Carlo error rate of $\text{O}(n^{-1/2})$, where n is the cost of the quadrature rule in terms of the number of points.

The work of Hickernell and Woźniakowski [16], and Sloan and Woźniakowski [22], on the weighted Korobov space of periodic functions on the unit cube, and the work of Wasilkowski and Woźniakowski [25] on WTP quadrature on non-periodic functions on the unit cube, combined with the observations above on strong tractability, suggests bounds on the worst case error for the case of quadrature with arbitrary weights on the space $\mathbb{H}_{d,\gamma}^{(r)}$. In the case of exponentially decreasing weights, as studied here, one might expect that for $r > 3/2$, given $\delta > 0$, there exist points on $(\mathbb{S}^2)^d$ yielding an optimal weight quadrature rule whose worst-case error would have an upper bound independent of d, of order $\text{O}(n^{-r/2+\delta})$. The specific form of this upper bound is suggested by the analysis of Sloan and Woźniakowski of QMC rules on the torus

[22, Theorem 3]. The analysis in Sections 4, 5 and 6 below shows that the DA algorithm satisfies the upper bound suggested here, given the sequences of point sets on $\mathbb{S}^2$ used in the numerical examples of Section 6.

3 Optimization Problem

We first describe the optimization problem in the general RKHS setting, as given in Section 2.

Assume that a sequence of distinct quadrature points $x_1, x_2, \ldots \in \mathcal{D}$, and a sequence of positive integers $n_0 < n_1 < \ldots$ are given and are the same for all spaces $\mathcal{H}^\gamma$. The quadrature rules for $\mathcal{H}^\gamma$ are then defined as some element of $V_j^\gamma := \operatorname{span}\{k_{x_1}^\gamma, \ldots, k_{x_{n_j}}^\gamma\} \subset \mathcal{H}^\gamma$. Note that $V_j^\gamma \subset V_{j+1}^\gamma$ for all $j \geqslant 0$.

Denote the optimal rule in V_j^γ by q_j^γ. Now define the pair-wise orthogonal spaces U_j^γ by $U_0^\gamma := V_0^\gamma$, and by the orthogonal decomposition $V_{j+1}^\gamma = V_j^\gamma \oplus U_{j+1}^\gamma$. Using the fact that the q_j^γ are optimal, it follows that

$$\delta_{j+1}^\gamma := q_{j+1}^\gamma - q_j^\gamma \in U_{j+1}^\gamma$$

and $\delta_0 := q_0^\gamma \in U_0^\gamma = V_0^\gamma$. Note that, while the dimension of U_{j+1}^γ is $n_{j+1} - n_j$, independently of γ,

$$U_{j+1}^\gamma \neq \operatorname{span}\{k_{x_{1+n_j}}^\gamma, \ldots, k_{x_{n_{j+1}}}^\gamma\}, \tag{7}$$

since the functions k_{x_m} and k_{x_n} for $m \neq n$ are, in general, not orthogonal to each other.

We use the notation $\mathbb{J} := \mathbb{N}^d$, treating elements of $\mathbb{J}$ as indices, with a partial order such that for $i, j \in \mathbb{J}$, $i \leqslant j$ if and only if $i_h \leqslant j_h$ for all components.

For a index $i \in \mathbb{J}$, let $\downarrow i$ denote the *down-set* of i, defined by [6, p. 13]

$$\downarrow i := \{j \in \mathbb{J} \mid j \leqslant i\}.$$

Subsets of $\mathbb{J}$ are partially ordered by set inclusion. For a subset $I \subset \mathbb{J}$, let $\downarrow I$ denote the down-set of I, defined by $\downarrow I := \bigcup_{i \in I} \downarrow i$. Then $\downarrow I$ is the smallest set $J \supseteq I$ such that if $i \in J$ and $j \leqslant i$ then $j \in J$. Thus $\downarrow\downarrow I = \downarrow I$.

A sparse grid quadrature rule is then of the form

$$q \in V_I := \sum_{j \in I} \bigotimes_{k=1}^d V_{j_k}^{\gamma_{d,k}}$$

for some index set I. The orthogonal decomposition $V_j^\gamma = \bigoplus_{i=1}^j U_i^\gamma$ and the observation (7) yield the multidimensional orthogonal decomposition

$$V_I = \bigoplus_{j \in \downarrow I} \bigotimes_{k=1}^d U_{j_k}^{\gamma_{d,k}}.$$

A short derivation shows that an optimal $q \in V_I$ is given by

$$q_I^* = \sum_{j \in \downarrow I} \bigotimes_{k=1}^{d} \delta_{j_k}^{\gamma_{d,k}}.$$

Thus both V_I and q_I^* are obtained in terms of the down-set $\downarrow I$, effectively restricting the choice of the set I to index sets which are also down-sets. (This construction is similar to the general sparse grid construction for functions on the d-dimensional torus, examined by Gerstner and Griebel [8, Section 3.1].)

This leads us to defining the concepts of an *admissible* index set, and an *optimal* index set. An admissible index set I satisfies the *admissibility condition* (similar to that of Gerstner and Griebel [8, Section 3.1])

$$I = \downarrow I. \tag{8}$$

An optimal index set is one which minimizes the error for a given cost, or minimizes the cost for a given error. Here, the cost is the number of quadrature points, which is the dimension of V_I. (This is a similar concept of optimality to that mentioned by Griebel and Knapek in the context of approximation spaces [10, Section 5].)

We now make the definition of an optimal index set more precise. We first define $\nu_{j_k} := \dim U_{j_k}^{\gamma_{d,k}}$ and $\delta_{j_k}^{(k)} := \delta_{j_k}^{\gamma_{d,k}}$. For the remainder of this section, we use $\varepsilon \in (0,1)$ to denote the required upper bound on quadrature error. The optimization problem then uses the following definitions.

Definition 1 *For index $j \in \mathbb{J}$, define*

$$\nu_j := \prod_{k=1}^{d} \nu_{j_k}, \quad \Delta_j := \bigotimes_{k=1}^{d} \delta_{j_k}^{(k)}, \quad p_j := \|\Delta_j\|^2, \quad r_j := p_j/\nu_j.$$

For subset $I \subset \mathbb{J}$, define

$$\nu(I) := \sum_{j \in I} \nu_j, \quad p(I) := \sum_{j \in I} p_j.$$

Also, define $P := 1 - \varepsilon^2$.

Here, j_k is the kth component of the index j.

We use these definitions in the following sense. The quadrature rule q_I is given by a sum of incremental rules Δ_j, indexed by the multi-index j. The "profit" p_j for each incremental rule is its squared norm. The cost ν_j of each incremental rule is the number of extra points the incremental rule contributes to the overall quadrature rule, assuming that the admissibility condition (8) applies. Thus the cost $\nu(I)$ of a quadrature rule, as given by the number of

function evaluations, is just the total cost of the incremental rules. The ratio $r_j = p_j/\nu_j$ is called the *efficiency* of the incremental rule Δ_j.

Due to the properties of ν_{j_k} and Δ_j, ν and p satisfy

$$\nu_j, \nu(I) \in \mathbb{N}_+, \quad 0 < p_j < 1, \quad 0 < p(I) < 1, \quad p(\mathbb{J}) = 1. \tag{9}$$

We now consider the following optimization problem, posed as a minimization problem on the variable $I \subset \mathbb{J}$.

Optimization Problem 1

$$\text{Minimize} \quad \nu(\downarrow I), \quad \text{subject to} \quad p(I) \geqslant P,$$

for some $0 < P < 1$, where ν and p satisfy (9).

In other words, given a required upper bound ε on the quadrature error, the problem is to find the subset $I \subset \mathbb{J}$ with the smallest cost $\nu(\downarrow I) = \sum_{j \in \downarrow I} \nu_j$, satisfying the constraint $1 - \sum_{j \in I} p_j \leqslant \varepsilon^2$.

Optimization Problem 1 can have multiple solutions, since for $H, I, J \subset \mathbb{J}$ if $J = \downarrow H = \downarrow I$ and both $p(H) \geqslant P$ and $p(I) \geqslant P$ then both H and I are solutions to Optimization Problem 1. The following problem breaks this tie.

Optimization Problem 2

$$\text{Maximize} \quad p(I) \quad \text{subject to } I \text{ solving Optimization Problem 1.}$$

The solution of Optimization Problem 2 satisfies the admissibility condition (8):

Lemma 1 *If I is a solution of Optimization Problem 2, then $I = \downarrow I$.*

Proof Let $J = \downarrow I$, where I is a solution of Optimization Problem 2, and therefore of Optimization Problem 1. Then $I \subset J$ and thus J satisfies the constraints of Optimization Problem 1, since $p_i > 0$. Therefore J is also a solution of Optimization Problem 1, since $\nu(\downarrow J) = \nu(\downarrow I)$. If $I \subsetneq J$, it follows from $p_i > 0$ that $p(J) > p(I)$, and so I cannot be optimal. Therefore $I = J$. $\square$

In view of the admissibility condition (8), we reformulate Optimization Problem 2 as:

Optimization Problem 3

$$\text{Minimize} \quad \nu(I), \quad \text{subject to} \quad I = \downarrow I, \quad \text{and} \quad p(I) \geqslant P,$$

for some $0 < P < 1$, where ν and p satisfy (9).

As pointed out by (e.g.) Griebel and Knapek [10,11], some sparse grid problems can be formulated and solved as knapsack problems. The resulting solution is optimal in terms of total profit for a given cost.

We call Optimization Problem 3 a *down-set-constrained* binary knapsack problem. Each item in the knapsack is an incremental rule. The relationships between Optimization Problem 3 and other more well-known knapsack problems are described in more detail in Section 4.

4 Algorithm

The dimension adaptive algorithm

The dimension adaptive (DA) algorithm to choose the set I in the setting where $\mathcal{D}$ is $\mathbb{S}^1$ is described in [13]. The algorithm is quite general, and applies equally well to the current setting, where $\mathcal{D}$ is $\mathbb{S}^2$. We repeat the algorithm here as Algorithm 1, with some changes in notation. This is a greedy algorithm for Optimization Problem 3.

The detailed description of Algorithm 1 uses the following notation. Given the down-set $I \subset \mathbb{J}$, we define $M(I)$ to be the set of minimal elements of $\mathbb{J} \setminus I$, in other words,

$$M(I) := \big\{ i \in \mathbb{J} \setminus I \mid I \cup \{i\} = \downarrow(I \cup \{i\}) \big\}.$$

Given $j \in \mathbb{J}$, define $S(j)$, the *forward neighbourhood* of j, [8, p. 71] as

$$S(j) := \{ i \in \mathbb{J} \mid j < i \text{ and } (j \leqslant \ell < i \Rightarrow \ell = j) \},$$

that is, $S(j)$ is the set of minimal elements of $\{ i \in \mathbb{J} \mid j < i \}$.

Algorithm 1: The dimension adaptive (DA) algorithm.

Data: error ε, incremental rules Δ_j and their costs ν_j for $j \in \mathbb{J}$
Result: ε approximation q and index set I
$I := I_{(0)} := \{0\}$; $q := q_{(0)} := \Delta_0$;
while $\|\mathbf{1} - q\| > \varepsilon$ **do**
    $j^{(t+1)} := \text{argmax}_i \{ r_i \mid i \in M(I_{(t)}) \}$;
    $I := I_{(t+1)} := I_{(t)} \cup \{j^{(t+1)}\}$; $q := q_{(t+1)} := q + \Delta_{j^{(t+1)}}$;

To show how the minimal set $M(I)$ can be effectively determined at each step of Algorithm 1, we use the subscript (t) to keep track of the steps.

We treat Algorithm 1 as starting with the down-set $I_{(0)} := \{0\}$, the index $j^{(0)} := 0$ and the minimal set $M_{(0)} := \{0\}$. At the start of step $t+1$ of the algorithm, we have previously computed the down-set $I_{(t)}$, the index $j^{(t)}$ and the minimal set $M_{(t)}$, such that $j^{(t)} \in M_{(t)}$, and (for $t > 0$) $M_{(t)} = M(I_{(t-1)})$ and $I_{(t)} = I_{(t-1)} \cup \{j^{(t)}\}$. To construct $M_{(t+1)} = M(I_{(t)})$, set

$$S_{(t+1)} := \big\{ i \in S(j^{(t)}) \mid I_{(t)} \cup \{i\} = \downarrow(I_{(t)} \cup \{i\}) \big\}$$

and

$$M_{(t+1)} := \Big(M_{(t)} \setminus \{j^{(t)}\} \Big) \cup S_{(t+1)}.$$

Note that

$$\Big(M_{(t)} \setminus \{j^{(t)}\} \Big) \cup S_{(t+1)} = \big(M_{(t)} \cup S_{(t+1)} \big) \setminus \{j^{(t)}\}.$$

As every minimal element of $\mathbb{J} \setminus I_{(t)}$ is either an element of $M_{(t)}$ (but not $j^{(t)}$) or an element of $S_{(t+1)}$ we see that $M_{(t+1)}$ is equal to $M(I_{(t)})$.

Remark 1 The details of the algorithm of Gerstner and Griebel [8, Section 3.2] are similar to those of Algorithm 1. Their *old index set* $\mathcal{O}$ corresponds to our down-set I, and their *active index set* $\mathcal{A}$ corresponds to our minimal set $M(I)$.

Complexity of Algorithm 1

Here we briefly examine the computational effort involved in determining the down-set $I_{(t)}$ using Algorithm 1. As noted above, Algorithm 1 is similar to the algorithm of Gerstner and Griebel. Thus an analysis of complexity along the lines of their [8] Section 4.4 is appropriate.

At step $t+1$ of Algorithm 1, the index $j^{(t)}$ is removed from the minimal set $M_{(t)}$, the backward neighbourhood of each of d potential new indices is checked, and up to d new indices are inserted to form the updated set $M_{(t+1)}$. Thus the set $M_{(t)}$ contains at most $t(d-1)$ elements. (Given $i \in \mathbb{J}$, the *backward neighbourhood* $R(i)$ of i is the set of maximal elements of $\{\ell \in \mathbb{J} \mid \ell < i\}$ [8, p. 71].)

The checking of the backward neighbourhood of each index $i \in S(j^{(t)})$ involves looking up each of the $d-1$ indices in $R(i) \setminus \{j\}$ to see if it is already contained in the index set $I_{(t)}$. Assuming that the set $I_{(t)}$ is implemented using a hash table [9], the effort involved in this lookup is independent of the size of $I_{(t)}$ and is therefore approximately $O(d^2)$.
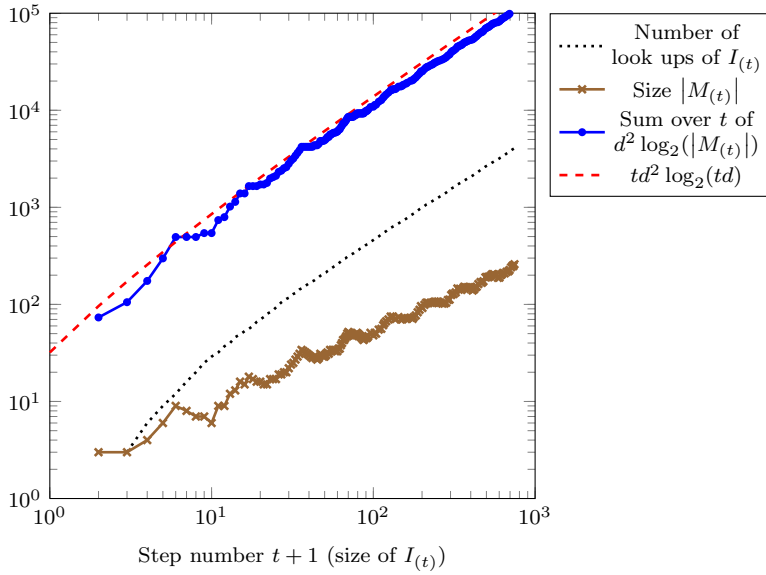

Fig. 2 Effort involved in calculating DA rules for $(\mathbb{S}^2)^4$, $r = 3$, $\gamma_{4,k} = 0.5^k$.

The index $j^{(t)}$ removed from $M_{(t)}$ is the index i with the maximum value of the efficiency r_i. Let us assume that $M_{(t)}$ is implemented as an implicit heap, or as a priority queue with the same time complexity for deletion of the largest element and for insertion of an element [17]. To find $j^{(t)}$ and remove it from $M_{(t)}$ takes $O(d \log_2(|M_{(t)}|)) = O(d \log(td))$ operations, such as comparison and index manipulation. Each insertion also takes at most $O(d \log(td))$ operations on $M_{(t)}$. Each insertion also requires the calculation of ν_i and p_i for each index i, each requiring effort $O(d)$, since each these are calculated via the product of one dimensional values. At each step, the total effort required is therefore at most $O(d^2 \log(td))$. The overall effort up to step $t+1$ is therefore at most $O(td^2 \log(td))$.

Figure 2 shows some components of the cumulative effort involved in calculating the DA rules for the numerical example of $(\mathbb{S}^2)^4$, $r = 3$, $\gamma_{4,k} = 0.5^k$, given the assumptions above. Figure 3 compares the cumulative effort for $M_{(t)}$ for the DA rules for the numerical example of $(\mathbb{S}^2)^d$, $r = 3$, $\gamma_{d,k} = 0.5^k$, for $d = 2, 4, 8, 16$, under the same assumptions. See also Figure 4 in Section 6, and the detailed description of the numerical examples given in that Section.

In general, the number of new indices added to the minimal set $M_{(t)}$ depends not only on the dimension d, but also on the dimension weights $\gamma_{d,k}$. Thus the dimension weights determine the set $M_{(t)}$ at each step. In particular, the size of $M_{(t)}$ may be much less than td. The dimension weights also determine the rate of convergence of the quadrature error $e(q_{(t)})$ to zero, via $\nu_{j^{(t)}}$ and $p_{j^{(t)}}$ for each index $j^{(t)}$, and hence determine the number of iterations t needed before the error is reduced to ε.

A fully detailed analysis of the computational effort of Algorithm 1, including an accurate estimate of $|M_{(t)}|$ as a function of d, $\gamma_{d,k}$, and t, is thus potentially very complicated, and is beyond the scope of this paper.

Optimality of Algorithm 1

Under certain conditions on ν and p, Algorithm 1 solves Optimization Problem 3. To see this, consider *monotonicity* with respect to the lattice partial ordering of $\mathbb{J}$.

Definition 2 *The function $p \in \mathbb{R}_+^{\mathbb{J}}$ is monotonically decreasing if $i < j$ implies that $p_i \geqslant p_j$. If $i < j$ implies that $p_i > p_j$, then $p \in \mathbb{R}_+^{\mathbb{J}}$ is strictly decreasing. The definitions of "monotonically increasing" and "strictly increasing" are similar.*

Using Definition 2, the following theorem holds.

Theorem 2 *If $p \in \mathbb{R}_+^{\mathbb{J}}$ is strictly decreasing and $\nu \in \mathbb{N}_+^{\mathbb{J}}$ is monotonically increasing, then Algorithm 1 yields a quadrature rule q and index set I such that I solves the down-set-constrained knapsack Optimization Problem 3, for $P = p(I) = \|\mathbf{1} - q\|^2$.*

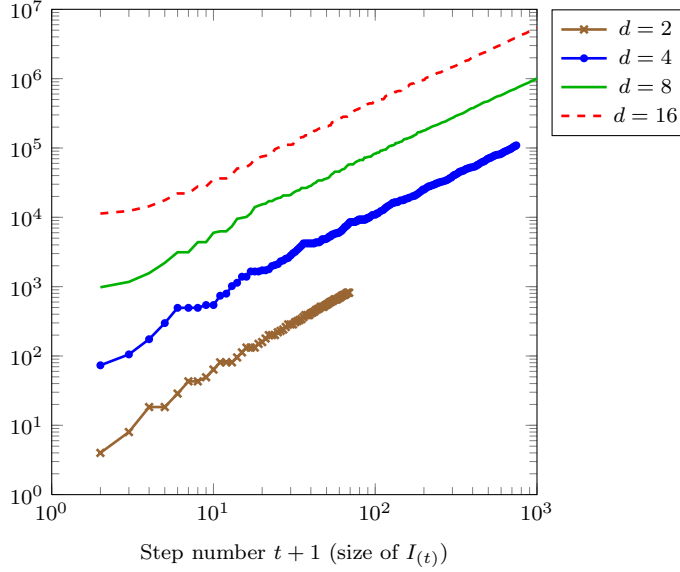The proof of Theorem 2 presented below proceeds in these stages.

Fig. 3 Sum over t of $d^2 \log_2(|M_{(t)}|)$ for DA rules for $(\mathbb{S}^2)^d$, $r = 3$, $\gamma_{d,k} = 0.5^k$.

1. We introduce a related binary knapsack problem, and show that if I is a solution of the binary knapsack problem, and I is also a down-set, then I is a solution of Optimization Problem 3.
2. We describe a greedy algorithm for the binary knapsack problem (Algorithm 2 below), and show that if the efficiency r_j is strictly decreasing, then each set I produced by the greedy algorithm is a solution of the binary knapsack problem, and is also a down-set, and therefore a solution of Optimization Problem 3.
3. We show that if the efficiency is strictly decreasing, then Algorithm 2 produces the same sequence of sets as Algorithm 1.

A binary (0/1) knapsack problem [5] related to Optimization Problem 3 is:

Optimization Problem 4

$$Minimize \quad \nu(I), \quad subject\ to \quad p(I) \geqslant P,$$

for some $0 < P < 1$, where ν and p satisfy (9).

Usually a binary knapsack problem is posed as a maximization problem, where the selection is from a finite set of items. Here we have a minimization problem and a countably infinite set. A finite minimization problem can always be posed as an equivalent maximization problem [20, p. 15]. In the case of Problem 4 this cannot be done, because the quantity to be maximized (the sum of the costs of the elements not in the knapsack) would be infinite. Instead, we must deal directly with the minimization form.

We now formulate a converse of Lemma 1.

Lemma 3 *If I is a solution of the Optimization Problem 4, and I also satisfies the admissibility condition $I = \downarrow I$, then I is a solution of Optimization Problem 3.*

Proof If I is a solution of the Optimization Problem 4, then $p(I) \geqslant P$. If I also satisfies the admissibility condition $I = \downarrow I$, then I satisfies the constraints of Optimization Problem 3 and thus minimizes p under these constraints, i.e., is a solution of Optimization Problem 3. $\square$

This justifies our calling Optimization Problem 3 a down-set-constrained knapsack problem.

If, in Optimization Problem 4 we identify each set $I \subset \mathbb{J}$ with its *indicator function* $\mathcal{I} \in \{0,1\}^{\mathbb{J}}$, where $\mathcal{I}_i = 1$ if and only if $i \in I$, we obtain a more usual formulation of the binary knapsack problem:

Optimization Problem 5

$$\text{Minimize} \quad \sum_{i \in \mathbb{J}} \nu_i \, \mathcal{I}_i, \quad \text{subject to} \quad \sum_{i \in \mathbb{J}} p_i \, \mathcal{I}_i \geqslant P, \quad \mathcal{I} \in \{0,1\}^{\mathbb{J}},$$

for some $0 < P < 1$, where ν and p satisfy (9).

Solving the binary knapsack problem is hard in general, but for certain values of the constraint P, a greedy algorithm yields the solution. These values are exactly the values for which the solution of the binary knapsack problem equals the solution of the *continuous* knapsack problem, which uses the same objective function ν as Optimization Problem 5, and relaxes the constraints $\mathcal{I}_i \in \{0,1\}$ to $\mathcal{I}_i \in [0,1]$. Dantzig [5, Figure 3, p. 274] illustrates this for the classical binary knapsack problem – the finite maximization problem. Martello and Toth [20, Theorem 2.1, p. 16] give an explicit solution for the continuous problem, and a more formal proof.

The greedy algorithm for Optimization Problem 4 is based on the efficiency $r_j = p_j/\nu_j$. The algorithms generates the initial values of an enumeration $j^{(t)}$ of $\mathbb{J}$, $t \in \mathbb{N}_+$, satisfying $r_{j^{(t)}} \geqslant r_{j^{(t+1)}}$. The algorithm recursively generates $I_{(t)}$ from $I_{(t-1)}$, until for some T the condition $p(I_{(T-1)}) < P \leqslant p(I_{(T)})$ holds, where

$$I_{(t)} := \bigcup_{s=1}^{t} j^{(s)}.$$

This greedy algorithm, listed as Algorithm 2, has the following properties.

Lemma 4 *For any $0 < P < 1$, Algorithm 2 terminates for some $t = T$.*

For each $t \geqslant 1$, the set generated by Algorithm 2, $I_{(t)}$ is the solution to Optimization Problem 4 for $P = p(I_{(t)})$.

Algorithm 2: The greedy algorithm for Optimization Problem 4.

Data: error ε, incremental rules Δ_j and their costs ν_j for $j \in \mathbb{J}$
Result: ε approximation q and index set I
$I := I_{(0)} := \{0\};\;\; q := q_{(0)} := \Delta_0;$
while $\|\mathbf{1} - q\| > \varepsilon$ **do**
$\quad\quad j^{(t+1)} := \mathrm{argmax}_i\{r_i \mid i \in \mathbb{J} \setminus I_{(t)}\};$
$\quad\quad I := I_{(t+1)} := I_{(t)} \cup \{j^{(t+1)}\};\;\; q := q_{(t+1)} := q + \Delta_{j^{(t+1)}}\;;$

Proof The algorithm terminates because $j^{(t)}$ is an enumeration of $\mathbb{J}$ and therefore $\sum_{t=1}^{\infty} p_{j^{(t)}} = 1$, since $\sum_{t=1}^{\infty} \Delta_{j^{(t)}} = \mathbf{1}$, but $P < 1$.

When $p(I_{(t)}) = P$ the constraints of Optimization Problem 4 are satisfied. Furthermore, as the method used the largest r_i, the objective function ν is minimised for Optimization Problem 4. A more detailed proof can be constructed along the lines of the proof of Theorem 2.1 of Martello and Toth [20]. □

The construction of the enumeration used in Algorithm 2 requires sorting an infinite sequence and is thus not feasible in general, but in the case where p is strictly decreasing and ν is monotonically increasing, the enumeration can be done recursively in finite time.

Lemma 5 *If p is strictly decreasing and ν is monotonically increasing, at each step $t > 0$ of Algorithm 2, the index $j^{(t+1)}$ produced by the algorithm is a minimal element of the set $\mathbb{J} \setminus I_{(t)}$. Also $j^{(0)} = 0$. Therefore $I_{(t+1)}$ is a down-set.*

Proof If p is strictly decreasing and ν is monotonically increasing, then r is monotonically decreasing. By construction, $r_{j^{(t)}} \geqslant r_{j^{(t+1)}}$, so the enumeration must have $j^{(t)} < j^{(t+1)}$. It follows that $j^{(0)} = 0$.

For $t > 0$, since $j^{(t+1)}$ is an enumeration of $\mathbb{J}$, no element occurs twice, and so $j^{(t+1)} \in \mathbb{J} \setminus I_{(t)}$. For $s > 1$, any later element $j^{(t+s)}$ in the enumeration cannot be smaller than $j^{(t+1)}$, so $j^{(t+1)}$ is a minimal element of $\mathbb{J} \setminus I_{(t)}$, that is $j^{(t+1)} \in M_{(t+1)} = M(I_{(t)})$. Since all elements smaller than $j^{(t+1)}$ occur earlier in the enumeration, we must have $\downarrow j^{(t+1)} \subset I_{(t)} \cup \{j^{(t+1)}\}$. Therefore, if $I_{(t)}$ is a down-set, then so is $I_{(t+1)}$. Since $I_{(0)} = \{0\}$, by induction, $I_{(t+1)}$ is always a down-set.   □

Corollary 6 *For each $T \geqslant 1$, the set $I_{(T)}$ generated by Algorithm 2 is the solution to Optimization Problem 3 for $P = p(I_{(T)})$.*

Proof This is an immediate consequence of Lemmas 3, 4 and 5.   □

In the case where r is strictly decreasing, we have the following result.

Lemma 7 *If the efficiency r is strictly decreasing, then Algorithm 2 produces the same sequence of sets I as Algorithm 1.*

Proof From the proof of Lemma 5, we have that if r is strictly decreasing, then for $t > 0$, $j^{(t+1)}$ as per Algorithm 2 is always the element of $M_{(t+1)}$ which maximizes r. This is exactly $j := \operatorname{argmax}_i \{r_i \mid i \in M(I_{(t)})\}$, as per Algorithm 1. For both algorithms, $I_{(0)} = \{0\}$. $\square$

All the pieces are now in place for the main proof of this Section.

Proof of Theorem 2. From Corollary 6 we see that if the efficiency r is strictly decreasing, then each set I produced by the greedy algorithm (Algorithm 2) is a solution of Optimization Problem 3. From Lemma 7 we see that if the efficiency r is strictly decreasing, then Algorithm 2 produces the same sequence of sets I as Algorithm 1.

If p is strictly decreasing and ν is monotonically increasing, then since $r_j = p_j / \nu_j$ then r is strictly decreasing. Therefore each set I in the sequence produced by Algorithm 1 is a solution of Optimization Problem 3. $\square$

5 Error bounds

We will now describe a second variant of WTP quadrature, $q^{(\mathrm{WW})}$ on $\mathbb{H}_{d,\gamma}^{(r)}$, based on the same set of incremental rules as that used for the sequence of quadrature rules $q^{(\mathrm{DA})}$ described in Section 3 above, but where the order in which the incremental rules are added to this variant is based on the construction of Wasilkowski and Woźniakowski [25, Section 5]. The construction of this variant uses criteria similar to those used by Wasilkowski and Woźniakowski [25, Theorem 3], but adapted to our setting. These criteria are

$$\|q_0\|_{\mathbb{H}_{1,\gamma}^{(r)}} \leqslant 1, \quad \|q_j - q_{j-1}\|_{\mathbb{H}_{1,\gamma}^{(r)}} \leqslant \sqrt{\gamma} C D^j \quad \text{for all } j \geqslant 1, \text{ and all } 0 < \gamma \leqslant 1 \tag{10}$$

(corresponding to Wasilkowski and Woźniakowski [25, (39)]), and

$$\nu_0 = 1, \quad \nu_j D^{j\rho} \leqslant 1 \quad \text{for all } j \geqslant 1 \tag{11}$$

(corresponding to Wasilkowski and Woźniakowski [25, (36)]),
for some $D \in (0, 1)$ and some positive C and ρ.

As a consequence of (10), we have

$$\|\Delta_j\|_{\mathbb{H}_{d,\gamma}^{(r)}} = \prod_{k=1}^d \left\|\delta_{j_k}^{(k)}\right\|_{\mathbb{H}_{1,\gamma_{d,k}}^{(r)}} \leqslant b(d, j), \quad \text{where}$$

$$b(d, j) := \prod_{k=1}^d \left(\sqrt{\gamma_{d,k}} \, C D^{j_k}\right)^{1 - \partial_{0,j_k}}. \tag{12}$$

Here, and below, we use

$$\partial_{a,b} := \begin{cases} 0, & a \neq b \\ 1, & a = b \end{cases}$$

to denote the Kronecker delta.

Let $(\xi_{d,k})$, $k = 1, \ldots, d$, be a sequence of positive numbers. In contrast to Wasilkowski and Woźniakowski [25, Section 5], we do not stipulate that $\xi_{d,1} = 1$, but instead, $\xi_{d,1} \geqslant \sqrt{1 - D^2}$. Define

$$\xi(d,j) := \prod_{k=1}^{d} \xi_{d,k}^{1 - \partial_{0,j_k}}. \tag{13}$$

We therefore have $b(d,j)/\xi(d,j) \to 0$ as $\|j\|_1 \to \infty$. We order the incremental rules in order of non-decreasing $b(d,j)/\xi(d,j)$ for each index j, creating an order on the indices $j^{(\text{WW})}(h)$. We adjust $\xi(d,k)$ so that this order agrees with the lattice partial ordering of the indices. We now define $I_N^{(\text{WW})} := \{j^{(\text{WW})}(1), \ldots, j^{(\text{WW})}(N)\}$, and define the quadrature rule

$$q_N^{(\text{WW})} := \sum_{j \in I_N^{(\text{WW})}} \Delta_j. \tag{14}$$

To obtain a quadrature error of at most $\varepsilon \in (0,1)$, we set

$$N(d, \varepsilon, \eta) := \left| \left\{ j \mid b(d,j)/\xi(d,j) > \left(\varepsilon/C_1(d,\eta) \right)^{1/(1-\eta)} \right\} \right|, \tag{15}$$

where $\eta \in (0,1)$ and

$$C_1(d, \eta) := \sqrt{\frac{\xi_{d,1}^{2(1-\eta)}}{1 - D^2} \prod_{k=2}^{d} \left(1 + (C^2 \gamma_{d,k})^\eta \, \xi_{d,k}^{2(1-\eta)} \frac{D^{2\eta}}{1 - D^{2\eta}} \right)}. \tag{16}$$

Finally, we define

$$q_{d,\varepsilon,\eta}^{(\text{WW})} := \sum_{j \in I_{N(d,\varepsilon,\eta)}}^{(\text{WW})} \Delta_j. \tag{17}$$

We can now present our version of Wasilkowski and Woźniakowski's main theorem on the error and cost of WTP quadrature [25, Theorem 3].

Theorem 8 *Let $\eta \in (0,1)$. Assume that a sequence of quadrature points $x_1, x_2, \ldots \in \mathbb{S}^2$, and a sequence of positive integers $1 = n_0 < n_1 < \ldots$ are given such that the corresponding optimal weight quadrature rules $q_j := q_j^1 \in \mathbb{H}_{1,1}^{(r)}$ satisfy (10) and (11) for some $D \in (0,1)$ and some positive C and ρ. Then*

the quadrature rule $q_{d,\varepsilon,\eta}^{(\mathrm{WW})}$ defined by (17) has worst-case quadrature error $e(q_{d,\varepsilon,\eta}^{(\mathrm{WW})}) \leqslant \varepsilon$, and its cost (in number of quadrature points) is bounded by

$$\mathrm{cost}(q_{d,\varepsilon,\eta}^{(\mathrm{WW})}) \leqslant C(d,\varepsilon,\eta) \left(\frac{1}{\varepsilon}\right)^{\rho/(1-\eta)}, \tag{18}$$

where

$$C(d,\varepsilon,\eta) := \frac{\max(\sqrt{\gamma_{d,1}},\xi_{d,1})^\rho \prod_{k=2}^{d} \left(1 + \dfrac{C^\rho \gamma_{d,k}^{\rho/2}}{\xi_{d,k}^\rho}\, g(k,\varepsilon,\eta)\right) f_{k,\eta}^\rho}{(1-D^\rho)(1-D^2)^{\rho/(2-2\eta)}} \tag{19}$$

$$f_{k,\eta} := \left(1 + C^{2\eta}\gamma_{d,k}^\eta \xi_{d,k}^{2(1-\eta)} \frac{D^{2\eta}}{1-D^{2\eta}}\right)^{1/(2(1-\eta))}, \tag{20}$$

$$g(k,\varepsilon,\eta) := \left\lfloor \frac{\log\left(\dfrac{C\gamma_{d,k}^{1/2}}{(1-D^2)^{1/(2-2\eta)}}\, \dfrac{\xi_{d,1}}{\xi_{d,k}} \left(\prod_{i=2}^{k} f_{i,\eta}\right) \left(\dfrac{1}{\varepsilon}\right)^{1/(1-\eta)}\right)}{\log\left(D^{-1}\right)} \right\rfloor_+ . \tag{21}$$

By $\lfloor x \rfloor_+$, we mean $\max(0,x)$.

Wasilkowski and Woźniakowski's proof, with $s := 2, \alpha := 1$, applies directly to our Theorem 8, once the change in $\xi_{d,1}$ is taken into account. For details of the proof, see Appendix A.

Corollary 1 of Wasilkowski and Woźniakowski [25, p. 434] presents a simpler bound for the cost of their WTP algorithm, and their simplification also applies here.

Corollary 9 *For every positive δ there exists a positive $c(d,\delta)$ such that the cost of the quadrature rule $q_{d,\varepsilon,\eta}^{(\mathrm{WW})}$ defined by (17) is bounded by*

$$\mathrm{cost}(q_{d,\varepsilon,\eta}^{(\mathrm{WW})}) \leqslant c(d,\delta) \left(\frac{1}{\varepsilon}\right)^{\rho+\delta} .$$

Following Theorem 4 of Wasilkowski and Woźniakowski [25] we now examine the dependence of our cost bound (18) on the dimension d. If $\xi_{d,k} = c$ for all $k \geqslant 2$, for some positive constant c, then for exponentially decreasing dimension weights $\gamma_{d,k} = g^k$ with $g < 1$, we have

$$\sup_d \left\{ \sum_{k=2}^{d} \left(\frac{\gamma_{d,k}^{1/2}}{\xi_{d,k}}\right)^\rho , \sum_{k=2}^{d} \left(\xi_{d,k}^s \frac{\gamma_{d,k}^{1/2}}{\xi_{d,k}}\right)^{s\eta} \right\} = \max\left\{ c^{-\rho}\frac{g^\rho}{1-g^{\rho/2}},\, c^{s-s\eta}\frac{g^{s\eta}}{1-g^{s\eta/2}} \right\}$$

$$< \infty.$$

Examining Theorem 4 of of Wasilkowski and Woźniakowski [25] and its proof in some detail, we see that a similar argument also applies to the $q^{(\mathrm{WW})}$ rules. This implies the following.

Theorem 10 *If $\xi_{d,k} = c$ for all $k \geqslant 2$, for some positive constant c, then for exponentially decreasing dimension weights $\gamma_{d,k} = g^k$ with $g < 1$, the algorithm giving the quadrature rule $q_{d,\varepsilon,\eta}^{(\mathrm{WW})}$ defined by (17) is strongly polynomial, and in particular, for each positive δ there exists c_δ such that*

$$\mathrm{cost}(q_{d,\varepsilon,\eta}^{(\mathrm{WW})}) \leqslant c_\delta \left(\frac{1}{\varepsilon}\right)^{\rho/(1-\eta)+\delta},$$

for all $\varepsilon \in (0,1)$ and all $d \geqslant 1$.

The sequence of rules $q^{(\mathrm{DA})}$ is at least as efficient as $q^{(\mathrm{WW})}$, in the sense that $q^{(\mathrm{DA})}$ is based on the optimal solution of the corresponding down-set-constrained continuous knapsack problem, as explained in Section 4. As a direct consequence of Theorem 8 and Corollary 9, we therefore have the following result.

Theorem 11 *Let $\eta \in (0,1)$. Assume that a sequence of quadrature points $x_1, x_2, \ldots \in \mathbb{S}^2$, and a sequence of positive integers $n_0 < n_1 < \ldots$ are given such that the corresponding optimal weight quadrature rules $q_j := q_j^1 \in \mathbb{H}_{1,1}^{(r)}$ satisfy (10) and (11) for some $D \in (0,1)$ and some positive C and ρ. Let $I_{(t)}, q_{(t)}^{(\mathrm{DA})}$ be the index set and corresponding quadrature rule generated by iteration t of Algorithm 1, based on the rules q_j, for sufficiently small error $\varepsilon = \varepsilon_0$.*

Then the quadrature rule $q_{(t)}^{(\mathrm{DA})}$ has worst-case quadrature error $e(q_{(t)}^{(\mathrm{DA})}) = \varepsilon_{(t)} := \sqrt{1 - p(I_{(t)})}$, and its cost $\nu(I_{(t)})$ is bounded by

$$\nu(I_{(t)}) \leqslant C(d, \varepsilon_{(t)}, \eta) \left(\frac{1}{\varepsilon_{(t)}}\right)^{\rho/(1-\eta)},$$

where $C(d, \varepsilon_{(t)}, \eta)$ is defined as per Theorem 8. As a consequence, for every positive δ there exists a positive $c(d, \delta)$ such that the cost of the quadrature rule $q_{(t)}^{(\mathrm{DA})}$ is bounded by

$$\nu(I_{(t)}) \leqslant c(d, \delta) \left(\frac{1}{\varepsilon_{(t)}}\right)^{\rho+\delta}.$$

6 Numerical results

With the estimates given by our analysis in hand, we now compare these to the results of a selection of numerical examples.

Since the underlying domain $\mathcal{D}$ is $\mathbb{S}^2$ rather than $\mathbb{S}^1$, we need to change some of the details of the algorithm in comparison to the algorithm used for the torus [13]. Specifically, we need a sequence of rules on a single sphere, which yields "good enough" worst case quadrature error with optimal weights. The numerical examples of this Section use a sequence of point sets consisting

of unions of spherical designs with increasing numbers of points and non-decreasing strengths.

For the unit sphere $\mathbb{S}^2$, a spherical design [7] of strength t and cardinality m is a set of m points $X = \{x_1, \ldots, x_m\} \subset \mathbb{S}^2$ such that the equal weight quadrature rule

$$Q_X(p) := \frac{1}{m} \sum_{h=1}^{m} p(x_h)$$

is exact for all spherical polynomials p of total degree at most t.

For the numerical examples of this Section, a combination of (approximate) extremal (E) and low cardinality (L) spherical designs are used, according to Table 1. These approximate spherical designs were all provided by Womersley [4,26].

Index j	0	1	2	3	4	5	6	7	8	9	10	11
Type	L	L	E	L	E	L	E	L	E	L	E	L
Strength t	0	1	1	3	3	7	7	15	15	31	31	63
Cardinality m	1	2	4	8	16	32	64	129	256	513	1024	2049

Table 1 Strength and cardinality of approximate spherical designs used with Algorithm 1 in the numerical examples.

If we let $m_j := |X_j|$, the cardinality of X_j, and let t_j be the strength of X_j, then the extremal spherical designs listed in Table 1 have $m_j = (t_j + 1)^2$, and the low cardinality spherical designs have $m_j = (t_j + 1)^2/2$ or $m_j = (t_j+1)^2/2+1$, and in all cases $t_j \geqslant \sqrt{m_j} - 1$. It is not yet known if a sequence of spherical designs satisfying this lower bound on strength can be extended indefinitely, but there is rigorous computational proof for t_j up to 100 [3]. Also, it is now known that an infinite sequence of spherical designs exists with the required asymptotic order of strength, that is, there is a sequence of spherical designs of cardinality $m_j = \mathcal{O}(t_j^2)$, [2] but the proof is not constructive and the corresponding implied constant is still unknown.

One difference between the constructions for $\mathbb{S}^1$ and $\mathbb{S}^2$ is that the nesting of spherical designs is not efficient. The union of two spherical designs of strengths t_1 and t_2 is in general, a spherical design whose strength is the *minimum* of t_1 and t_2. For the approximate spherical designs listed in Table 1, the first design of strength 0 is a single point. The next design of strength 1 consists of two antipodal points, so nesting is efficient in this case. After this, the resulting unions of spherical designs, in general, have strength no greater than 1. This is one motivation for using optimal weight rules rather than equal weight rules on a single sphere. In the optimal weight case, adding points to a rule does not increase its worst case error.

We now turn to estimates for rules on a single sphere, in order to use them with Theorem 8. On a single unit sphere, the DA sparse grid quadrature

rule whose worst case error is to be estimated is an optimal weight rule $q_j = q_r^\gamma(S_j)$, based on the increasing union of the approximate spherical designs from Table 1, $S_j := \bigcup_{i=0}^{j} X_i$. Its worst case error is therefore no larger than that of the optimal weight rule $q_r^\gamma(X_j)$ based on X_j, the largest spherical design contained in the union, which is, in turn no greater than the worst case error of the equal weight rule $q_r^{\gamma,(\mathrm{QMC})}(X_j)$ based on X_j, with weights $1/|X_j|$,

$$e^2(q_j) < e^2\big(q_r^\gamma(X_j)\big) \leqslant e^2\big(q_r^{\gamma,(\mathrm{QMC})}(X_j)\big).$$

According to Hesse, Kuo and Sloan [15, Theorem 4], for $\gamma = 1$, we have the bound

$$e^2\big(q_r^{1,(\mathrm{QMC})}(X_j)\big) \leqslant c t_j^{-2r},$$

and therefore

$$e^2\big(q_r^1(X_j)\big) \leqslant c(\sqrt{m_j} - 1)^{-2r} \leqslant C_1 m_j^{-r},$$

for some $C_1 > 0$. For general γ, as per Kuo and Sloan [19], we have

$$e^2\big(q_r^{\gamma,(\mathrm{QMC})}(X_j)\big) = -1 + \frac{1}{m_j^2} \sum_{h=1}^{m_j} \sum_{i=1}^{m_j} K_{1,\gamma}^{(r)}(x_{j,h}, x_{j,i})$$

$$= \gamma \frac{1}{m_j^2} \sum_{h=1}^{m_j} \sum_{i=1}^{m_j} A_r(x_{j,h} \cdot x_{j,i}) \leqslant \gamma C_1 m_j^{-r}. \qquad (22)$$

For the sequence of spherical designs given by Table 1, we also have

$$2^j \leqslant m_j \leqslant 2^j + 1.$$

We therefore have

$$e^2(q_j) \leqslant \gamma C_2\, 2^{-rj}, \qquad (23)$$

for some $C_2 > 0$. Recall that

$$e^2(q_j) = 1 - \|q_j\|_{\mathbb{H}_{1,\gamma}^{(r)}}^2$$

$$= 1 - \|q_{j-1}\|_{\mathbb{H}_{1,\gamma}^{(r)}}^2 - \|q_j - q_{j-1}\|_{\mathbb{H}_{1,\gamma}^{(r)}}^2$$

$$= e^2(q_{j-1}) - \|q_j - q_{j-1}\|_{\mathbb{H}_{1,\gamma}^{(r)}}^2,$$

for $j \geqslant 1$, since $q_j - q_{j-1}$ is orthogonal to q_{j-1}. Therefore

$$e^2(q_{j-1}) = e^2(q_j) + \|q_j - q_{j-1}\|_{\mathbb{H}_{1,\gamma}^{(r)}}^2.$$

Since $e^2(q_j) \geqslant 0$, using (23) we obtain $\|q_j - q_{j-1}\|_{\mathbb{H}_{1,\gamma}^{(r)}}^2 \leqslant \gamma C_2\, 2^{-r(j-1)}$. This, in turn implies that

$$\|q_j - q_{j-1}\|_{\mathbb{H}_{1,\gamma}^{(r)}}^2 \leqslant \gamma C_3\, 2^{-rj},$$

where $C_3 = 2^r C_2$.

All the approximate spherical designs listed in Table 1 have one point in common, the "north pole" $(0,0,1)$. Therefore the number of points $n_j = |S_j|$ of the DA quadrature rule with this sequence of point sets satisfies $n_j = 1 + \sum_{i=0}^{j}(m_i - 1) = -j + \sum_{i=0}^{j} m_i$. Since $2^i \leqslant m_i \leq 2^i + 1$, we have $2^{j+1} - j - 1 \leqslant n_j \leqslant 2^{j+1}$, and so $m_j \leqslant n_j \leqslant m_{j+1}$, for $j \geqslant 0$. In view of (22), and the preceding argument, criteria (10) and (11) hold with $D = 2^{-r/2}$, $C = \sqrt{C_3}$ and $\rho = 2/r$.

Consider now that the spherical design for $j = 0$ consists of only one point, the north pole, and the spherical design for $j = 1$ consists of two points, the north pole and its antipode, the "south pole" $(0,0,-1)$. From (22) we have

$$e^2\big(q_r^{\gamma,(\mathrm{QMC})}(X_0)\big) = \gamma A_r(1), \text{ and}$$

$$e^2\big(q_r^{\gamma,(\mathrm{QMC})}(X_1)\big) = \frac{\gamma}{4} \sum_{h=1}^{2} \sum_{i=1}^{2} A_r(x_{1,h} \cdot x_{1,i}) = \frac{\gamma}{2}\big(A_r(1) + A_r(-1)\big), \text{ so that}$$

$$\begin{aligned}
\|q_j - q_{j-1}\|_{\mathbb{H}_{1,\gamma}^{(r)}}^2 &= e^2\big(q_r^{\gamma,(\mathrm{QMC})}(X_0)\big) - e^2\big(q_r^{\gamma,(\mathrm{QMC})}(X_1)\big) \\
&= \frac{\gamma}{2}\big(A_r(1) - A_r(-1)\big) \leqslant \gamma C_3 \, 2^{-r},
\end{aligned}$$

yielding $C_3 \geqslant 2^{r-1}\big(A_r(1) - A_r(-1)\big)$. In other words,

$$C \geqslant 2^{(r-1)/2}\sqrt{A_r(1) - A_r(-1)}.$$

Note that the sequence of point sets used on a single sphere, as listed in Table 1 does not vary with γ, as would be the case in the original setting of Wasilkowski and Woźniakowski [25]. The estimates above show that in the setting of this paper it is not necessary to vary the sequence in this way.

The numerical examples of this Section use exponentially decreasing dimension weights, partly because this guarantees strong tractability [19, Theorem 4], and partly because this allows comparison with the numerical examples used by Hesse, Kuo and Sloan in the QMC case [15, Section 7]. To see how the rules $q^{(\mathrm{DA})}$ and $q^{(\mathrm{WW})}$ behave as the decay of the dimension weights is varied, the numerical examples used here use $r = 3$ with $\gamma_k = g^k$, for $g = 0.1$, 0.5, and 0.9,

For the $q^{(\mathrm{WW})}$ rules, we use the definition (14) with $\xi_{d,k} := CD$, where $D = 2^{-r/2}$ as above, and

$$C := 2^{(r-1)/2}\sqrt{A_r(1) - A_r(-1)} = \sqrt{A_3(1) - A_3(-1)} \simeq 1.7426.$$

To further validate these values, we use the $q^{(\mathrm{DA})}$ rules with $r = 3$, $d = 1$ and $\gamma_1 = 1$, to verify numerically that (10) holds for all of the sets S_j generated from the designs in Table 1. In fact, numerically, (10) also holds for D as above and $C = \|q_1 - q_0\|_{\mathbb{H}_{1,1}^{(3)}} = \sqrt{e^2(q_0) - e^2(q_1)} \simeq 1.453$. That is,

$$2^{rj/2}\|q_j - q_{j-1}\|_{\mathbb{H}_{1,1}^{(3)}} \leqslant \|q_1 - q_0\|_{\mathbb{H}_{1,1}^{(3)}}, \text{ for } j = 2\dots 11.$$

Each of the numerical examples described below uses a particular dimension d, from $d = 1$ to 16; a particular maximum 1-norm for indices, typically 20; and a particular maximum number of points, up to 100 000.

The numerical results are potentially affected by three problems. First, if γ is close to zero, and the number of points is large, then the matrix used to compute the weights becomes ill-conditioned, and the weights may become inaccurate. In this case, a least squares solution is used to obtain a best approximation to the weights. Second, if the current squared error is close to zero, and the squared norm for the current index is close to machine epsilon, then severe cancellation may occur. Third, the sequence of spherical designs used in our numerical examples is finite, so it is quite possible that our algorithm generates an index corresponding to a spherical design which is not included in our finite set. In these last two cases, the calculation of the quadrature rule is terminated.
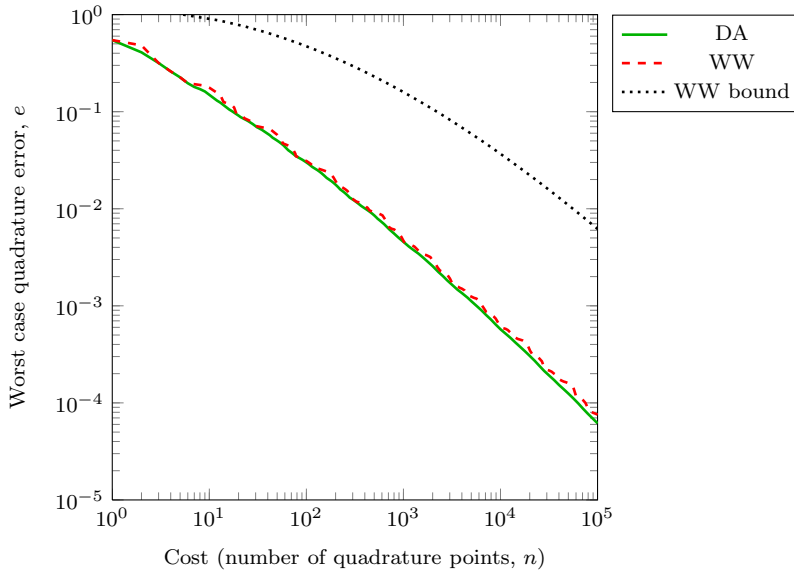

Fig. 4 Error of DA and WW rules vs WW bound for $(\mathbb{S}^2)^4$, $r = 3$, $\gamma_{4,k} = 0.5^k$.

Figure 4 is a log-log plot displaying the typical convergence behaviour of the DA and WW rules for the cases examined. The case shown is that of $(\mathbb{S}^2)^4$, $r = 3$, $\gamma_{4,k} = 0.5^k$. The number of points used varies from $n = 1$ to 100 000. The cost axis is horizontal and the error axis is vertical, to match the figures shown in the torus paper [13]. The curve in Figure 4 labelled "WW bound" is the minimum of the bounds given by Theorem 8, as the parameter η is varied over a finite number of values between 0 and 1.

In general, the DA algorithm has a cost no greater than that of the WW algorithm. Both are bounded by the WW bound of Theorem 8. The WW cost bound

itself has an asymptotic rate of convergence of $\mathcal{O}(\varepsilon^{-\rho}) = \mathcal{O}(\varepsilon^{-2/r}) = \mathcal{O}(\varepsilon^{-2/3})$ for all of our cases. In other words, the asymptotic bound has quadrature error of order $\mathcal{O}(n^{-r/2}) = \mathcal{O}(n^{-3/2})$.

Judging from the slopes of the curves in Figure 4, the rates of convergence of both algorithms appear consistent with that of the bound, but the asymptotic rate is not achieved by either algorithm or by the bound itself, for the number of quadrature points displayed in the plot.
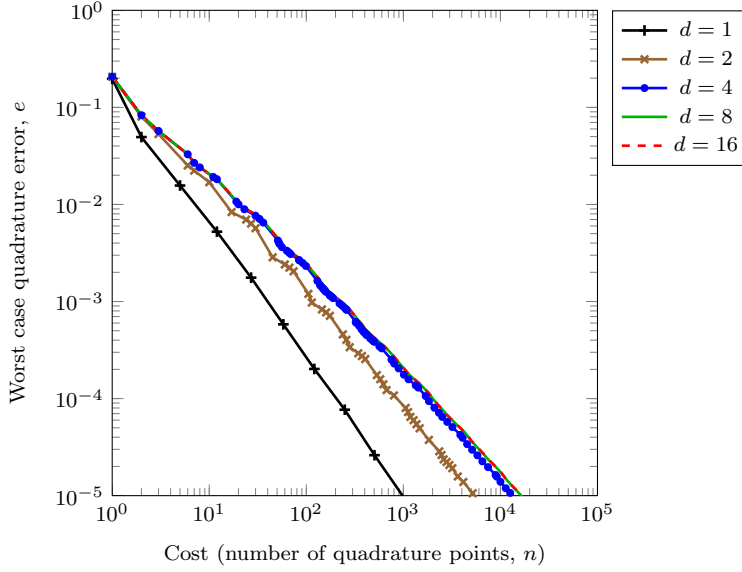

Fig. 5 Error of DA rules for $(\mathbb{S}^2)^d$, $d = 1, 2, 4, 8, 16$; $r = 3$, $\gamma_{d,k} = 0.1^k$.

For $\gamma_{d,k} = 0.1^k$, Figure 5 shows how the convergence rate of the error of the DA quadrature rules varies with dimension d, for $d = 1$, 2, 4, 8, and 16. The curve for $d = 1$ appears consistent with the asymptotic error rate $\varepsilon = \mathcal{O}(N^{-3/2})$. The cases $d = 8$ and $d = 16$ are almost indistinguishable on this Figure. This is an example of the convergence in dimension.

Figure 6 shows the equivalent results for the DA quadrature rules for $\gamma_{d,k} = 0.9^k$. The curve for $d = 1$ again appears consistent with the asymptotic error rate $\varepsilon = \mathcal{O}(N^{-3/2})$, but as d increases to 16, the initial rate of convergence to zero of the error becomes much slower than that for $\gamma_{d,k} = 0.1^k$. For example, for $d = 8$, more than 1000 quadrature points are needed to reduce the quadrature error from 1 to 0.1, and for $d = 16$, more than 100 000 quadrature points are needed. A similar phenomenon is observed for sparse grid quadrature with RKHS on the torus [13].
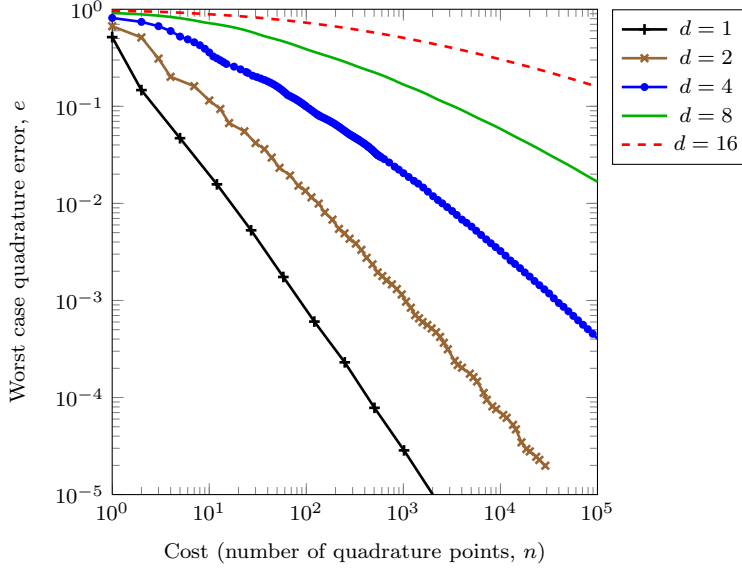
Fig. 6 Error of DA rules for $(\mathbb{S}^2)^d$, $d = 1, 2, 4, 8, 16$; $r = 3$, $\gamma_{d,k} = 0.9^k$.

Appendix A   Proof of Theorem 8

Assume that a sequence of quadrature points $x_1, x_2, \ldots \in \mathbb{S}^2$ and a sequence of positive integers $1 = n_0 < n_1 < \ldots$ are given such that for all $\gamma \in (0, 1]$ the corresponding optimal weight quadrature rules

$$q_j^\gamma := \sum_{\ell=1}^{n_j} w_\ell^\gamma k_{x_\ell}^\gamma \in \mathbb{H}_{1,\gamma}^{(r)} \quad \text{satisfy}$$

$$\|q_0^\gamma\|_{\mathbb{H}_{1,\gamma}^{(r)}} \leqslant 1, \quad \left\|q_j^\gamma - q_{j-1}^\gamma\right\|_{\mathbb{H}_{1,\gamma}^{(r)}} \leqslant \sqrt{\gamma} CD^j, \quad \text{and} \tag{24}$$

$$\nu_j D^{j\rho} \leqslant 1, \quad \text{for all } j \geqslant 1, \tag{25}$$

for some $D \in (0, 1)$ and some positive C and ρ, where $\nu_0 := 1$ and $\nu_j := n_j - n_{j-1}$ for all $j \geqslant 1$.

Note that while the points x_ℓ do not change with γ, the weights w_ℓ^γ *do* change, and so each rule q_j^γ depends on γ in general.

Define

$$b_j^\gamma := \begin{cases} 1, & j = 0, \\ \sqrt{\gamma} CD^j, & j \geqslant 1; \end{cases} \quad \delta_j^\gamma := \begin{cases} q_0^\gamma, & j = 0, \\ q_j^\gamma - q_{j-1}^\gamma, & j \geqslant 1. \end{cases}$$

Then (24) implies that

$$\left\|\delta_j^\gamma\right\|_{\mathbb{H}_{1,\gamma}^{(r)}} \leqslant b_j^\gamma \quad \text{for all } j \geqslant 0.$$

As a consequence, and since $\|q_0^\gamma\|_{\mathbb{H}_{1,\gamma}^{(r)}} \leqslant 1$, it holds that

$$\left\| \Delta_j^\gamma \right\|_{\mathbb{H}_{d,\gamma}^{(r)}} \leqslant b(d,j), \tag{26}$$

where

$$b(d,j) := \prod_{k=1}^d b_{j_k}^{\gamma_{d,k}} = \prod_{k=1}^d \left(\sqrt{\gamma_{d,k}}\, CD^{j_k} \right)^{1-\partial_{0,j_k}},$$

as per (12).

Let $(\xi_{d,k})$, $k = 1, \ldots, d$, be a sequence of positive numbers, with $\xi_{d,1} \geqslant \sqrt{1 - D^2}$, and define

$$\xi(d,j) := \prod_{k=1}^d \xi_{d,k}^{1-\partial_{0,j_k}}.$$

as per (13). We intend to sort the incremental rules in order of decreasing $b(d,j)/\xi(d,j)$. For this order to agree with the lattice ordering, it must be the case that $b_1^{\gamma_{d,k}}/\xi_{d,k} \leqslant 1$, that is, $\xi_{d,k} \geqslant \sqrt{\gamma_{d,k}}CD$ for all k.

Let

$$I(d,\varepsilon,\eta) := \left\{ j \;\middle|\; b(d,j)/\xi(d,j) > \left(\varepsilon/C_1(d,\eta) \right)^{1/(1-\eta)} \right\},$$

where $\eta \in (0,1)$ and

$$C_1(d,\eta) := \sqrt{ \frac{\xi_{d,1}^{2(1-\eta)}}{1-D^2} \prod_{k=2}^d \left(1 + (C^2 \gamma_{d,k})^\eta\, \xi_{d,k}^{2(1-\eta)} \frac{D^{2\eta}}{1-D^{2\eta}} \right) },$$

as per (16).

Let $N(d,\varepsilon,\eta) := |I(d,\varepsilon,\eta)|$, and let $j^{(\mathrm{WW})} = (j_{(1)}^{(\mathrm{WW})}, j_{(2)}^{(\mathrm{WW})}, \ldots)$ be an ordering of indices the indices in $\mathbb{J}$ by decreasing $b(d,j)/\xi(d,j)$. Then

$$I_{N(d,\varepsilon,\eta)}^{(\mathrm{WW})} := \left\{ j_{(1)}^{(\mathrm{WW})}, j_{(2)}^{(\mathrm{WW})}, \ldots, j_{(N(d,\varepsilon,\eta))}^{(\mathrm{WW})} \right\} = I(d,\varepsilon,\eta).$$

Define

$$q_{d,\varepsilon,\eta}^{(\mathrm{WW})} := \begin{cases} \sum_{j \in I(d,\varepsilon,\eta)} \Delta_j, & \varepsilon < 1, \\ 0, & \varepsilon \geqslant 1, \end{cases} \tag{27}$$

in agreement with (17) for $\varepsilon < 1$.

The proof now proceeds in two main steps, closely following Wasilkowski and Woźniakowski's proof of their [25] Theorem 3, for comparison purposes. Firstly, we prove the error bound $e(q_{d,\varepsilon,\eta}^{(\mathrm{WW})}) \leqslant \varepsilon$. Secondly, we prove the cost bound (18).

For $\varepsilon \geqslant 1$, from (5) and (27) we have $e(q_{d,\varepsilon,\eta}^{(\mathrm{WW})}) = 1$. Let

$$\alpha(d,\varepsilon,\eta) := \left(\varepsilon/C_1(d,\eta)\right)^{1/(1-\eta)} \tag{28}$$

so that $I(d,\varepsilon,\eta) = \{j \mid b(d,j) > \xi(d,j)\,\alpha(d,\varepsilon,\eta)\}$. From (26), and using the orthogonality of the rules Δ_j, we therefore have

$$e^2(q_{d,\varepsilon,\eta}^{(\mathrm{WW})}) \leqslant e^2(d,\varepsilon,\eta) := \begin{cases} \sum_{b(d,j)\leqslant\xi(d,j)\,\alpha(d,\varepsilon,\eta)} b^2(d,j), & \varepsilon < 1, \\ 0, & \varepsilon \geqslant 1. \end{cases}$$

For $h = 1, 2, \ldots, d$, we also define

$$\overline{e}^2(h,\varepsilon,\eta) := \begin{cases} \sum_{\overline{b}(h,j)\leqslant\overline{\xi}(h,j)\,\overline{\alpha}(h,\varepsilon,\eta)} \overline{b}^2(h,j), & \varepsilon < 1, \\ 0, & \varepsilon \geqslant 1, \end{cases}$$

where

$$\overline{b}(h,j) := \prod_{k=1}^{h} b_{j_k}^{\gamma_{d,k}} = \prod_{k=1}^{h} \left(\sqrt{\gamma_{d,k}}\,CD^{j_k}\right)^{1-\partial_{0,j_k}},$$

$$\overline{\xi}(h,j) := \prod_{k=1}^{h} \xi_{d,k}^{1-\partial_{0,j_k}}, \quad \overline{\alpha}(h,\varepsilon,\eta) := \left(\varepsilon/\overline{C_1}(h,\eta)\right)^{1/(1-\eta)}, \tag{29}$$

$$\overline{C_1}(h,\eta) := \sqrt{\frac{\xi_{d,1}^{2(1-\eta)}}{1-D^2} \prod_{k=2}^{h} \left(1 + (C^2\gamma_{d,k})^\eta\,\xi_{d,k}^{2(1-\eta)}\frac{D^{2\eta}}{1-D^{2\eta}}\right)}.$$

Note that $\overline{e}(d,\varepsilon,\eta) = e(d,\varepsilon,\eta)$.

We show by induction on h that $\overline{e}(h,\varepsilon,\eta) \leqslant \varepsilon$ for all $\varepsilon > 0$. For $h = 1$, if $\varepsilon \geqslant 1$, this is trivially true, since $\overline{e}(1,\varepsilon,\eta) = 1$. For $h = 1$ and $\varepsilon < 1$, let

$$N_\varepsilon^* := \left|\{j \mid \overline{b}(1,j) > \overline{\xi}(1,j)\,\overline{\alpha}(1,\varepsilon,\eta)\}\right|. \tag{30}$$

Since $\overline{b}(1,0) = 1$, $\overline{\xi}(1,0) = 1$, and $\overline{\alpha}(1,\varepsilon,\eta) = \left(\varepsilon/\overline{C_1}(1,\eta)\right)^{1/(1-\eta)}$, it holds that if $\overline{C_1}(1,\eta) \geqslant 1$, then $N_\varepsilon^* \geqslant 1$. But

$$\overline{C_1}(1,\eta) = \frac{\xi_{d,1}^{1-\eta}}{\sqrt{1-D^2}},$$

and we have specified that $\xi_{d,1} \geqslant \sqrt{1-D^2}$, so it follows that $\overline{C_1}(1,\eta) \geqslant 1$. Now

$$\overline{e}^2(1,\varepsilon,\eta) = \sum_{j\geqslant N_\varepsilon^*} \overline{b}^2(1,j) = \sum_{j\geqslant N_\varepsilon^*} \gamma_{d,1}\,C^2 D^{2j}$$

$$= \gamma_{d,1}\,C^2 D^{2N_\varepsilon^*} \sum_{j\geqslant 0} D^{2j} = \frac{\gamma_{d,1}\,C^2 D^{2N_\varepsilon^*}}{1-D^2} = \frac{\overline{b}^2(1,N_\varepsilon^*)}{1-D^2},$$

but

$$\overline{b}^2(1, N_\varepsilon^*) \leqslant \overline{\xi}^2(1, N_\varepsilon^*)\, \overline{\alpha}^2(1, \varepsilon, \eta) = \xi_{d,1}^2 \left(\frac{\varepsilon \sqrt{1 - D^2}}{\xi_{d,1}^{1-\eta}} \right)^{2/(1-\eta)}$$

$$= (\varepsilon \sqrt{1 - D^2})^{2/(1-\eta)} = \varepsilon^{2/(1-\eta)}\, (1 - D^2)^{1/(1-\eta)},$$

so

$$\overline{e}^2(1, \varepsilon, \eta) = \frac{\overline{b}^2(1, N_\varepsilon^*)}{1 - D^2} \leqslant \varepsilon^{2/(1-\eta)}\, (1 - D^2)^{\eta/(1-\eta)} \leqslant \varepsilon^2.$$

For the inductive step, let $1 < h \leqslant d$ and assume that $\overline{e}(h-1, \varepsilon, \eta) \leqslant \varepsilon$. The following estimates show that $\overline{e}(h, \varepsilon, \eta) \leqslant \varepsilon$.

Noting that, from their definitions (29), $\overline{b}(h, j) = \overline{b}(h-1, j)\, b_{j_h}^{\gamma_{d,h}}$ and $\overline{\xi}(h, j) = \overline{\xi}(h-1, j)\, \xi_{d,h}^{1-\partial_{0,j_h}}$, let

$$\overline{\beta}_\ell(h, \varepsilon, \eta) := \frac{\xi_{d,h}^{1-\partial_{0,j_h}}\, \overline{\alpha}(h, \varepsilon, \eta)}{b_\ell^{\gamma_{d,h}}}, \tag{31}$$

with $\overline{\alpha}$ as per (29). Since $\overline{b}(h, j) = \overline{b}(h-1, j)\, b_\ell^{\gamma_{d,h}}$ if and only if $\ell = j_h$, and since

$$\overline{\xi}(h, j)\, \overline{\alpha}(h, \varepsilon, \eta) = \overline{\xi}(h-1, j)\, b_\ell^{\gamma_{d,h}}\, \overline{\beta}_\ell(h, \varepsilon, \eta),$$

we see that for $\varepsilon < 1$,

$$\overline{e}^2(h, \varepsilon, \eta) = \sum_{\ell=0}^{\infty} \sum_{\overline{b}(h-1,j) \leqslant \overline{\xi}(h-1,j)\, \overline{\beta}_\ell(h,\varepsilon,\eta)} \overline{b}^2(h-1, j)\, (b_\ell^{\gamma_{d,h}})^2. \tag{32}$$

Now let

$$a_{h,\eta} := \frac{1}{\sqrt{1 + (C^2 \gamma_{d,h})^\eta\, \xi_{d,h}^{2(1-\eta)}\, \frac{D^{2\eta}}{1 - D^{2\eta}}}}. \tag{33}$$

Then $\overline{C_1}(h, \eta) = \overline{C_1}(h-1, \eta)/a_{h,\eta}$, so that

$$\overline{\alpha}(h, \varepsilon, \eta) = \left(\frac{a_{h,\eta}\, \varepsilon}{\overline{C_1}(h-1, \eta)} \right)^{1/(1-\eta)} = \overline{\alpha}(h-1, a_{h,\eta}\, \varepsilon, \eta),$$

and so

$$\overline{\beta}_\ell(h, \varepsilon, \eta) = \frac{\xi_{d,h}^{1-\partial_{0,j_h}}\, \overline{\alpha}(h, \varepsilon, \eta)}{b_\ell^{\gamma_{d,h}}} = \overline{\alpha}\left(h-1, a_{h,\eta}\, \varepsilon\, \left(\frac{\xi_{d,h}^{1-\partial_{0,j_h}}}{b_\ell^{\gamma_{d,h}}} \right)^{1-\eta}, \eta \right). \tag{34}$$

In particular, $\overline{\beta}_0(h, \varepsilon, \eta) = \overline{\alpha}(h - 1, a_{h,\eta}\,\varepsilon, \eta)$, and therefore

$$\sum_{\overline{b}(h-1,j) \leqslant \overline{\xi}(h-1,j)\ \overline{\beta}_0(h,\varepsilon,\eta)} \overline{b}^2(h-1,j) = \sum_{\overline{b}(h-1,j) \leqslant \overline{\xi}(h-1,j)\ \overline{\alpha}(h-1,a_{h,\eta}\,\varepsilon,\eta)} \overline{b}^2(h-1,j)$$

$$= \overline{e}^2(h - 1, a_{h,\eta}\,\varepsilon, \eta).$$

This implies that

$$\overline{e}^2(h, \varepsilon, \eta) =$$

$$\sum_{\ell=0}^{\infty}(b_\ell^{\gamma_{d,h}})^2 \sum_{\overline{b}(h-1,j) \leqslant \overline{\xi}(h-1,j)\ \overline{\alpha}\left(h-1,a_{h,\eta}\,\varepsilon\left(\frac{\xi_{d,h}^{1-\partial_{0,j_h}}}{b_\ell^{\gamma_{d,h}}}\right)^{1-\eta},\eta\right)} \overline{b}^2(h-1,j)$$

$$= \overline{e}^2(h - 1, a_{h,\eta}\,\varepsilon, \eta) + \sum_{\ell=1}^{\infty}(b_\ell^{\gamma_{d,h}})^2\,\overline{e}^2\left(h - 1, a_{h,\eta}\,\varepsilon\left(\frac{\xi_{d,h}}{b_\ell^{\gamma_{d,h}}}\right)^{1-\eta}, \eta\right)$$

$$\leqslant (a_{h,\eta}\,\varepsilon)^2\left(1 + \sum_{\ell=1}^{\infty}(b_\ell^{\gamma_{d,h}})^2\left(\frac{\xi_{d,h}}{b_\ell^{\gamma_{d,h}}}\right)^{2(1-\eta)}\right)$$

$$= (a_{h,\eta}\,\varepsilon)^2\left(1 + \xi_{d,h}^{2(1-\eta)}\sum_{\ell=1}^{\infty}(b_\ell^{\gamma_{d,h}})^{2\eta}\right)$$

$$= (a_{h,\eta}\,\varepsilon)^2\left(1 + \xi_{d,h}^{2(1-\eta)}\gamma_{d,h}^\eta\,C^{2\eta}\frac{D^{2\eta}}{1 - D^{2\eta}}\right) = \frac{(a_{h,\eta}\,\varepsilon)^2}{a_{h,\eta}^2} = \varepsilon^2.$$

We now analyze the cost of $q_{d,\varepsilon,\eta}^{(\mathrm{WW})}$. As per Definition 1, the cost of the incremental rule $\Delta_j = \bigotimes_{k=1}^d \delta_{j_k}^{(k)}$ is $\nu_j := \prod_{k=1}^d \nu_{j_k}$, and therefore

$$\mathrm{cost}(q_{d,\varepsilon,\eta}^{(\mathrm{WW})}) \quad = c(d,\varepsilon,\eta) := \begin{cases} \sum_{b(d,j) \leqslant \xi(d,j)\ \alpha(d,\varepsilon,\eta)} \prod_{k=1}^d \nu_{j_k}, & \varepsilon < 1, \\ 0, & \varepsilon \geqslant 1, \end{cases}$$

with $\alpha(d, \varepsilon, \eta)$ as per (28). For $h \in \{1, \ldots, d\}$, define

$$\overline{c}(h,\varepsilon,\eta) := \begin{cases} \sum_{\overline{b}(h,j) \leqslant \overline{\xi}(h,j)\ \overline{\alpha}(h,\varepsilon,\eta)} \prod_{k=1}^h \nu_{j_k}, & \varepsilon < 1, \\ 0, & \varepsilon \geqslant 1, \end{cases}$$

with $\overline{b}$, $\overline{\xi}$ and $\overline{\alpha}$ as per (29). Clearly, $\overline{c}(d, \varepsilon, \eta) = c(d, \varepsilon, \eta)$.

For $h = 1$, we use N_ε^* as per (30), so for $\varepsilon < 1$ we have

$$\overline{c}(1, \varepsilon, \eta) = \sum_{j=0}^{N_\varepsilon^* - 1} \nu_j.$$

The condition (25) then results in the bound

$$
\overline{c}(1,\varepsilon,\eta) \leqslant \sum_{j=0}^{N_\varepsilon^*-1} D^{-j\rho} = \frac{D^{-\rho N_\varepsilon^*}-1}{D^{-\rho}-1} = D^\rho \frac{D^{-\rho N_\varepsilon^*}-1}{1-D^\rho}
$$

$$
\leqslant \frac{D^{-\rho(N_\varepsilon^*-1)}}{1-D^\rho}. \tag{35}
$$

From (30), we have

$$
N_\varepsilon^* = \left| \{j \mid \overline{b}(1,j) > \overline{\xi}(1,j)\,\overline{\alpha}(1,\varepsilon,\eta)\} \right|
$$
$$
= \left| \{j \mid \overline{b}(1,j) > \xi_{d,1}^{1-\partial_{0,j}}\,\overline{\alpha}(1,\varepsilon,\eta)\} \right|.
$$

Using (29), we see that N_ε^* is the smallest integer $j > 0$ such that

$$
\sqrt{\gamma_{d,1}}\,CD^j \leqslant \xi_{d,1}\,\overline{\alpha}(1,\varepsilon,\eta).
$$

This, in turn, implies that either $N_\varepsilon^* = 1$, or, using the definitions of $\overline{\alpha}$ and $\overline{C_1}$ from (29),

$$
D^{N_\varepsilon^*-1} > \frac{\xi_{d,1}}{\sqrt{\gamma_{d,1}}\,C}\,\overline{\alpha}(1,\varepsilon,\eta) = \frac{\xi_{d,1}}{\sqrt{\gamma_{d,1}}\,C}\left(\varepsilon/\overline{C_1}(1,\eta)\right)^{1/(1-\eta)}
$$
$$
= \frac{(1-D^2)^{1/(2-2\eta)}}{\sqrt{\gamma_{d,1}}}\,\varepsilon^{1/(1-\eta)}.
$$

If $N_\varepsilon^* = 1$, then since $\xi_{d,1} \geqslant \sqrt{1-D^2}$, it holds that

$$
1 = D^{N_\varepsilon^*-1} > \overline{\alpha}(1,\varepsilon,\eta) = \frac{(1-D^2)^{1/(2-2\eta)}}{\xi_{d,1}}\,\varepsilon^{1/(1-\eta)}.
$$

Let

$$
C_1^* := (1-D^2)^{1/(2-2\eta)}\,\min\left(\frac{1}{\sqrt{\gamma_{d,1}}},\frac{1}{\xi_{d,1}}\right),
$$

so that $D^{N_\varepsilon^*-1} > C_1^*\varepsilon^{1/(1-\eta)}$. The bound (35) then becomes

$$
\overline{c}(1,\varepsilon,\eta) \leqslant \frac{D^{-\rho(N_\varepsilon^*-1)}}{1-D^\rho} < C_1(1/\varepsilon)^{\rho/(1-\eta)},
$$

where

$$
C_1 := \frac{(C_1^*)^{-\rho}}{1-D^\rho} = \frac{\max(\sqrt{\gamma_{d,1}},\xi_{d,1})^\rho}{(1-D^\rho)(1-D^2)^{\rho/(2-2\eta)}}. \tag{36}
$$

For $h \geqslant 2$ assume that $\overline{c}(h-1,\varepsilon,\eta) \leqslant C_{h-1}(\rho/\varepsilon)^{1/(1-\eta)}$ for some $C_{h-1} > 0$. An argument similar to that used to establish (32) shows that

$$
\overline{c}(h,\varepsilon,\eta) = \sum_{\ell=0}^{\infty} \nu_\ell \sum_{\overline{b}(h-1,j)\leqslant\overline{\xi}(h-1,j)\,\overline{\beta}_\ell(h,\varepsilon,\eta)} \prod_{k=1}^{h-1} \nu_{j_k}
$$

with $\overline{b}$ and $\overline{\xi}$ as per (29) and $\overline{\beta}_\ell$ as per (31). Using (33) and (34), we see that

$$
\overline{c}(h,\varepsilon,\eta) = \overline{c}(h-1, a_{h,\eta}\,\varepsilon, \eta) + \sum_{\ell=1}^{\overline{g}(h,\varepsilon,\eta)} \nu_\ell\, \overline{c}\left(h-1, a_{h,\eta}\,\varepsilon\left(\frac{\xi_{d,h}}{C\sqrt{\gamma_{d,h}}\,D^\ell}\right)^{1-\eta}, \eta\right),
$$

(37)

where $\overline{g}(h,\varepsilon,\eta)$ is the largest integer j such that

$$
C\sqrt{\gamma_{d,h}}\,D^j > \xi_{d,h}\,\overline{\alpha}(h,\varepsilon,\eta),
$$

and therefore

$$
\overline{g}(h,\varepsilon,\eta) = \left\lfloor \frac{\log\left(C\sqrt{\gamma_{d,h}}/\left(\xi_{d,h}\,\overline{\alpha}(h,\varepsilon,\eta)\right)\right)}{\log(D^{-1})} \right\rfloor_+.
$$

From (29) we know that

$$
1/\overline{\alpha}(h,\varepsilon,\eta) = \overline{C_1}(h,\eta)^{1/(1-\eta)}\varepsilon^{-1/(1-\eta)}.
$$

From (29) and (33) it also follows that

$$
\overline{C_1}(h,\eta) = \frac{\xi_{d,1}^{1-\eta}}{\sqrt{1-D^2}\prod_{k=2}^{h} a_{k,\eta}}.
$$

From (20) and (33), we see that $a_{h,\eta}^{-1/(1-\eta)} = f_{h,\eta}$, and therefore

$$
1/\overline{\alpha}(h,\varepsilon,\eta) = \xi_{d,1}\,(1-D^2)^{-1/(2-2\eta)}\left(\prod_{k=2}^{h} a_{k,\eta}^{-1/(1-\eta)}\right)\varepsilon^{-1/(1-\eta)}
$$

$$
= \xi_{d,1}\,(1-D^2)^{-1/(2-2\eta)}\left(\prod_{k=2}^{h} f_{k,\eta}\right)\varepsilon^{-1/(1-\eta)}.
$$

Therefore

$$
\overline{g}(h,\varepsilon,\eta) = \left\lfloor \frac{\log\left(\frac{C\gamma_{d,h}^{1/2}}{(1-D^2)^{1/(2-2\eta)}}\,\frac{\xi_{d,1}}{\xi_{d,h}}\left(\prod_{k=2}^{h} f_{k,\eta}\right)\left(\frac{1}{\varepsilon}\right)^{1/(1-\eta)}\right)}{\log\left(D^{-1}\right)} \right\rfloor_+
$$

$$
= g(h,\varepsilon,\eta) \quad \text{as per (21)}.
$$

From the inductive assumption and the sum (37), it follows that

$$
\overline{c}(h,\varepsilon,\eta) \leqslant C_{h-1}(a_{h,\eta}\,\varepsilon)^{-\rho/(1-\eta)}\left(1+\sum_{\ell=1}^{\overline{g}(h,\varepsilon,\eta)} \nu_\ell\left(\frac{C\sqrt{\gamma_{d,h}}\,D^\ell}{\xi_{d,h}}\right)^\rho\right)
$$

$$
= C_{h-1}(a_{h,\eta}\,\varepsilon)^{-\rho/(1-\eta)}\left(1+\frac{C^\rho\gamma_{d,h}^{\rho/2}}{\xi_{d,h}^\rho}\sum_{\ell=1}^{g(h,\varepsilon,\eta)} \nu_\ell\,D^{\ell\rho}\right).
$$

Therefore the bound (25) implies the bound

$$\overline{c}(h,\varepsilon,\eta) \leqslant C_{h-1}(a_{h,\eta}\,\varepsilon)^{-\rho/(1-\eta)}\left(1 + \frac{C^\rho \gamma_{d,h}^{\rho/2}}{\xi_{d,h}^\rho}\,g(h,\varepsilon,\eta)\right) = C_h\ \left(\frac{1}{\varepsilon}\right)^{\rho/(1-\eta)},$$

where (using (36))

$$C_h := C_{h-1}\,a_{h,\eta}^{-\rho/(1-\eta)}\left(1 + \frac{C^\rho \gamma_{d,h}^{\rho/2}}{\xi_{d,h}^\rho}\,g(h,\varepsilon,\eta)\right)$$

$$= C_{h-1}\left(1 + \frac{C^\rho \gamma_{d,h}^{\rho/2}}{\xi_{d,h}^\rho}\,g(h,\varepsilon,\eta)\right)f_{h,\eta}^\rho.$$

By induction, we therefore have

$$C_d = C_1 \prod_{k=2}^d \left(1 + \frac{C^\rho \gamma_{d,k}^{\rho/2}}{\xi_{d,k}^\rho}\,g(k,\varepsilon,\eta)\right)f_{k,\eta}^\rho$$

$$= \frac{\max(\sqrt{\gamma_{d,1}},\xi_{d,1})^\rho \prod\limits_{k=2}^d \left(1 + \frac{C^\rho \gamma_{d,k}^{\rho/2}}{\xi_{d,k}^\rho}\,g(k,\varepsilon,\eta)\right)f_{k,\eta}^\rho}{(1-D^\rho)(1-D^2)^{\rho/(2-2\eta)}}$$

$$= C(d,\varepsilon,\eta), \quad \text{as per (19).} \qquad \square$$

Acknowledgements

References

1. K. Altmann. Space-time distortion of laser pulses due to multiple scattering in particulate media. *Applied Optics*, 27:2451–2460, 1988.
2. A. Bondarenko, D. Radchenko, and M. Viazovska. Optimal asymptotic bounds for spherical designs. *Annals of Mathematics*, 178:1–10, 2013.
3. X. Chen, A. Frommer, and B. Lang. Computational existence proofs for spherical *t*-designs. *Numerische Mathematik*, 117:289–305, 2011.
4. X. Chen and R. Womersley. Existence of solutions to underdetermined equations and spherical designs. *SIAM Journal of Numerical Analysis*, 44(6):2326–2341, 2006.
5. G. B. Dantzig. Discrete-variable extremum problems. *Operations Research*, 5(2):266–277, April 1957.

6. B. A. Davey and H. A. Priestley. *Introduction to Lattices and Order*. Cambridge University Press, Cambridge, 1990.
7. P. Delsarte, J. M. Goethals, and J. J. Seidel. Spherical codes and designs. *Geometriae Dedicata*, 6:363–388, 1977.
8. T. Gerstner and M. Griebel. Dimension-adaptive tensor product quadrature. *Computing*, 71:65–87, 2003.
9. M. Griebel. Adaptive sparse grid multilevel methods for elliptic PDEs based on finite differences. *Computing*, 61(2):151–179, 1998.
10. M. Griebel and S. Knapek. Optimized tensor-product approximation spaces. *Constructive Approximation*, 16(4):525–540, 2000.
11. M. Griebel and S. Knapek. Optimized general sparse grid approximation spaces for operator equations. *Mathematics of Computations*, 78(268):2223–2257, October 2009. Also available as SFB611 preprint No 402.
12. M. Hegland. Adaptive sparse grids. *ANZIAM Journal*, 44 (E):C335–C353, 2003.
13. M. Hegland and P. Leopardi. The rate of convergence of sparse grid quadrature on the torus. In W. McLean and A. J. Roberts, editors, *Proceedings of the 15th Biennial Computational Techniques and Applications Conference, CTAC-2010*, volume 52 of *ANZIAM J.*, pages C500–C517, July 2011.
14. M. Hegland and P. C. Leopardi. The rate of convergence of sparse grid quadrature on products of spheres. Presented at MCQMC 2010, Warsaw, August 2010.
15. K. Hesse, F. Y. Kuo, and I. H. Sloan. A component-by-component approach to efficient numerical integration over products of spheres. *Journal of Complexity*, 23(1):25–51, February 2007.
16. F. J. Hickernell and H. Woźniakowski. Tractability of multivariate integration for periodic functions. *Journal of Complexity*, 17(4):660–682, 2001.
17. D. W. Jones. An empirical comparison of priority-queue and event-set implementations. *Communications of the ACM*, 29(4):300–311, 1986.
18. B. Kaplan, G. Ledanois, and B. Drévillon. Mueller matrix of dense polystyrene latex sphere suspensions: Measurements and Monte Carlo simulation. *Applied Optics*, 40:2769–2777, 2001.
19. F. Y. Kuo and I. H. Sloan. Quasi-Monte Carlo methods can be efficient for integration over products of spheres. *Journal of Complexity*, 21(2):196–210, 2005.
20. S. Martello and P. Toth. *Knapsack problems*. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons Ltd., Chichester, 1990.
21. H. Sato. Is the single scattering model invalid for the coda excitation at long lapse time? *Pure And Applied Geophysics*, 128:43–47, 1988.
22. I. H. Sloan and H. Woźniakowski. Tractability of multivariate integration for weighted Korobov classes. *Journal of Complexity*, 17(4):697–721, 2001.
23. A. V. Starkov, M. Noormohammadian, and U. G. Oppel. A stochastic model and a variance-reduction Monte-Carlo method for the calculation of light transport. *Applied Physics B: Lasers and Optics*, 60:335–340, 1995.
24. G. H. Vineyard. Multiple scattering of neutrons. *Physical Review*, 96(1):93–98, October 1954.
25. G. W. Wasilkowski and H. Woźniakowski. Weighted tensor product algorithms for linear multivariate problems. *Journal of Complexity*, 15(3):402–447, 1999.
26. R. S. Womersley. Spherical designs with close to the minimal number of points. Applied Mathematics Report AMR09/26, The University of New South Wales, October 2009. Preprint submitted to Elsevier.
27. S. Zakowicz, Z. Harman, N. Grün, and W. Scheid. Angular distribution of hypersatellite and satellite radiation emitted after resonant transfer and excitation into u^{91+} ions. *Physical Review A*, 68(4):042711, October 2003.