Available online at www.sciencedirect.com

ScienceDirect

Journal of Complexity 24 (2008) 283–323

Journal of COMPLEXITY

www.elsevier.com/locate/jco

Lattice rule algorithms for multivariate approximation in the average case setting

Frances Y. Kuo[a,*], Ian H. Sloan[a], Henryk Woźniakowski[b,c]

[a] *School of Mathematics and Statistics, University of New South Wales, Sydney NSW 2052, Australia*

[b] *Department of Computer Science, Columbia University, New York, NY 10027, USA*

[c] *Institute of Applied Mathematics and Mechanics, University of Warsaw, ul. Banacha 2, 02-097 Warszawa, Poland*

Received 3 April 2006; accepted 26 October 2006

Available online 22 April 2007

Abstract

We study multivariate approximation for continuous functions in the average case setting. The space of d variate continuous functions is equipped with the zero mean Gaussian measure whose covariance function is the reproducing kernel of a weighted Korobov space with the smoothness parameter $\alpha > 1$ and weights $\gamma_{d,j}$ for $j = 1, 2, \ldots, d$. The weight $\gamma_{d,j}$ moderates the behavior of functions with respect to the jth variable, and small $\gamma_{d,j}$ means that functions depend weakly on the jth variable. We study lattice rule algorithms which approximate the Fourier coefficients of a function based on function values at lattice sample points. The generating vector for these lattice points is constructed by the component-by-component algorithm, and it is tailored for the approximation problem. Our main interest is when d is large, and we study tractability and strong tractability of multivariate approximation. That is, we want to reduce the initial average case error by a factor ε by using a polynomial number of function values in ε^{-1} and d in the case of tractability, and only polynomial in ε^{-1} in the case of strong tractability. Necessary and sufficient conditions on tractability and strong tractability are obtained by applying known general tractability results for the class of arbitrary linear functionals and for the class of function values. Strong tractability holds for the two classes in the average case setting iff $\sup_{d\geqslant 1}\sum_{j=1}^{d}\gamma_{d,j}^{s} < \infty$ for some positive $s < 1$, and tractability holds iff $\sup_{d\geqslant 1}\sum_{j=1}^{d}\gamma_{d,j}^{t}/\log(d+1) < \infty$ for some positive $t < 1$. The previous results for the class of function values have been non-constructive. We provide a construction in this paper and prove tractability and strong tractability error bounds for lattice rule algorithms. This paper can be viewed as a continuation of our previous paper where we studied multivariate approximation for weighted Korobov spaces in the worst case setting. Many technical results from that paper are also useful for the average case setting. The exponents of ε^{-1} and d corresponding to our error bounds are not sharp.

*Corresponding author. Fax: +61 2 93857123.

E-mail addresses: f.kuo@unsw.edu.au (F.Y. Kuo), i.sloan@unsw.edu.au (I.H. Sloan), henryk@cs.columbia.edu (H. Woźniakowski).

0885-064X/$ - see front matter © 2007 Published by Elsevier Inc.

doi:10.1016/j.jco.2006.10.006

However, for α close to 1 and for slow decaying weights, we obtain almost the minimal exponent of ε^{-1}. We also compare the results from the worst case and the average case settings in weighted Korobov spaces.

Keywords: Multivariate approximation; Lattice rule algorithm; Average case error; Tractability

1. Introduction

We are interested in approximating real-valued continuous functions defined on the cube $[0, 1]^d$. The number of variables is d, and $d = 1, 2, \ldots$. Our main interest is when d is large, even in the hundreds or thousands, as is the case for many applications in mathematical finance, quantum physics and chemistry, see [16].

We assume that the space of continuous functions is equipped with a zero mean Gaussian measure μ_d whose covariance function is the reproducing kernel of a *weighted Korobov space* with the smoothness parameter $\alpha > 1$. In particular, the μ_d-measure of the classical Korobov space with smoothness parameter β is 1 as long as $\beta < \alpha - 1$, see Section 7. Hence, with probability 1, all functions have partial derivatives up to the order $(\alpha - 1)/2$ with respect to each variable.

The reproducing kernel depends on the non-negative weights $\gamma_{d,j}$ for $j = 1, 2, \ldots, d$. The classical Korobov space corresponds to $\gamma_{d,j} = 1$ in which case all variables play the same role. As long as d is relatively small, this is a reasonable assumption. If, however, d is large, the role of successive variables or groups of variables may be quite different. This corresponds to more general weights. In this paper we study the so-called product weights, leaving the case of general weights for future research. For product weights, we have $1 \geqslant \gamma_{d,1} \geqslant \gamma_{d,2} \geqslant \cdots \geqslant \gamma_{d,d}$, and each weight $\gamma_{d,j}$ moderates the behavior of functions with respect to the jth variable. Small $\gamma_{d,j}$ means that functions depend weakly on the jth variable, and the limiting case when $\gamma_{d,j} = 0$ means that functions are constant with respect to the jth variable.

In this paper we study *lattice rule algorithms*. Lattice rules are traditionally used for multivariate integration, see [10]. Recently, there has been a significant progress in the efficient construction of generating vectors of lattice rules for multivariate integration. The generating vectors can be constructed by the component-by-component algorithm, which works for arbitrarily large d, see [3,11–13]. The essence of this algorithm is that each component of the generating vector is computed by a one-dimensional search, with all the previous components kept unchanged. The cost of constructing an n-point lattice rule for d variables is proportional to $d\, n\, \log n$, see [8].

Lattice rules may also be used for multivariate approximation, see [4] and papers cited there. This paper can be viewed as a continuation of [4]. We study lattice rule algorithms which approximate the Fourier coefficients of functions based on function values at lattice sample points. The generating vector for these lattice points is constructed by the component-by-component algorithm, and is especially suited for multivariate approximation. The paper [4] considered the approximation problem in the worst case setting for functions from weighted Korobov spaces, while here we study the problem in the average case setting for a much larger class of functions—the space of continuous functions equipped with a zero mean Gaussian measure. Due to the assumption that the covariance function of the Gaussian measure is the reproducing kernel of the weighted Korobov space, many technical results from [4] can be applied to the average case analysis.

The second major theme of this paper is tractability and strong tractability of multivariate approximation in the average case setting. Tractability has recently become a popular research

subject in information-based complexity, see [7] and papers cited there. The essence of this study is to find necessary and sufficient conditions for which the minimal number of information evaluations needed to reduce the initial error by a factor ε is polynomial in ε^{-1} and d in the case of tractability, and only polynomial in ε^{-1} in the case of strong tractability. When these conditions hold we want to find algorithms enjoying tractability or strong tractability error bounds.

Two classes of information evaluations are typically studied. The first class Λ^{all} consists of all continuous functionals, whereas the second class Λ^{std} consists of only function values. The initial error is defined as the smallest error which can be achieved without sampling the functions. Tractability and strong tractability can be studied in various settings including the worst case and the average case settings. Most papers on tractability have been devoted to the worst case setting. The average case setting has been studied in [2], where only strong tractability is addressed, and in [1], where both strong tractability and tractability are addressed. Necessary and sufficient conditions on tractability and strong tractability are typically expressed in terms of weights of the underlying problem. Classical spaces correspond to equal weights, $\gamma_{d,j} = 1$, and in this case tractability does *not* hold, and, even worse, the minimal number of information evaluations depends *exponentially* on d. This is referred to as the *curse of dimensionality*.

Tractability can usually be obtained for sufficiently decaying weights. For multivariate approximation in the average case setting, necessary and sufficient conditions on tractability and strong tractability can be obtained from [1,2]. For both classes, Λ^{all} and Λ^{std}, strong tractability holds iff

$$\sup_{d \geqslant 1} \sum_{j=1}^{d} \gamma_{d,j}^{s} < \infty \tag{1}$$

for some positive $s < 1$. When this holds then the minimal number of evaluations is of order

$$O(\varepsilon^{-p}),$$

with the factor in the big O notation independent of ε and d but dependent on the exponent p. For both classes Λ^{all} and Λ^{std}, the exponent p can be arbitrarily close to the so-called exponent of strong tractability

$$p^* = \frac{2\max(1/\alpha, s_\gamma)}{1 - \max(1/\alpha, s_\gamma)},$$

where s_γ is the infimum of the positive numbers s for which (1) holds. Note that p^* can be arbitrarily large if α or s_γ is close to 1.

For both classes Λ^{all} and Λ^{std}, tractability holds iff

$$\sup_{d \geqslant 1} \frac{\sum_{j=1}^{d} \gamma_{d,j}^{t}}{\log(d+1)} < \infty \tag{2}$$

for some positive $t < 1$. (All logarithms in the paper are natural logarithms.) When this holds then the minimal number of evaluations is of order

$$O(d^a \varepsilon^{-p}),$$

with the factor in the big O notation independent of ε and d but dependent on the positive numbers a and p. For both classes Λ^{all} and Λ^{std}, there is a trade-off between the exponents a and p. Indeed, let $\tau \in (\max(1/\alpha, t_\gamma), 1)$, where t_γ is the infimum of the positive numbers t for which (2) holds. Then the exponents a and p can be equal to (for the class Λ^{all}) or can be arbitrarily close to

(for the class Λ^{std})

$$\frac{2\zeta(\alpha\tau)R_\tau}{1-\tau} \quad \text{and} \quad \frac{2\tau}{1-\tau},$$

respectively. Here and elsewhere in the paper $\zeta(x) = \sum_{h=1}^{\infty} h^{-x}$ for $x > 1$ is the Riemann zeta function, and $R_\tau = \limsup_{d\to\infty} \sum_{j=1}^{d} \gamma_{d,j}^{\tau}/\log(d+1)$ is finite.

Now let τ go to $\max(1/\alpha, t_\gamma)$. Then the exponent p is minimized and goes to

$$p^{**} = \frac{2\max(1/\alpha, t_\gamma)}{1-\max(1/\alpha, t_\gamma)},$$

but the exponent a goes to infinity when $1/\alpha \geqslant t_\gamma$ or when $1/\alpha < t_\gamma$ and R_τ goes to infinity. Obviously, for given d and ε we should choose τ such that $d^a\varepsilon^{-p}$ is minimized. We stress that in the case of tractability, the minimal exponent p can be arbitrarily large if α or t_γ is close to 1.

The results for the class Λ^{all} are constructive, whereas the proof presented in [2] for the class Λ^{std} uses a non-constructive argument and therefore the results are non-constructive. We provide a construction for the class Λ^{std} in this paper. This construction is based on lattice rule algorithms. We prove that the average case errors of lattice rule algorithms achieve tractability or strong tractability error bounds. We have the following results.

For lattice rule algorithms with the generating vector especially constructed for multivariate approximation in the average case setting, we prove in Theorem 10 that the number of function values required to reduce the average case error from its initial value by a factor of ε is of order

$$\begin{array}{ll} O\left(\varepsilon^{-[p+2p/(2+p)]}\right) & \text{in the case of strong tractability,} \\ O\left(d^a\varepsilon^{-[p+2p/(2+p)]}\right) & \text{in the case of tractability,} \end{array}$$

with the factors in the big O notation independent of ε and d but dependent on p, and on a and p, respectively. Here p can be, as before, arbitrarily close to p^* or p^{**}, depending on whether strong tractability or tractability holds, and there is a trade-off between p and a, see the second part of Theorem 10.

Clearly these bounds are not optimal. As already mentioned p can be very large when α, s_γ or t_γ is close to 1. In this case, the extra term $2p/(2+p)$, which is always less than 2, is insignificant since $p + 2p/(2+p) = p[1 + 2/(2+p)] \approx p$. On the other hand, p is small for large α and small s_γ and t_γ. In this case, the extra term is more important but does not cause serious concern since $p + 2p/(2+p) = p[1 + 2/(2+p)] \leqslant 2p$. In other words, for a hard problem we have essentially the (already large) minimal exponent of ε^{-1}. When a problem is easy we are at worst doubling the (small) minimal exponent of ε^{-1}.

As a comparison we also consider the generating vector constructed for multivariate integration. We prove in Theorem 14 that the number of function values required is of order

$$\begin{array}{ll} O\left(\varepsilon^{-2p}\right) & \text{in the case of strong tractability,} \\ O\left(d^a\varepsilon^{-2p}\right) & \text{in the case of tractability,} \end{array}$$

with the factors in the big O notation independent of ε and d but dependent on p, and on a and p, respectively. Here p can be arbitrarily close to p^* or p^{**}, depending on whether strong tractability or tractability holds, and again there is a trade-off between p and a, see the second part of Theorem 14. The essence of these estimates is that we roughly double the exponent of ε^{-1} as compared to the minimal one. Thus, not surprisingly the generating vector constructed especially for approximation is better than the vector constructed for integration.

In Section 7 we compare the worst case and the average case settings for multivariate approximation defined over weighted Korobov spaces. We show how much easier the average case setting is, and show that much smaller tractability and strong tractability exponents can be obtained in the average case setting, compared to the worst case setting. Finally, in Section 8 we discuss the implementation issues associated with the construction of generating vectors and present numerical results.

2. Formulation of the problem

We consider the approximation problem in the average case setting for the space $\mathcal{F} := C([0,1]^d)$ of continuous real functions defined on $[0,1]^d$. *Multivariate approximation*, or simply *approximation*, is defined in terms of the operator which is the embedding from $\mathcal{F}$ to the space $\mathcal{G} := L_2([0,1]^d)$, i.e., $\mathrm{EMB}_d : \mathcal{F} \to \mathcal{G}$ is given by

$$\mathrm{EMB}_d\, f = f.$$

We assume that the space $\mathcal{F}$ is equipped with a Gaussian probability measure μ_d whose mean element is zero and whose covariance function is

$$\int_{\mathcal{F}} f(\boldsymbol{x}) f(\boldsymbol{y}) \mu_d(\mathrm{d}f) = K_{d,\alpha,\boldsymbol{\gamma}_d}(\boldsymbol{x},\boldsymbol{y}) \quad \forall \boldsymbol{x},\boldsymbol{y} \in [0,1]^d, \tag{3}$$

where $K_{d,\alpha,\boldsymbol{\gamma}_d}(\boldsymbol{x},\boldsymbol{y})$ is the reproducing kernel for the weighted Korobov space with smoothness parameter $\alpha > 1$. The reader is referred to [17] for Gaussian measures and to [6,14] for weighted Korobov spaces. More precisely, the reproducing kernel $K_{d,\alpha,\boldsymbol{\gamma}_d}$ has the form

$$K_{d,\alpha,\boldsymbol{\gamma}_d}(\boldsymbol{x},\boldsymbol{y}) = \sum_{\boldsymbol{h}\in\mathbb{Z}^d} \frac{e^{2\pi i \boldsymbol{h}\cdot(\boldsymbol{x}-\boldsymbol{y})}}{r_d(\alpha,\boldsymbol{\gamma}_d,\boldsymbol{h})}, \quad i = \sqrt{-1}, \tag{4}$$

where $\boldsymbol{\gamma}_d = (\gamma_{d,1}, \gamma_{d,2}, \ldots, \gamma_{d,d})$ is a vector of positive [1] *weights* satisfying

$$1 \geqslant \gamma_{d,1} \geqslant \gamma_{d,2} \geqslant \cdots \geqslant \gamma_{d,d} > 0,$$

and

$$r_d(\alpha,\boldsymbol{\gamma}_d,\boldsymbol{h}) = \prod_{j=1}^{d} r(\alpha,\gamma_{d,j},h_j) \quad \text{with} \quad r(\alpha,\gamma_{d,j},h_j) = \begin{cases} 1 & \text{if } h_j = 0, \\ \gamma_{d,j}^{-1}|h_j|^{\alpha} & \text{otherwise.} \end{cases} \tag{5}$$

Note that $r_d(\alpha,\boldsymbol{\gamma}_d,\boldsymbol{h}) = r_d(\alpha,\boldsymbol{\gamma}_d,-\boldsymbol{h})$ and therefore $K_{d,\alpha,\boldsymbol{\gamma}_d}$ takes real values and can be rewritten as

$$K_{d,\alpha,\boldsymbol{\gamma}_d}(\boldsymbol{x},\boldsymbol{y}) = \sum_{\boldsymbol{h}\in\mathbb{Z}^d} \frac{\cos(2\pi\boldsymbol{h}\cdot(\boldsymbol{x}-\boldsymbol{y}))}{r_d(\alpha,\boldsymbol{\gamma}_d,\boldsymbol{h})}.$$

To simplify our notation, we write $r_d(\boldsymbol{h}) := r_d(\alpha,\boldsymbol{\gamma}_d,\boldsymbol{h})$ from this point on, except when there is a need to show the dependence on α or $\boldsymbol{\gamma}_d$.

[1] The zero weight $\gamma_{d,j} = 0$ can be treated as the limiting case of $\gamma_{d,j}$ tending to zero.

Let $\nu_d = \mu_d \mathrm{EMB}_d^{-1}$. Then ν_d is a Gaussian measure on the whole [2] space $\mathcal{G}$, with mean element zero and with the covariance operator C_{ν_d} given by

$$(C_{\nu_d} f)(\boldsymbol{x}) = \int_{[0,1]^d} K_{d,\alpha,\gamma_d}(\boldsymbol{x},\boldsymbol{y}) f(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}, \tag{6}$$

see e.g., [15, pp. 218–222]. It is easy to obtain the eigenpairs of the operator C_{ν_d}. We have

$$C_{\nu_d} \cos(2\pi \boldsymbol{h}\cdot\boldsymbol{x}) = \frac{1}{r_d(\boldsymbol{h})} \cos(2\pi \boldsymbol{h}\cdot\boldsymbol{x}) \quad \text{for all } \boldsymbol{h} \in \mathbb{Z}^d,$$

$$C_{\nu_d} \sin(2\pi \boldsymbol{h}\cdot\boldsymbol{x}) = \frac{1}{r_d(\boldsymbol{h})} \sin(2\pi \boldsymbol{h}\cdot\boldsymbol{x}) \quad \text{for all non-zero } \boldsymbol{h} \in \mathbb{Z}^d.$$

Hence, the normalized eigenfunction corresponding to the eigenvalue $1/r_d(\boldsymbol{0}) = 1$ is just the function 1, and for $\boldsymbol{h} \neq \boldsymbol{0}$ the two normalized eigenfunctions corresponding to the double eigenvalue $1/r_d(\boldsymbol{h}) = 1/r_d(-\boldsymbol{h})$ are $\sqrt{2}\cos(2\pi \boldsymbol{h}\cdot\boldsymbol{x})$ and $\sqrt{2}\sin(2\pi \boldsymbol{h}\cdot\boldsymbol{x})$. The eigenfunctions are orthonormalized in $\mathcal{G}$, and additionally the L_∞ norms of all the eigenfunctions are uniformly bounded by $\sqrt{2}$. The last property will be needed later.

It is convenient to label the vectors $\boldsymbol{h} = \boldsymbol{h}^{(i)} \in \mathbb{Z}^d$ so that the corresponding eigenvalues of C_{ν_d} are in non-increasing order, i.e.,

$$\lambda_{d,i} = \frac{1}{r_d(\boldsymbol{h}^{(i)})} \quad \text{with } 1 = \lambda_{d,1} \geqslant \lambda_{d,2} \geqslant \cdots > 0. \tag{7}$$

Clearly such labeling is not unique since we can have repeated eigenvalues.

Without loss of generality, see [15, Chapter 6], we approximate f by linear algorithms

$$A_{n,d}(f) = \sum_{k=1}^{n} a_k L_k(f), \tag{8}$$

where L_k belongs either to the class Λ^{all} of all continuous linear functionals or to the class Λ^{std} of function evaluations. The *average case error* of the algorithm $A_{n,d}$ is defined as

$$e^{\mathrm{avg}}(A_{n,d}) := \left(\int_{\mathcal{F}} \|f - A_{n,d}(f)\|_{\mathcal{G}}^2 \mu_d(\mathrm{d}f) \right)^{1/2},$$

and the *initial error* associated with $A_{0,d} = 0$ is

$$e_{0,d}^{\mathrm{avg}} := \left(\int_{\mathcal{F}} \|f\|_{\mathcal{G}}^2 \mu_d(\mathrm{d}f) \right)^{1/2} = \left(\int_{[0,1]^d} K_{d,\alpha,\gamma_d}(\boldsymbol{x},\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \right)^{1/2}$$
$$= \left(\sum_{\boldsymbol{h}\in\mathbb{Z}^d} \frac{1}{r_d(\boldsymbol{h})} \right)^{1/2} = \prod_{j=1}^{d} \left(1 + 2\zeta(\alpha)\gamma_{d,j}\right)^{1/2}.$$

Hence, the square of the initial error is equal to the sum of the eigenvalues of the operator C_{ν_d}, and

$$e_{0,d}^{\mathrm{avg}} = \left(\sum_{i=1}^{\infty} \lambda_{d,i} \right)^{1/2} = \prod_{j=1}^{d} \left(1 + 2\zeta(\alpha)\gamma_{d,j}\right)^{1/2}.$$

[2] The inverse operator EMB_d^{-1} is defined for an arbitrary Borel set A of $\mathcal{G}$ by $\mathrm{EMB}_d^{-1} A = \{f \in \mathcal{F} : \mathrm{EMB}_d f \in A\}$.

Let $\Lambda \in \{\Lambda^{\text{all}}, \Lambda^{\text{std}}\}$. For $\varepsilon \in (0, 1)$ and $d \geqslant 1$, define

$$n^{\text{avg}}(\varepsilon, d, \Lambda) := \min\{n : \exists A_{n,d} \text{ with } L_k \in \Lambda \text{ such that } e^{\text{avg}}(A_{n,d}) \leqslant \varepsilon e^{\text{avg}}_{0,d}\}$$

as the minimal number of evaluations from the class Λ which is needed to reduce the initial error in the average case setting by a factor ε. The approximation problem in the average case setting is *tractable* in the class Λ iff

$$n^{\text{avg}}(\varepsilon, d, \Lambda) \leqslant C\varepsilon^{-p}\, d^a \quad \forall d = 1, 2, \ldots, \forall \varepsilon \in (0, 1), \tag{9}$$

where C, p and a are non-negative numbers independent of ε and d. The approximation problem is *strongly tractable* if (9) holds with $a = 0$. In this case, the infimum of p from (9) is called the *exponent of strong tractability*, and is denoted by $p^{\text{avg}}(\Lambda)$.

In Theorem 1, we summarize known results on tractability and strong tractability in the average case setting with our specific covariance kernel $K_{d,\alpha,\boldsymbol{\gamma}_d}$. This theorem can be derived from already established results in [1,2]. The details of the derivation will be given in the next section.

Theorem 1. *Consider multivariate approximation* $\text{EMB}_d : \mathcal{F} \to \mathcal{G}$ *in the average case setting, where* $\mathcal{F}$ *is equipped with a zero mean Gaussian measure whose covariance kernel is* $K_{d,\alpha,\boldsymbol{\gamma}_d}$ *given by* (4)*, with the smoothness parameter* $\alpha > 1$*. For* $\boldsymbol{\gamma} = \{\boldsymbol{\gamma}_d\}_{d \geqslant 1}$ *a given infinite sequence of weight vectors* $\boldsymbol{\gamma}_d = (\gamma_{d,1}, \gamma_{d,2}, \ldots, \gamma_{d,d})$ *satisfying* $1 \geqslant \gamma_{d,1} \geqslant \gamma_{d,2} \geqslant \cdots \geqslant \gamma_{d,d} > 0$*, define the sum exponents*

$$s_{\boldsymbol{\gamma}} := \inf\left\{ s > 0 : \sup_{d \geqslant 1} \sum_{j=1}^{d} \gamma_{d,j}^{s} < \infty \right\},$$

$$t_{\boldsymbol{\gamma}} := \inf\left\{ t > 0 : \sup_{d \geqslant 1} \frac{\sum_{j=1}^{d} \gamma_{d,j}^{t}}{\log(d+1)} < \infty \right\},$$

with the convention that $\inf \emptyset = \infty$*. For* $\tau > 0$*, define also*

$$R_\tau := \limsup_{d \to \infty} \frac{\sum_{j=1}^{d} \gamma_{d,j}^{\tau}}{\log(d+1)}.$$

(a) *The approximation problem in the average case setting is strongly tractable in either class* Λ^{all} *or* Λ^{std} *iff* $s_{\boldsymbol{\gamma}} < 1$*. When this holds, then for any* $\tau \in (\max(1/\alpha, s_{\boldsymbol{\gamma}}), 1)$ *and any integer* k*, we have*

$$n^{\text{avg}}(\varepsilon, d, \Lambda^{\text{all}}) = O\left(\varepsilon^{-\frac{2\tau}{1-\tau}}\right),$$

$$n^{\text{avg}}(\varepsilon, d, \Lambda^{\text{std}}) = O\left(\varepsilon^{-\frac{2\tau}{1-\tau}[1-(1-\tau)^k]^{-1}}\right),$$

with the implied factors in the big O notation independent of ε *and* d *but dependent on* τ *and, for the class* Λ^{std}*, on* k*. The exponents of strong tractability are*

$$p^{\text{avg}}(\Lambda^{\text{all}}) = p^{\text{avg}}(\Lambda^{\text{std}}) = \frac{2\max(1/\alpha, s_{\boldsymbol{\gamma}})}{1 - \max(1/\alpha, s_{\boldsymbol{\gamma}})}.$$

(b) *The approximation problem in the average case setting is tractable in either class* Λ^{all} *or* Λ^{std} *iff* $t_\gamma < 1$. *When this holds, then for any* $\tau \in (\max(1/\alpha, t_\gamma), 1)$, *any* $a > 2\zeta(\alpha\tau)R_\tau/\tau$, *and any integer* k, *we have*

$$n^{\mathrm{avg}}(\varepsilon, d, \Lambda^{\mathrm{all}}) = O\left(d^{\frac{a\tau}{1-\tau}}\, \varepsilon^{-\frac{2\tau}{1-\tau}}\right),$$

$$n^{\mathrm{avg}}(\varepsilon, d, \Lambda^{\mathrm{std}}) = O\left(d^{\frac{a\tau}{1-\tau}[1-(1-\tau)^k]^{-1}}\, \varepsilon^{-\frac{2\tau}{1-\tau}[1-(1-\tau)^k]^{-1}}\right),$$

with the implied factors in the big O notation independent of ε *and* d *but dependent on* τ, a, *and, for the class* Λ^{std}, *on* k.

First we comment on the sum exponents s_γ and t_γ, and the numbers R_τ. Clearly $s_\gamma \geqslant t_\gamma$. In the special case where the weights do not depend on d, we have the following result.

Lemma 2. *Let* $\gamma_{d,j} = \gamma_j$ *for* $1 \leqslant j \leqslant d$ *and* $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots)$ *with* $1 \geqslant \gamma_1 \geqslant \gamma_2 \geqslant \cdots > 0$. *Suppose that* $t_\gamma < \infty$. *Then*

$$s_\gamma = t_\gamma \quad \textit{and} \quad R_\tau = 0 \ \ \forall \tau > t_\gamma.$$

Proof. Since $s_\gamma \geqslant t_\gamma$, we only need to show that $s_\gamma \leqslant t_\gamma$. For any $t > t_\gamma$, let $X := \sup_{d\geqslant 1} \sum_{j=1}^{d} \gamma_j^t / \log(d+1) < \infty$. Since the weights are non-increasing, we have

$$\frac{d\gamma_d^t}{\log(d+1)} \leqslant \frac{\sum_{j=1}^{d} \gamma_j^t}{\log(d+1)} \leqslant X \quad \forall d \geqslant 1,$$

implying

$$\gamma_d \leqslant \left(\frac{\log(d+1)}{d} X\right)^{1/t} \quad \forall d \geqslant 1.$$

Thus, for any $s > t$ we have

$$\sum_{j=1}^{\infty} \gamma_j^s \leqslant X^{s/t} \sum_{j=1}^{\infty} \left(\frac{\log(j+1)}{j}\right)^{s/t} < \infty.$$

Since s can be arbitrarily close to t, and t can be arbitrarily close to t_γ, we conclude that $s_\gamma \leqslant t_\gamma$. Hence $s_\gamma = t_\gamma$. Furthermore, for any $\tau > t_\gamma = s_\gamma$ we have $Y := \sum_{j=1}^{\infty} \gamma_j^\tau < \infty$ and thus

$$R_\tau = \limsup_{d\to\infty} \frac{\sum_{j=1}^{d} \gamma_j^\tau}{\log(d+1)} \leqslant \limsup_{d\to\infty} \frac{Y}{\log(d+1)} = 0,$$

which implies $R_\tau = 0$. □

Combining Lemma 2 and Theorem 1, we see that *tractability and strong tractability are equivalent when the weights are independent of* d.

For weights that do depend on d, we could have $t_\gamma < \infty$ and $s_\gamma \neq t_\gamma$. Also R_τ need not be zero for any $\tau > t_\gamma$. Indeed, take for example, $\gamma_{d,j} = 1$ for all $j = 1, 2, \ldots, \lceil \log(d+1) \rceil$, and

$\gamma_{d,j} = 0$ otherwise. Then we have $s_\gamma = \infty$, since for any $s > 0$,

$$\sum_{j=1}^{d} \gamma_{d,j}^{s} = \lceil \log(d+1) \rceil \to \infty \quad \text{as } d \to \infty.$$

On the other hand we have $t_\gamma = 0$ since for any $t > 0$,

$$1 \leqslant \frac{\sum_{j=1}^{d} \gamma_{d,j}^{t}}{\log(d+1)} = \frac{\lceil \log(d+1) \rceil}{\log(d+1)} < 1 + \frac{1}{\log(d+1)} \leqslant 1 + \frac{1}{\log 2} \quad \forall d \geqslant 1.$$

Clearly, in this example we have $R_\tau = 1$ for all $\tau > t_\gamma$.

We now comment on the big O bounds in Theorem 1, keeping in mind that tractability and strong tractability are equivalent if the weights do not depend on d. Assume first that $s_\gamma < 1$. Then we have strong tractability, and for all $\tau \in (\max(1/\alpha, s_\gamma), 1)$,

$$\text{the exponent of } \varepsilon^{-1} \text{ is, or is arbitrarily close to, } \frac{2\tau}{1-\tau}.$$

(For the class Λ^{std}, the exponent of ε^{-1} is arbitrarily close to $2\tau/(1-\tau)$ by taking k arbitrarily large.) We need to take τ as small as possible to minimize the exponent of ε^{-1}. As we shall see in the derivation of Theorem 1, to ensure that the factors in the big O bounds are finite we must have $\tau > 1/\alpha$ to guarantee that $\zeta(\alpha\tau) < \infty$, and $\tau > s_\gamma$ to guarantee that $\sup_{d \geqslant 1} \sum_{j=1}^{d} \gamma_{d,j}^{\tau} < \infty$. Suppose that $1/\alpha < s_\gamma < 1$. Then we can take $\tau = s_\gamma$ if we additionally assume that $\sup_{d \geqslant 1} \sum_{j=1}^{d} \gamma_{d,j}^{s_\gamma} < \infty$. In this case, the minimal exponent of ε^{-1} is or is arbitrary close to $2s_\gamma/(1-s_\gamma)$.

Now let $t_\gamma < 1$. Then we have tractability, and for all $\tau \in (\max(1/\alpha, t_\gamma), 1)$,

$$\text{the exponent of } \varepsilon^{-1} \text{ is, or is arbitrarily close to, } \frac{2\tau}{1-\tau},$$

and

$$\text{the exponent of } d \text{ is arbitrarily close to } \frac{2\zeta(\alpha\tau)R_\tau}{1-\tau}.$$

Note that we need $\tau > 1/\alpha$ to guarantee that $\zeta(\alpha\tau) < \infty$, and $\tau > t_\gamma$ is needed to guarantee $R_\tau < \infty$. Clearly, there is a trade-off between the exponents of ε^{-1} and d. Just as for strong tractability, we need τ as small as possible to minimize the exponent of ε^{-1}. On the other hand, the exponent of d can be very large if τ is small. Suppose that $1/\alpha < t_\gamma < 1$. Then we can even take $\tau = t_\gamma$ to obtain the minimal exponent of ε^{-1}, provided we assume additionally that $R_{t_\gamma} < \infty$. In this case, the exponent of ε^{-1} is, or is arbitrarily close to, $2t_\gamma/(1-t_\gamma)$, and the exponent of d is arbitrarily close to $2\zeta(\alpha t_\gamma)R_{t_\gamma}/(1-t_\gamma)$.

3. Derivation of Theorem 1

In this section we derive Theorem 1 from already established results in [1,2]. We consider the two classes Λ^{all} and Λ^{std} separately.

3.1. Tractability and strong tractability in the class $\Lambda^{\rm all}$

In the class $\Lambda^{\rm all}$, the optimal algorithm is well known and $n^{\rm avg}(\varepsilon, d, \Lambda^{\rm all})$ is fully characterized by the eigenvalues of the covariance operator C_{ν_d} given by (6), see [15, Chapter 6].

We recall that the vectors $\boldsymbol{h} = \boldsymbol{h}^{(i)} \in \mathbb{Z}^d$ are labeled such that the corresponding eigenvalues $\lambda_{d,i} = 1/r_d(\boldsymbol{h}^{(i)})$ are in non-increasing order. The optimal algorithm in the class $\Lambda^{\rm all}$ is the truncated Fourier series

$$A_{n,d}(f)(\boldsymbol{x}) := \sum_{i=1}^{n} \hat{f}(\boldsymbol{h}^{(i)}) e^{2\pi \mathrm{i} \boldsymbol{h}^{(i)} \cdot \boldsymbol{x}}.$$

The average case error of $A_{n,d}$ is

$$e^{\rm avg}(A_{n,d}) = \left(\sum_{i=n+1}^{\infty} \frac{1}{r_d(\boldsymbol{h}^{(i)})} \right)^{1/2} = \left(\sum_{i=n+1}^{\infty} \lambda_{d,i} \right)^{1/2}.$$

Here, optimality means that the algorithm has the minimal average case error among all algorithms that use at most n evaluations from $\Lambda^{\rm all}$. Thus,

$$n^{\rm avg}(\varepsilon, d, \Lambda^{\rm all}) = \min \left\{ n : \sum_{i=n+1}^{\infty} \lambda_{d,i} \leqslant \varepsilon^2 \sum_{i=1}^{\infty} \lambda_{d,i} \right\}.$$

Theorem 1 of [1] states necessary and sufficient conditions on tractability and strong tractability of the approximation problem in the class $\Lambda^{\rm all}$, see also [2] for strong tractability. Specifically, the approximation problem is strongly tractable in the class $\Lambda^{\rm all}$ iff

$$M_\tau := \sup_{d \geqslant 1} \frac{\left(\sum_{i=1}^{\infty} \lambda_{d,i}^{\tau}\right)^{1/\tau}}{\sum_{i=1}^{\infty} \lambda_{d,i}} < \infty \quad \text{for some } \tau \in (0, 1).$$

When this holds, then

$$n^{\rm avg}(\varepsilon, d, \Lambda^{\rm all}) \leqslant \left\lceil \left(\frac{\tau M_\tau}{1-\tau} \right)^{\frac{\tau}{1-\tau}} \varepsilon^{-\frac{2\tau}{1-\tau}} \right\rceil. \tag{10}$$

The exponent of strong tractability is

$$p^{\rm avg}(\Lambda^{\rm all}) = \frac{2\tau^*}{1-\tau^*},$$

where τ^* is the infimum of all numbers τ for which (10) holds. The approximation problem is tractable in the class $\Lambda^{\rm all}$ iff

$$M_{\tau,a} := \sup_{d \geqslant 1} \frac{\left(\sum_{i=1}^{\infty} \lambda_{d,i}^{\tau}\right)^{1/\tau}}{d^a \sum_{i=1}^{\infty} \lambda_{d,i}} < \infty \quad \text{for some } \tau \in (0, 1) \text{ and } a \geqslant 0.$$

When this holds, then

$$n^{\rm avg}(\varepsilon, d, \Lambda^{\rm all}) \leqslant \left\lceil \left(\frac{\tau M_{\tau,a}}{1-\tau} \right)^{\frac{\tau}{1-\tau}} d^{\frac{a\tau}{1-\tau}} \varepsilon^{-\frac{2\tau}{1-\tau}} \right\rceil. \tag{11}$$

In our current setting, we have for $\tau \in (1/\alpha, 1)$,

$$\frac{\left(\sum_{i=1}^{\infty} \lambda_{d,i}^{\tau}\right)^{1/\tau}}{\sum_{i=1}^{\infty} \lambda_{d,i}} = \prod_{j=1}^{d} \frac{\left(1 + 2\zeta(\alpha\tau)\gamma_{d,j}^{\tau}\right)^{1/\tau}}{1 + 2\zeta(\alpha)\gamma_{d,j}}, \tag{12}$$

where we made use of $[r_d(\boldsymbol{h})]^{\tau} = [r_d(\alpha, \boldsymbol{\gamma}_d, \boldsymbol{h})]^{\tau} = r_d(\alpha\tau, \boldsymbol{\gamma}_d^{\tau}, \boldsymbol{h})$ with $\boldsymbol{\gamma}_d^{\tau} := (\gamma_{d,1}^{\tau}, \gamma_{d,2}^{\tau}, \ldots, \gamma_{d,d}^{\tau})$. We now show that the right-hand side of (12) has upper and lower bounds given by

$$\exp\left(b_{\tau} \sum_{j=1}^{d} \gamma_{d,j}^{\tau}\right) \leqslant \prod_{j=1}^{d} \frac{\left(1 + 2\zeta(\alpha\tau)\gamma_{d,j}^{\tau}\right)^{1/\tau}}{1 + 2\zeta(\alpha)\gamma_{d,j}} \leqslant \exp\left(\frac{2\zeta(\alpha\tau)}{\tau} \sum_{j=1}^{d} \gamma_{d,j}^{\tau}\right), \tag{13}$$

where

$$b_{\tau} := \frac{1}{\tau} \log\left(1 + \frac{(2 - 2^{\tau})\zeta(\alpha\tau)}{1 + 2^{\tau}\zeta(\alpha\tau)}\right).$$

To prove (13) we note that $\prod_{j=1}^{d} (1+x_j) = \exp(\sum_{j=1}^{d} \log(1+x_j))$ for all $x_j \geqslant 0$. Furthermore, it can be checked that for any $x^* > 0$ we have

$$\frac{\log(1 + x^*)}{x^*} x \leqslant \log(1 + x) \leqslant x \quad \forall x \in (0, x^*]. \tag{14}$$

The upper bound in (13) follows easily by applying the upper estimate of (14) with $x = 2\zeta(\alpha\tau)\gamma_{d,j}^{\tau}$ to the numerator and estimating the denominator by 1. To prove the lower bound in (13), we use Jensen's inequality [3] in the denominator:

$$\begin{aligned} 1 + 2\zeta(\alpha)\gamma_{d,j} = 1 + 2\gamma_{d,j} \sum_{h=1}^{\infty} h^{-\alpha} &\leqslant \left(1 + 2^{\tau}\gamma_{d,j}^{\tau} \sum_{h=1}^{\infty} h^{-\alpha\tau}\right)^{1/\tau} \\ &= \left(1 + 2^{\tau}\zeta(\alpha\tau)\gamma_{d,j}^{\tau}\right)^{1/\tau}. \end{aligned}$$

Then

$$\begin{aligned} \frac{\left(1 + 2\zeta(\alpha\tau)\gamma_{d,j}^{\tau}\right)^{1/\tau}}{1 + 2\zeta(\alpha)\gamma_{d,j}} \geqslant \left(\frac{1 + 2\zeta(\alpha\tau)\gamma_{d,j}^{\tau}}{1 + 2^{\tau}\zeta(\alpha\tau)\gamma_{d,j}^{\tau}}\right)^{1/\tau} &= \left(1 + \frac{(2 - 2^{\tau})\zeta(\alpha\tau)\gamma_{d,j}^{\tau}}{1 + 2^{\tau}\zeta(\alpha\tau)\gamma_{d,j}^{\tau}}\right)^{1/\tau} \\ &\geqslant \left(1 + \frac{(2 - 2^{\tau})\zeta(\alpha\tau)}{1 + 2^{\tau}\zeta(\alpha\tau)}\gamma_{d,j}^{\tau}\right)^{1/\tau}. \end{aligned}$$

From this and the lower estimate of (14) with $x^* = (2 - 2^{\tau})\zeta(\alpha\tau)/(1 + 2^{\tau}\zeta(\alpha\tau))$, we obtain the lower bound in (13). This completes the proof of (13).

From (12) and (13) we conclude that

$$\exp\left(b_{\tau} \sup_{d \geqslant 1} \sum_{j=1}^{d} \gamma_{d,j}^{\tau}\right) \leqslant M_{\tau} \leqslant \exp\left(\frac{2\zeta(\alpha\tau)}{\tau} \sup_{d \geqslant 1} \sum_{j=1}^{d} \gamma_{d,j}^{\tau}\right).$$

[3] Jensen's inequality states that $\left(\sum_k a_k\right)^{\tau} \leqslant \sum_k a_k^{\tau}$ for any non-negative a_k and any $\tau \in (0, 1]$.

Thus, M_τ is finite for some $\tau \in (1/\alpha, 1)$ iff

$$\sup_{d \geqslant 1} \sum_{j=1}^{d} \gamma_{d,j}^{\tau} < \infty \quad \text{for some } \tau \in (1/\alpha, 1).$$

This, in turn, holds iff $s_\gamma < 1$, where s_γ is defined as in Theorem 1. Hence, strong tractability holds iff $s_\gamma < 1$. When $s_\gamma < 1$, then for $\tau \in (\max(1/\alpha, s_\gamma), 1)$ we have $M_\tau < \infty$. This, together with (10), proves the big O bound on $n^{\mathrm{avg}}(\varepsilon, d, \Lambda^{\mathrm{all}})$ in the case of strong tractability.

We now consider tractability for the class Λ^{all}. For $\tau \in (1/\alpha, 1)$ and $a \geqslant 0$, we use $e^x = (d+1)^{x/\log(d+1)}$, and conclude from (12) and (13) that

$$\sup_{d \geqslant 1} (d+1)^{b_\tau \frac{\sum_{j=1}^{d} \gamma_{d,j}^{\tau}}{\log(d+1)} - a} \leqslant M_{\tau,a} \leqslant \sup_{d \geqslant 1} \left(1 + \tfrac{1}{d}\right)^{a} (d+1)^{\frac{2\zeta(\alpha\tau)}{\tau} \frac{\sum_{j=1}^{d} \gamma_{d,j}^{\tau}}{\log(d+1)} - a}.$$

Thus, $M_{\tau,a}$ is finite for some $\tau \in (1/\alpha, 1)$ and $a \geqslant 0$ iff

$$\sup_{d \geqslant 1} \frac{\sum_{j=1}^{d} \gamma_{d,j}^{\tau}}{\log(d+1)} < \infty \quad \text{for some } \tau \in (1/\alpha, 1),$$

which holds iff $t_\gamma < 1$, where t_γ is defined as in Theorem 1. Hence, tractability holds iff $t_\gamma < 1$. When $t_\gamma < 1$, then for any $\tau \in (\max(1/\alpha, t_\gamma), 1)$ we may choose $a > 2\zeta(\alpha\tau) R_\tau/\tau$, where $R_\tau < \infty$ is defined as in Theorem 1. This choice of a ensures that the exponent of $d+1$ in the upper bound of $M_{\tau,a}$ is negative for large d, and hence $M_{\tau,a} < \infty$. This and (11) together prove the big O bound on $n^{\mathrm{avg}}(\varepsilon, d, \Lambda^{\mathrm{all}})$ in Theorem 1 in the case of tractability.

3.2. Tractability and strong tractability in the class Λ^{std}

We now turn to the class Λ^{std}. Since $\Lambda^{\mathrm{std}} \subset \Lambda^{\mathrm{all}}$, we know that $s_\gamma < 1$ is a necessary condition for strong tractability in the class Λ^{std}, and $t_\gamma < 1$ is a necessary condition for tractability in the class Λ^{std}.

Using the fact that the eigenfunctions of C_{ν_d} are uniformly bounded in the L_∞ norm[4] by $\sqrt{2}$, we find out from the proof of Theorem 1 in [2] that for any $\tau \in (\max(1/\alpha, s_\gamma), 1)$ in the case of strong tractability, and any $\tau \in (\max(1/\alpha, t_\gamma), 1)$ and any $a > 2\zeta(\alpha\tau) R_\tau/\tau$ in the case of tractability, for any positive integer k there is an algorithm $A_{kn,d}$ using kn function values whose average case error satisfies

$$e^{\mathrm{avg}}(A_{kn,d}) \leqslant \left(4^k + p_d(4^k - 1)/3\right)^{1/2} n^{-q_k} e_{0,d}^{\mathrm{avg}},$$

where

$$p_d = \begin{cases} \dfrac{\tau M_\tau}{1-\tau} & \text{in the case of strong tractability,} \\ \dfrac{\tau M_{\tau,a}}{1-\tau}\, d^a & \text{in the case of tractability,} \end{cases}$$

$$q_k = \frac{1-\tau}{2\tau}\left(1 - (1-\tau)^k\right).$$

[4] This assumption is removed in [1].

If we define

$$n = \left\lceil \left(4^k + p_d(4^k - 1)/3\right)^{1/(2q_k)} \varepsilon^{-1/q_k} \right\rceil$$

then $e^{\mathrm{avg}}(A_{kn,d}) \leqslant \varepsilon e^{\mathrm{avg}}_{0,d}$, and therefore $n^{\mathrm{avg}}(\varepsilon, d, \Lambda^{\mathrm{std}}) \leqslant kn$, which proves the big O bounds of $n^{\mathrm{avg}}(\varepsilon, d, \Lambda^{\mathrm{std}})$ in Theorem 1.

This also shows that $s_\gamma < 1$ implies strong tractability in the class Λ^{std}, and that the exponent of strong tractability in this class is the same as in the class Λ^{all}. Furthermore, we see that $t_\gamma < 1$ implies tractability in the class Λ^{std}. This completes the derivation of Theorem 1.

A closer examination of the argument above leads to the conclusion that the necessary and sufficient conditions for tractability and strong tractability remain the same if we study the absolute error $e^{\mathrm{avg}}_{n,d}$ instead of the normalized error $e^{\mathrm{avg}}_{n,d}/e^{\mathrm{avg}}_{0,d}$. When the absolute error $e^{\mathrm{avg}}_{n,d}$ is considered, the factors in the big O bounds of $n^{\mathrm{avg}}(\varepsilon, d, \Lambda)$, $\Lambda \in \{\Lambda^{\mathrm{all}}, \Lambda^{\mathrm{std}}\}$, are larger, although the exponents of d and ε^{-1} remain unchanged for both tractability and strong tractability.

We stress here that the proof in [2] was non-constructive, giving no clue as to how to find an algorithm $A_{n,d}$ that achieves the estimate of $n^{\mathrm{avg}}(\varepsilon, d, \Lambda^{\mathrm{std}})$ in Theorem 1. We provide a construction in this paper.

4. Lattice rule algorithms in the class Λ^{std}

A *rank*-1 *lattice rule*, see [10], is an equal weight integration rule which approximates the integral of a function f over the unit cube $[0, 1]^d$ by

$$\frac{1}{n} \sum_{k=1}^{n} f\left(\left\{\frac{kz}{n}\right\}\right).$$

Here z is an integer vector, known as the *generating vector*, which has no factor in common with n, and the braces around a vector indicate that each component of the vector is to be replaced by its fractional part. In recent years there have been many theoretical advances on lattice rules for multivariate integration. The most significant achievement has been the development of *component-by-component* algorithms for choosing good generating vectors that lead to the optimal rate of convergence for integrands belonging to weighted Korobov or weighted Sobolev spaces, see e.g., [3,11–13].

In this section we study *lattice rule algorithms* for multivariate approximation. By a lattice rule algorithm we mean a linear algorithm that uses function values at the lattice sample points $\{kz/n\}$ for $k = 1, 2, \ldots, n$. For simplicity we assume that n is prime and

$$z \in \mathcal{Z}_n^d := \{1, 2, \ldots, n-1\}^d.$$

4.1. The optimal lattice rule algorithm

We start by finding the optimal lattice rule algorithm $A^{(\mathrm{opt})}_{n,d}$ for a given $z \in \mathcal{Z}_n^d$. More precisely,

$$A^{(\mathrm{opt})}_{n,d}(f)(\boldsymbol{x}) := \sum_{k=1}^{n} a_k(\boldsymbol{x}) f\left(\left\{\frac{kz}{n}\right\}\right),$$

where $a_1, a_2 \ldots, a_n$ are functions from $\mathcal{G}$ chosen to minimize the average case error

$$e_{n,d}^{\text{avg (opt)}}(\boldsymbol{z}) := \left(\int_{\mathcal{F}} \|f - A_{n,d}^{(\text{opt})}(f)\|_{\mathcal{G}}^2 \mu_d(\mathrm{d}f) \right)^{1/2}.$$

Note that $e_{n,d}^{\text{avg(opt)}}(\boldsymbol{z})$ is a quadratic form in terms of the functions a_k and therefore it is possible to find its minimum by the standard technique.

Lemma 3. *Let n be prime and let $\boldsymbol{z} \in \mathcal{Z}_n^d$ be given. The optimal lattice rule algorithm $A_{n,d}^{(\text{opt})}$ has the explicit form*

$$\begin{aligned} &A_{n,d}^{(\text{opt})}(f)(\boldsymbol{x}) \\ &\quad = \sum_{k=1}^{n} \left(\frac{1}{n} \sum_{\boldsymbol{h} \in \mathbb{Z}^d} \frac{1}{1 + r_d(\boldsymbol{h}) F_d(\boldsymbol{h}, \boldsymbol{z})} e^{-2\pi \mathrm{i} k \boldsymbol{h} \cdot \boldsymbol{z}/n + 2\pi \mathrm{i} \boldsymbol{h} \cdot \boldsymbol{x}} \right) f\left(\left\{ \tfrac{k\boldsymbol{z}}{n} \right\} \right) \\ &\quad = \sum_{\boldsymbol{h} \in \mathbb{Z}^d} \left(\frac{1}{n} \frac{1}{1 + r_d(\boldsymbol{h}) F_d(\boldsymbol{h}, \boldsymbol{z})} \sum_{k=1}^{n} f\left(\left\{ \tfrac{k\boldsymbol{z}}{n} \right\} \right) e^{-2\pi \mathrm{i} k \boldsymbol{h} \cdot \boldsymbol{z}/n} \right) e^{2\pi \mathrm{i} \boldsymbol{h} \cdot \boldsymbol{x}}, \end{aligned} \tag{15}$$

where

$$F_d(\boldsymbol{h}, \boldsymbol{z}) := \sum_{\substack{\boldsymbol{\ell} \in \mathbb{Z}^d \setminus \{\boldsymbol{0}\} \\ \boldsymbol{\ell} \cdot \boldsymbol{z} \equiv 0 (\mathrm{mod}\, n)}} \frac{1}{r_d(\boldsymbol{h} + \boldsymbol{\ell})}. \tag{16}$$

The average case error of $A_{n,d}^{(\text{opt})}$ is

$$e_{n,d}^{\text{avg(opt)}}(\boldsymbol{z}) = \left(\sum_{\boldsymbol{h} \in \mathbb{Z}^d} \frac{F_d(\boldsymbol{h}, \boldsymbol{z})}{1 + r_d(\boldsymbol{h}) F_d(\boldsymbol{h}, \boldsymbol{z})} \right)^{1/2}. \tag{17}$$

Proof. For each $k = 1, 2, \ldots, n$, let $\hat{a}_k(\boldsymbol{h})$ denote the Fourier coefficient of a_k. Then the pointwise error from the approximation $A_{n,d}^{(\text{opt})}(f)$ is

$$(f - A_{n,d}^{(\text{opt})}(f))(\boldsymbol{x}) = \sum_{\boldsymbol{h} \in \mathbb{Z}^d} \left(\hat{f}(\boldsymbol{h}) - \sum_{k=1}^{n} \hat{a}_k(\boldsymbol{h}) f\left(\left\{ \tfrac{k\boldsymbol{z}}{n} \right\} \right) \right) e^{2\pi \mathrm{i} \boldsymbol{h} \cdot \boldsymbol{x}}.$$

In effect, $A_{n,d}^{(\text{opt})}(f)$ approximates each Fourier coefficient of f by

$$\hat{f}(\boldsymbol{h}) \approx \sum_{k=1}^{n} \hat{a}_k(\boldsymbol{h}) f\left(\left\{ \tfrac{k\boldsymbol{z}}{n} \right\} \right).$$

The average case error is

$$e_{n,d}^{\text{avg (opt)}}(\boldsymbol{z}) = \left(\sum_{\boldsymbol{h} \in \mathbb{Z}^d} e_{\boldsymbol{h}} \right)^{1/2}, \quad e_{\boldsymbol{h}} := \int_{\mathcal{F}} \left| \hat{f}(\boldsymbol{h}) - \sum_{k=1}^{n} \hat{a}_k(\boldsymbol{h}) f\left(\left\{ \tfrac{k\boldsymbol{z}}{n} \right\} \right) \right|^2 \mu_d(\mathrm{d}f).$$

Note that $e_{\boldsymbol{h}}$ can be viewed as the squared average case integration error in approximating the Fourier coefficient $\hat{f}(\boldsymbol{h})$. The optimal choice of $a_1, a_2, \ldots, a_n$ can be obtained by choosing the Fourier coefficients $\hat{a}_1(\boldsymbol{h}), \hat{a}_2(\boldsymbol{h}), \ldots, \hat{a}_n(\boldsymbol{h})$ to minimize $e_{\boldsymbol{h}}$ for each $\boldsymbol{h}$.

It follows from the property (3) of the covariance kernel K_{d,α,γ_d} that

$$\begin{aligned}\int_{\mathcal{F}} \hat{f}(\boldsymbol{h})\overline{\hat{f}(\boldsymbol{h})}\mu_d(\mathrm{d}f) &= \int_{\mathcal{F}}\int_{[0,1]^d}\int_{[0,1]^d} f(\boldsymbol{x})f(\boldsymbol{y})e^{-2\pi\mathrm{i}\boldsymbol{h}\cdot(\boldsymbol{x}-\boldsymbol{y})}\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{y}\,\mu_d(\mathrm{d}f)\\ &= \int_{[0,1]^d}\int_{[0,1]^d} K_{d,\alpha,\gamma_d}(\boldsymbol{x},\boldsymbol{y})e^{-2\pi\mathrm{i}\boldsymbol{h}\cdot(\boldsymbol{x}-\boldsymbol{y})}\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{y} = \frac{1}{r_d(\boldsymbol{h})},\end{aligned}$$

and

$$\begin{aligned}\int_{\mathcal{F}} \hat{f}(\boldsymbol{h})f\left(\left\{\tfrac{k\boldsymbol{z}}{n}\right\}\right)\mu_d(\mathrm{d}f) &= \int_{\mathcal{F}}\int_{[0,1]^d} f(\boldsymbol{x})f\left(\left\{\tfrac{k\boldsymbol{z}}{n}\right\}\right)e^{-2\pi\mathrm{i}\boldsymbol{h}\cdot\boldsymbol{x}}\,\mathrm{d}\boldsymbol{x}\mu_d(\mathrm{d}f)\\ &= \int_{[0,1]^d} K_{d,\alpha,\gamma_d}\left(\boldsymbol{x},\left\{\tfrac{k\boldsymbol{z}}{n}\right\}\right)e^{-2\pi\mathrm{i}\boldsymbol{h}\cdot\boldsymbol{x}}\,\mathrm{d}\boldsymbol{x} = \frac{e^{-2\pi\mathrm{i}k\boldsymbol{h}\cdot\boldsymbol{z}/n}}{r_d(\boldsymbol{h})}.\end{aligned}$$

Similarly,

$$\int_{\mathcal{F}} \overline{\hat{f}(\boldsymbol{h})}f\left(\left\{\tfrac{k\boldsymbol{z}}{n}\right\}\right)\mu_d(\mathrm{d}f) = \frac{e^{2\pi\mathrm{i}k\boldsymbol{h}\cdot\boldsymbol{z}/n}}{r_d(\boldsymbol{h})},$$

and

$$\int_{\mathcal{F}} f\left(\left\{\tfrac{k\boldsymbol{z}}{n}\right\}\right)f\left(\left\{\tfrac{t\boldsymbol{z}}{n}\right\}\right)\mu_d(\mathrm{d}f) = \sum_{\boldsymbol{p}\in\mathbb{Z}^d}\frac{e^{2\pi\mathrm{i}(k-t)\boldsymbol{p}\cdot\boldsymbol{z}/n}}{r_d(\boldsymbol{p})}.$$

Thus,

$$\begin{aligned}e_{\boldsymbol{h}} &= \frac{1}{r_d(\boldsymbol{h})} - \sum_{k=1}^{n}\hat{a}_k(\boldsymbol{h})\frac{e^{2\pi\mathrm{i}k\boldsymbol{h}\cdot\boldsymbol{z}/n}}{r_d(\boldsymbol{h})} - \sum_{k=1}^{n}\overline{\hat{a}_k(\boldsymbol{h})}\frac{e^{-2\pi\mathrm{i}k\boldsymbol{h}\cdot\boldsymbol{z}/n}}{r_d(\boldsymbol{h})}\\ &\quad + \sum_{k=1}^{n}\sum_{t=1}^{n}\hat{a}_k(\boldsymbol{h})\overline{\hat{a}_t(\boldsymbol{h})}\sum_{\boldsymbol{p}\in\mathbb{Z}^d}\frac{e^{2\pi\mathrm{i}(k-t)\boldsymbol{p}\cdot\boldsymbol{z}/n}}{r_d(\boldsymbol{p})}\\ &= \frac{1}{r_d(\boldsymbol{h})} - \boldsymbol{a}^{\mathrm{T}}\overline{\boldsymbol{b}} - \boldsymbol{b}^{\mathrm{T}}\overline{\boldsymbol{a}} + \boldsymbol{a}^{\mathrm{T}}\mathcal{K}\overline{\boldsymbol{a}},\end{aligned}$$

where $\boldsymbol{a} = [\hat{a}_t(\boldsymbol{h})]_{1\leqslant t\leqslant n}^{\mathrm{T}}$,

$$\boldsymbol{b} = \left[\frac{e^{-2\pi\mathrm{i}k\boldsymbol{h}\cdot\boldsymbol{z}/n}}{r_d(\boldsymbol{h})}\right]_{1\leqslant k\leqslant n}^{\mathrm{T}} \quad\text{and}\quad \mathcal{K} = \left[\sum_{\boldsymbol{p}\in\mathbb{Z}^d}\frac{e^{2\pi\mathrm{i}(k-t)\boldsymbol{p}\cdot\boldsymbol{z}/n}}{r_d(\boldsymbol{p})}\right]_{\substack{1\leqslant k\leqslant n\\ 1\leqslant t\leqslant n}}.$$

Since $\mathcal{K}$ is hermitian and positive definite, we see that $e_{\boldsymbol{h}}$ is minimized when $\boldsymbol{a}$ is the solution of $\boldsymbol{b} = \overline{\mathcal{K}}\boldsymbol{a}$, and the minimum is

$$e_{\boldsymbol{h}} = \frac{1}{r_d(\boldsymbol{h})} - \boldsymbol{a}^{\mathrm{T}}\overline{\boldsymbol{b}}.$$

Suppose that $\hat{a}_t(\boldsymbol{h}) = \rho(\boldsymbol{h})\, e^{-2\pi \mathrm{i} t \boldsymbol{h}\cdot \boldsymbol{z}/n}$. We check for which $\rho(\boldsymbol{h})$ we have $\boldsymbol{b} = \overline{\mathcal{K}}\boldsymbol{a}$. For each $k = 1, 2, \ldots, n$ we compute the kth component of $\overline{\mathcal{K}}\boldsymbol{a}$, getting

$$\begin{aligned}
&\sum_{t=1}^{n} \sum_{\boldsymbol{p}\in\mathbb{Z}^d} \frac{e^{-2\pi \mathrm{i}(k-t)\boldsymbol{p}\cdot\boldsymbol{z}/n}}{r_d(\boldsymbol{p})} \rho(\boldsymbol{h}) e^{-2\pi \mathrm{i} t \boldsymbol{h}\cdot\boldsymbol{z}/n} \\
&\quad = \rho(\boldsymbol{h}) \sum_{\boldsymbol{p}\in\mathbb{Z}^d} \frac{e^{-2\pi \mathrm{i} k \boldsymbol{p}\cdot\boldsymbol{z}/n}}{r_d(\boldsymbol{p})} \left(\sum_{t=1}^{n} e^{2\pi \mathrm{i} t (\boldsymbol{p}-\boldsymbol{h})\cdot\boldsymbol{z}/n} \right) = \rho(\boldsymbol{h})\, n \sum_{\substack{\boldsymbol{p}\in\mathbb{Z}^d \\ (\boldsymbol{p}-\boldsymbol{h})\cdot\boldsymbol{z}\equiv 0 (\mathrm{mod}\, n)}} \frac{e^{-2\pi \mathrm{i} k \boldsymbol{p}\cdot\boldsymbol{z}/n}}{r_d(\boldsymbol{p})} \\
&\quad = \rho(\boldsymbol{h}) n e^{-2\pi \mathrm{i} k \boldsymbol{h}\cdot\boldsymbol{z}/n} \sum_{\substack{\boldsymbol{\ell}\in\mathbb{Z}^d \\ \boldsymbol{\ell}\cdot\boldsymbol{z}\equiv 0 (\mathrm{mod}\, n)}} \frac{1}{r_d(\boldsymbol{h}+\boldsymbol{\ell})} \\
&\quad = \rho(\boldsymbol{h}) n e^{-2\pi \mathrm{i} k \boldsymbol{h}\cdot\boldsymbol{z}/n} \left(\frac{1}{r_d(\boldsymbol{h})} + F_d(\boldsymbol{h}, z) \right),
\end{aligned}$$

which equals the kth component of $\boldsymbol{b}$ if we take

$$\rho(\boldsymbol{h}) = \frac{1}{n} \frac{1}{r_d(\boldsymbol{h})} \left(\frac{1}{r_d(\boldsymbol{h})} + F_d(\boldsymbol{h}, z) \right)^{-1} = \frac{1}{n} \frac{1}{1 + r_d(\boldsymbol{h}) F_d(\boldsymbol{h}, z)}.$$

This leads to

$$\hat{a}_k(\boldsymbol{h}) = \frac{e^{-2\pi \mathrm{i} k \boldsymbol{h}\cdot\boldsymbol{z}/n}}{n} \frac{1}{1 + r_d(\boldsymbol{h}) F_d(\boldsymbol{h}, z)},$$

and

$$e_{\boldsymbol{h}} = \frac{1}{r_d(\boldsymbol{h})} \left(1 - \frac{1}{1 + r_d(\boldsymbol{h}) F_d(\boldsymbol{h}, z)} \right) = \frac{F_d(\boldsymbol{h}, z)}{1 + r_d(\boldsymbol{h}) F_d(\boldsymbol{h}, z)}.$$

The lemma follows easily from the last two expressions. □

From Lemma 3 we see that the optimal algorithm $A_{n,d}^{(\mathrm{opt})}$ approximates the Fourier coefficient $\hat{f}(\boldsymbol{h})$, or equivalently the integral of $f(\boldsymbol{x}) e^{-2\pi \mathrm{i} \boldsymbol{h}\cdot\boldsymbol{x}}$, by a lattice rule with equal integration weights

$$\frac{1}{n} \frac{1}{1 + r_d(\boldsymbol{h}) F_d(\boldsymbol{h}, z)}$$

as opposed to the typical weights $1/n$. The algorithm $A_{n,d}^{(\mathrm{opt})}$ turns out to be essentially the same as the spline algorithm considered in [19], where, instead of an average case error, a modified worst case error criterion was used.

4.2. Two lattice rule algorithms

In practice we cannot include every Fourier component as in the optimal algorithm in (15). Here we introduce two modified algorithms which include only a finite number of Fourier components.

Clearly each term inside the sum of (17) satisfies

$$\frac{F_d(\boldsymbol{h}, z)}{1 + r_d(\boldsymbol{h}) F_d(\boldsymbol{h}, z)} \leqslant \min \left(\frac{1}{r_d(\boldsymbol{h})}, F_d(\boldsymbol{h}, z) \right).$$

It makes sense to use the first bound when $r_d(\boldsymbol{h})$ is large and the second bound when $r_d(\boldsymbol{h})$ is small. This motivates the definition of the set

$$\mathcal{A}_d(M) := \{\boldsymbol{h} \in \mathbb{Z}^d : r_d(\boldsymbol{h}) \leqslant M\}. \tag{18}$$

Using (17), we can write

$$e_{n,d}^{\mathrm{avg(opt)}}(\boldsymbol{z}) \leqslant \left(\sum_{\boldsymbol{h} \notin \mathcal{A}_d(M)} \frac{1}{r_d(\boldsymbol{h})} + \sum_{\boldsymbol{h} \in \mathcal{A}_d(M)} \frac{F_d(\boldsymbol{h}, \boldsymbol{z})}{1 + r_d(\boldsymbol{h}) F_d(\boldsymbol{h}, \boldsymbol{z})} \right)^{1/2} \tag{19}$$

$$\leqslant \left(\sum_{\boldsymbol{h} \notin \mathcal{A}_d(M)} \frac{1}{r_d(\boldsymbol{h})} + \sum_{\boldsymbol{h} \in \mathcal{A}_d(M)} F_d(\boldsymbol{h}, \boldsymbol{z}) \right)^{1/2}. \tag{20}$$

Both upper bounds above have interpretations as the average case errors for more realistic algorithms. Firstly, if in (15), instead of approximating all Fourier coefficients of f, we approximate only those Fourier coefficients with $\boldsymbol{h} \in \mathcal{A}_d(M)$ and simply omit the terms with $\boldsymbol{h} \notin \mathcal{A}_d(M)$, then we have a modified algorithm

$$A_{n,d,M}^{(1)}(f)(\boldsymbol{x}) := \sum_{\boldsymbol{h} \in \mathcal{A}_d(M)} \left(\frac{1}{n} \frac{1}{1 + r_d(\boldsymbol{h}) F_d(\boldsymbol{h}, \boldsymbol{z})} \sum_{k=1}^{n} f\left(\left\{\frac{k\boldsymbol{z}}{n}\right\}\right) e^{-2\pi \mathrm{i} k \boldsymbol{h} \cdot \boldsymbol{z}/n} \right) e^{2\pi \mathrm{i} \boldsymbol{h} \cdot \boldsymbol{x}}, \tag{21}$$

whose average case error is exactly the right-hand side of (19). Indeed, $A_{n,d,M}^{(1)}$ corresponds to taking

$$\hat{a}_k(\boldsymbol{h}) = 0 \quad \text{and} \quad e_{\boldsymbol{h}} = \frac{1}{r_d(\boldsymbol{h})} \ \text{ for } \boldsymbol{h} \notin \mathcal{A}_d(M)$$

in the proof of Lemma 3. If we further change the optimal algorithm by taking

$$\hat{a}_k(\boldsymbol{h}) = \frac{e^{-2\pi \mathrm{i} k \boldsymbol{h} \cdot \boldsymbol{z}/n}}{n} \quad \text{and} \quad e_{\boldsymbol{h}} = F_d(\boldsymbol{h}, \boldsymbol{z}) \ \text{ for } \boldsymbol{h} \in \mathcal{A}_d(M),$$

which corresponds to using the typical integration weights $1/n$ in the lattice rule approximation of $\hat{f}(\boldsymbol{h})$ for $\boldsymbol{h} \in \mathcal{A}_d(M)$, then we have another modified algorithm

$$A_{n,d,M}^{(2)}(f)(\boldsymbol{x}) := \sum_{\boldsymbol{h} \in \mathcal{A}_d(M)} \left(\frac{1}{n} \sum_{k=1}^{n} f\left(\left\{\frac{k\boldsymbol{z}}{n}\right\}\right) e^{-2\pi \mathrm{i} k \boldsymbol{h} \cdot \boldsymbol{z}/n} \right) e^{2\pi \mathrm{i} \boldsymbol{h} \cdot \boldsymbol{x}}, \tag{22}$$

whose average case error is exactly the expression in (20). We summarize this discussion in the following corollary.

Corollary 4. *The average case errors for the algorithms $A_{n,d,M}^{(1)}$ and $A_{n,d,M}^{(2)}$ defined by* (21) *and* (22)*, respectively, are given by*

$$e_{n,d,M}^{\mathrm{avg}\,(i)}(\boldsymbol{z}) = \left(\sum_{\boldsymbol{h} \notin \mathcal{A}_d(M)} \frac{1}{r_d(\boldsymbol{h})} + E_d^{(i)}(\boldsymbol{z}) \right)^{1/2}, \quad i = 1, 2,$$

where

$$E_d^{(1)}(z) := \sum_{\boldsymbol{h}\in\mathcal{A}_d(M)} \frac{F_d(\boldsymbol{h},z)}{1+r_d(\boldsymbol{h})F_d(\boldsymbol{h},z)} \quad \text{and} \quad E_d^{(2)}(z) := \sum_{\boldsymbol{h}\in\mathcal{A}_d(M)} F_d(\boldsymbol{h},z).$$

For both algorithms, it is clear that only the second term of the error, $E_d^{(i)}(z)$, depends on z, and we should choose z such that $E_d^{(i)}(z)$ is as small as possible. For fixed z we have $E_d^{(1)}(z) \leqslant E_d^{(2)}(z)$ and $e_{n,d}^{\mathrm{avg(opt)}}(z) \leqslant e_{n,d,M}^{\mathrm{avg}(1)}(z) \leqslant e_{n,d,M}^{\mathrm{avg}(2)}(z)$. Note, however, that a good choice of z for one algorithm is not necessarily good for the other two algorithms.

The algorithm proposed in [4] is our $A_{n,d,M}^{(2)}$ here. There it was shown that the *worst case error* of $A_{n,d,M}^{(2)}$, for the unit ball of the weighted Korobov space with the smoothness parameter $\alpha > 1$, is bounded from above by

$$\left(\frac{1}{M} + E_d^{(2)}(z)\right)^{1/2},$$

and a vector z^* was constructed component-by-component to minimize $E_d^{(2)}(z)$. We will describe this construction in the next section.

One important point to note is that the weights in [4] do not depend on the dimension d. Thus, we must be cautious when we make use of results from [4] that are inductive in the dimension. Throughout this paper (except when we discuss tractability and strong tractability), we shall assume that d, and the corresponding weight vector $\boldsymbol{\gamma}_d = (\gamma_{d,1}, \gamma_{d,2}, \ldots, \gamma_{d,d})$, are given and fixed. We shall use s as a running index from 1 to d in any inductive argument. Formally, for each $s = 1, 2, \ldots, d$, we define the s-dimensional truncated variants of $r_d(\boldsymbol{h})$, $\mathcal{A}_d(M)$, $F_d(\boldsymbol{h},z)$, $E_d^{(1)}(z)$ and $E_d^{(2)}(z)$ as follows:

$$r_{d,s}(\boldsymbol{h}) := \prod_{j=1}^{s} r(\alpha,\gamma_{d,j},h_j), \quad \mathcal{A}_{d,s}(M) := \{\boldsymbol{h}\in\mathbb{Z}^s : r_{d,s}(\boldsymbol{h}) \leqslant M\},$$

$$F_{d,s}(\boldsymbol{h},z) := \sum_{\substack{\boldsymbol{\ell}\in\mathbb{Z}^s\setminus\{\boldsymbol{0}\}\\ \boldsymbol{\ell}\cdot z\equiv 0 \,(\mathrm{mod}\, n)}} \frac{1}{r_{d,s}(\boldsymbol{h}+\boldsymbol{\ell})},$$

$$E_{d,s}^{(1)}(z) := \sum_{\boldsymbol{h}\in\mathcal{A}_{d,s}(M)} \frac{F_{d,s}(\boldsymbol{h},z)}{1+r_{d,s}(\boldsymbol{h})F_{d,s}(\boldsymbol{h},z)}, \quad E_{d,s}^{(2)}(z) := \sum_{\boldsymbol{h}\in\mathcal{A}_{d,s}(M)} F_{d,s}(\boldsymbol{h},z).$$

These expressions involve only the first s dimensions, but they make use of the dth weight vector $\boldsymbol{\gamma}_d$. When $s = d$ we recover the original expressions, for example, $\mathcal{A}_{d,d}(M) = \mathcal{A}_d(M)$.

Now we examine the set $\mathcal{A}_d(M)$. Clearly, $\mathcal{A}_d(M) = \emptyset$ if $M < 1$. Furthermore, we observe that $\mathcal{A}_d(M) = \{\boldsymbol{0}\}$ if $\gamma_{d,1} \neq 1$ and $M \in [1, 1/\gamma_{d,1})$, since if $\boldsymbol{h} \in \mathcal{A}_d(M)$ then necessarily $|h_j| \leqslant (\gamma_{d,j}M)^{1/\alpha}$ for all $j = 1, 2, \ldots, d$. Thus, we need $M \geqslant 1/\gamma_{d,1}$ to guarantee a non-trivial set $\mathcal{A}_d(M)$, and if $\gamma_{d,1}$ is tiny then M would have to be huge. Fortunately, the problem is trivial when $\gamma_{d,1}$ is small, and it no longer matters if the set contains just the zero vector. For simplicity, we shall assume for the remainder of this paper that

$$\gamma_{d,1} \geqslant \gamma_0 > 0.$$

We end this section with two lemmas that we shall need later.

Lemma 5. *For $M \geqslant 1$, the cardinality of the set $\mathcal{A}_d(M)$ satisfies*

$$(\gamma_{d,1} M)^{1/\alpha} \leqslant |\mathcal{A}_d(M)| \leqslant M^q \prod_{j=1}^{d} \left(1 + 2\zeta(\alpha q)\gamma_{d,j}^q\right) \quad \forall q > \tfrac{1}{\alpha}.$$

Proof. The proof is by induction using the properties $|\mathcal{A}_{d,1}(M)| = 1 + 2\lfloor(\gamma_{d,1}M)^{1/\alpha}\rfloor \geqslant (\gamma_{d,1}M)^{1/\alpha}$ and

$$\mathcal{A}_{d,s+1}(M) = \{(\boldsymbol{h}, 0) : \boldsymbol{h} \in \mathcal{A}_{d,s}(M)\} \cup \bigcup_{\substack{h_{s+1}=-\infty \\ h_{s+1}\neq 0}}^{\infty} \left\{(\boldsymbol{h}, h_{s+1}) : \boldsymbol{h} \in \mathcal{A}_{d,s}\left(\tfrac{\gamma_{d,s+1}M}{|h_{s+1}|^\alpha}\right)\right\}$$

for all $s = 1, 2, \ldots, d-1$. See [4, Lemma 1] for the full details. □

Lemma 6. *For $M \geqslant 1$ we have*

$$\sum_{\boldsymbol{h}\notin\mathcal{A}_d(M)} \frac{1}{r_d(\boldsymbol{h})} \leqslant \frac{c_{0,d,\tau}}{|\mathcal{A}_d(M)|^{1/\tau-1}} \quad \forall \tau \in \left(\tfrac{1}{\alpha}, 1\right),$$

where

$$c_{0,d,\tau} := \frac{\tau}{1-\tau} \prod_{j=1}^{d} \left(1 + 2\zeta(\alpha\tau)\gamma_{d,j}^\tau\right)^{1/\tau}. \tag{23}$$

Proof. We recall that the vectors $\boldsymbol{h} = \boldsymbol{h}^{(i)} \in \mathbb{Z}^d$ are indexed in such a way that the numbers $\lambda_{d,i} = 1/r_d(\boldsymbol{h}^{(i)})$ are in non-increasing order. Thus,

$$\mathcal{A}_d(M) = \{\boldsymbol{h}^{(i)} \in \mathbb{Z}^d : i = 1, 2, \ldots, |\mathcal{A}_d(M)|\},$$

and we have

$$\sum_{\boldsymbol{h}\notin\mathcal{A}_d(M)} \frac{1}{r_d(\boldsymbol{h})} = \sum_{i=|\mathcal{A}_d(M)|+1}^{\infty} \lambda_{d,i}.$$

Since the numbers $\lambda_{d,i}$ are non-increasing, we have for all $i \geqslant 1$,

$$\lambda_{d,i}^\tau \leqslant \frac{1}{i}\sum_{\ell=1}^{i} \lambda_{d,\ell}^\tau \leqslant \frac{1}{i}\sum_{\ell=1}^{\infty} \lambda_{d,\ell}^\tau = \frac{1}{i}\prod_{j=1}^{d}\left(1 + 2\zeta(\alpha\tau)\gamma_{d,j}^\tau\right) \quad \forall \tau > \tfrac{1}{\alpha},$$

which leads to

$$\sum_{i=|\mathcal{A}_d(M)|+1}^{\infty} \lambda_{d,i} \leqslant \sum_{i=|\mathcal{A}_d(M)|+1}^{\infty} i^{-1/\tau} \prod_{j=1}^{d}\left(1 + 2\zeta(\alpha\tau)\gamma_{d,j}^\tau\right)^{1/\tau}$$

$$\leqslant \int_{|\mathcal{A}_d(M)|}^{\infty} x^{-1/\tau}\,\mathrm{d}x \prod_{j=1}^{d}\left(1+2\zeta(\alpha\tau)\gamma_{d,j}^{\tau}\right)^{1/\tau}$$

$$= \frac{1}{|\mathcal{A}_d(M)|^{1/\tau-1}}\frac{\tau}{1-\tau}\prod_{j=1}^{d}\left(1+2\zeta(\alpha\tau)\gamma_{d,j}^{\tau}\right)^{1/\tau} \quad \forall \tau \in \left(\tfrac{1}{\alpha},1\right).$$

This completes the proof. □

5. Generating vectors constructed for approximation

In this section we introduce component-by-component constructions to find good generating vectors for lattice rule algorithms. We prove that the average case errors of these algorithms achieve tractability or strong tractability error bounds.

5.1. Component-by-component constructions

We see from Corollary 4 that it is enough to construct a generating vector $\boldsymbol{z}$ for which $E_d^{(i)}(\boldsymbol{z})$ is as small as possible. We stress once again that we must use the same weight vector $\boldsymbol{\gamma}_d = (\gamma_{d,1}, \gamma_{d,2}, \ldots, \gamma_{d,d})$ throughout the entire construction process. The two constructions, one for each algorithm $A_{n,d,M}^{(i)}$, are given below.

Algorithm 7. *Let n be a prime number and $M \geqslant 1$. For $i = 1$ or 2,*

1. *Set $z_1 = 1$.*
2. *For $s = 2, 3, \ldots, d$, find z_s in $\{1, 2, \ldots, n-1\}$ to minimize*

$$E_{d,s}^{(i)}(z_1, \ldots, z_{s-1}, z_s).$$

Lemma 8. *Let $\kappa > 1$ and $M \geqslant 1$. Suppose n is a prime number satisfying $n \geqslant \kappa M^{1/\alpha}$. Then for each $i = 1, 2$ the average case error of the algorithm $A_{n,d,M}^{(i)}$ with $\boldsymbol{z}^{(i)} \in \mathcal{Z}_n^d$ constructed by Algorithm 7(i) satisfies*

$$e_{n,d,M}^{\mathrm{avg}\,(i)}(\boldsymbol{z}^{(i)}) \leqslant \left(\frac{c_{0,d,\tau}}{|\mathcal{A}_d(M)|^{1/\tau-1}} + \frac{c_{1,d,\lambda,\delta}|\mathcal{A}_d(M)|^{1/\lambda}}{(n-1)^{1/\lambda}}\right)^{1/2}$$

for all $\tau \in (1/\alpha, 1)$, $\lambda \in (1/\alpha, 1]$, and $\delta \in (0, (1-1/\kappa)^{\alpha}]$, where $c_{0,d,\tau}$ is defined as in (23)*, and*

$$c_{1,d,\lambda,\delta} := \frac{1}{\delta}\prod_{j=1}^{d}\left(1+2(1+\delta^{\lambda})\zeta(\alpha\lambda)\gamma_{d,j}^{\lambda}\right)^{1/\lambda}.$$

Proof. Using Corollary 4 and Lemma 6, we see that the result is proved if we can show by induction that for each $i = 1, 2$, the vector $\boldsymbol{z}^{(i)} \in \mathcal{Z}_n^d$ constructed by Algorithm 7(i) satisfies, for each $s = 1, 2, \ldots, d$,

$$E_{d,s}^{(i)}(z_1^{(i)}, \ldots, z_s^{(i)}) \leqslant \left(\frac{1}{\delta^{\lambda}}\frac{|\mathcal{A}_{d,s}(M)|}{n-1}\prod_{j=1}^{s}\left(1+2(1+\delta^{\lambda})\zeta(\alpha\lambda)\gamma_{d,j}^{\lambda}\right)\right)^{1/\lambda} \tag{24}$$

for all $\lambda \in (1/\alpha, 1]$ and $\delta \in (0, (1-1/\kappa)^\alpha]$. Here we only give an outline of the proof. See the proof of [4, Lemma 6] for the full details. Note that our error bound here is tighter than in [4, Lemma 6] since it depends on the cardinality of $\mathcal{A}_{d,s}(M)$ instead of its upper bound.

For each $i = 1, 2$, suppose we have already obtained $\boldsymbol{z}^{(i,s)} = (z_1^{(i)}, \ldots, z_s^{(i)}) \in \mathcal{Z}_n^s$ from Algorithm 7(i), with $s < d$, for which (24) is satisfied. First we show that

$$E_{d,s+1}^{(i)}(\boldsymbol{z}^{(i,s)}, z_{s+1}) \leqslant \left(1 + 2\zeta(\alpha)\gamma_{d,s+1}\right) E_{d,s}^{(i)}(\boldsymbol{z}^{(i,s)}) + \theta(\boldsymbol{z}^{(i,s)}, z_{s+1}),$$

where $\theta(\boldsymbol{z}^{(i,s)}, z_{s+1})$ corresponds to the $\ell_{s+1} \neq 0$ terms in $E_{d,s+1}^{(2)}(\boldsymbol{z}^{(i,s)}, z_{s+1})$. Then we apply Jensen's inequality and show that for all $\lambda \in (1/\alpha, 1]$,

$$\frac{1}{n-1} \sum_{z_{s+1}=1}^{n-1} \left[\theta(\boldsymbol{z}^{(i,s)}, z_{s+1})\right]^\lambda \leqslant W_1 + W_2 + W_3,$$

with

$$W_1 \leqslant \frac{2\zeta(\alpha\lambda)\gamma_{d,s+1}^\lambda}{n-1} \left(\prod_{j=1}^{s} \left(1 + 2\zeta(\alpha\lambda)\gamma_{d,j}^\lambda\right)\right) |\mathcal{A}_{d,s}(M)|,$$

$$W_2 \leqslant \frac{1 + 2\zeta(\alpha\lambda)\gamma_{d,s+1}^\lambda}{n-1} \left(\prod_{j=1}^{s} \left(1 + 2\zeta(\alpha\lambda)\gamma_{d,j}^\lambda\right)\right) 2 \sum_{h_{s+1}=1}^{\infty} \left|\mathcal{A}_{d,s}\left(\tfrac{\gamma_{d,s+1}M}{h_{s+1}^\alpha}\right)\right|,$$

$$W_3 \leqslant \frac{2(1-1/\kappa)^{-\alpha\lambda}\zeta(\alpha\lambda)\gamma_{d,s+1}^\lambda}{n-1} \left(\prod_{j=1}^{s} \left(1 + 2\zeta(\alpha\lambda)\gamma_{d,j}^\lambda\right)\right) 2 \sum_{h_{s+1}=1}^{\infty} \left|\mathcal{A}_{d,s}\left(\tfrac{\gamma_{d,s+1}M}{h_{s+1}^\alpha}\right)\right|.$$

These estimates are tedious to obtain. In particular, we need to use the property that $\boldsymbol{h} \in \mathcal{A}_{d,s+1}(M)$ implies $|h_{s+1}| \leqslant (\gamma_{d,s+1}M)^{1/\alpha} \leqslant M^{1/\alpha} \leqslant \kappa^{-1}n$.

Since $z_{s+1}^{(i)}$ is chosen to minimize $E_{d,s+1}^{(i)}(\boldsymbol{z}^{(i,s)}, z_{s+1})$, we have

$$\begin{aligned} &E_{d,s+1}^{(i)}(\boldsymbol{z}^{(i,s)}, z_{s+1}^{(i)}) \\ &\quad\leqslant \left(1 + 2\zeta(\alpha)\gamma_{d,s+1}\right) E_{d,s}^{(i)}(\boldsymbol{z}^{(i,s)}) + \left(\min_{z_{s+1} \in \mathcal{Z}_n} [\theta(\boldsymbol{z}^{(i,s)}, z_{s+1})]^\lambda\right)^{1/\lambda} \\ &\quad\leqslant \left(1 + 2\zeta(\alpha)\gamma_{d,s+1}\right) E_{d,s}^{(i)}(\boldsymbol{z}^{(i,s)}) + \left(\frac{1}{n-1} \sum_{z_{s+1}=1}^{n-1} [\theta(\boldsymbol{z}^{(i,s)}, z_{s+1})]^\lambda\right)^{1/\lambda}. \end{aligned}$$

Now we combine all of the estimates above and use the induction hypothesis and the property

$$|\mathcal{A}_{d,s+1}(M)| = |\mathcal{A}_{d,s}(M)| + 2 \sum_{h_{s+1}=1}^{\infty} \left|\mathcal{A}_{d,s}\left(\tfrac{\gamma_{d,s+1}M}{h_{s+1}^\alpha}\right)\right|.$$

After multiple applications of Jensen's inequality, we finally conclude that the error bound (24) holds with s replaced by $s+1$. □

5.2. Choosing M and n

Given $\varepsilon > 0$, we want to find small M and n for which the upper bound of the errors $e^{\text{avg}\,(i)}_{n,d,M}(\boldsymbol{z}^{(i)})$, $i = 1, 2$, given in Lemma 8 is at most $\varepsilon e^{\text{avg}}_{0,d}$. First we estimate the error bound from above by replacing $|\mathcal{A}_d(M)|$ with its lower and upper bounds from Lemma 5. This gives

$$e^{\text{avg}\,(i)}_{n,d,M}(\boldsymbol{z}^{(i)}) \leqslant \left(\frac{c_{2,d,\tau}}{M^{(1-\tau)/(\alpha\tau)}} + \frac{c_{3,d,q,\lambda,\delta}\, M^{q/\lambda}}{(n-1)^{1/\lambda}}\right)^{1/2}$$

for all $q > 1/\alpha$, $\tau \in (1/\alpha, 1)$, $\lambda \in (1/\alpha, 1]$, and $\delta \in (0, (1 - 1/\kappa)^{\alpha}]$, where

$$c_{2,d,\tau} := \left(\frac{1}{\gamma_{d,1}}\right)^{\frac{1-\tau}{\alpha\tau}} \frac{\tau}{1-\tau} \prod_{j=1}^{d} \left(1 + 2\zeta(\alpha q)\gamma^{\tau}_{d,j}\right)^{1/\tau},$$

$$c_{3,d,q,\lambda,\delta} := \frac{1}{\delta} \prod_{j=1}^{d} \left(1 + 2\zeta(\alpha q)\gamma^{q}_{d,j}\right)^{1/\lambda} \left(1 + 2(1+\delta^{\lambda})\zeta(\alpha\lambda)\gamma^{\lambda}_{d,j}\right)^{1/\lambda}. \tag{25}$$

Now we take $q = \lambda = \tau$ and choose $M \geqslant 1$ such that the first term in the error bound is at most $\varepsilon^2[e^{\text{avg}}_{0,d}]^2/2$. This can be achieved by taking

$$M = \max\left(\left(\frac{2c_{2,d,\tau}}{\varepsilon^2[e^{\text{avg}}_{0,d}]^2}\right)^{\frac{\alpha\tau}{1-\tau}}, 1\right). \tag{26}$$

We then choose a prime n satisfying $n \geqslant \kappa M^{1/\alpha}$ such that the second term in the error bound is no greater than the first. This can be achieved by taking

$$n = \text{pr}\left(\max\left(\left(\frac{c_{3,d,\tau,\tau,\delta} M^{\frac{1-\tau+\alpha\tau}{\alpha\tau}}}{c_{2,d,\tau}}\right)^{\tau} + 1, \kappa M^{\frac{1}{\alpha}}\right)\right), \tag{27}$$

where $\text{pr}(x)$ denotes the smallest prime number which is no less than x. From Chebyshev's theorem we know that $\lceil x \rceil \leqslant \text{pr}(x) \leqslant 2\lceil x \rceil$.

Substituting (26) into (27), we obtain

$$n = \text{pr}\left(\max\left(\mathcal{C}_{1,d,\tau,\delta}\, \varepsilon^{-\frac{2\tau}{1-\tau}[1+\tau(\alpha-1)]} + 1, \mathcal{C}_{2,d,\tau}\varepsilon^{-\frac{2\tau}{1-\tau}}, \mathcal{C}_{3,d,\tau,\delta}, \kappa\right)\right),$$

where

$$\begin{aligned}
\mathcal{C}_{1,d,\tau,\delta} &:= \left(\frac{c_{3,d,\tau,\tau,\delta}}{c_{2,d,\tau}} \left(\frac{2\, c_{2,d,\tau}}{[e^{\text{avg}}_{0,d}]^2}\right)^{\frac{1-\tau+\alpha\tau}{1-\tau}}\right)^{\tau} \\
&\leqslant 2^{\tau+\frac{\alpha\tau^2}{1-\tau}} \left(\frac{\tau}{1-\tau}\right)^{\frac{\alpha\tau^2}{1-\tau}} \left(\frac{1}{\delta}\right)^{\tau} \left(\frac{1}{\gamma_0}\right)^{\tau} \exp\left(2\zeta(\alpha\tau)\left(2 + \frac{\alpha\tau}{1-\tau} + \delta^{\tau}\right) \sum_{j=1}^{d} \gamma^{\tau}_{d,j}\right) \qquad (28) \\
&= 2^{\tau+\frac{\alpha\tau^2}{1-\tau}} \left(\frac{\tau}{1-\tau}\right)^{\frac{\alpha\tau^2}{1-\tau}} \left(\frac{1}{\delta}\right)^{\tau} \left(\frac{1}{\gamma_0}\right)^{\tau} (d+1)^{2\zeta(\alpha\tau)\left(2+\frac{\alpha\tau}{1-\tau}+\delta^{\tau}\right)\frac{\sum_{j=1}^{d}\gamma^{\tau}_{d,j}}{\log(d+1)}}, \qquad (29)
\end{aligned}$$

$$\mathcal{C}_{2,d,\tau} := \kappa \left(\frac{2\, c_{2,d,\tau}}{[e_{0,d}^{\mathrm{avg}}]^2} \right)^{\frac{\tau}{1-\tau}}$$

$$\leqslant \kappa 2^{\frac{\tau}{1-\tau}} \left(\frac{\tau}{1-\tau}\right)^{\frac{\tau}{1-\tau}} \left(\frac{1}{\gamma_0}\right)^{\frac{1}{\alpha}} \exp\left(2\zeta(\alpha\tau) \tfrac{1}{1-\tau} \sum_{j=1}^{d} \gamma_{d,j}^{\tau} \right) \tag{30}$$

$$= \kappa 2^{\frac{\tau}{1-\tau}} \left(\frac{\tau}{1-\tau}\right)^{\frac{\tau}{1-\tau}} \left(\frac{1}{\gamma_0}\right)^{\frac{1}{\alpha}} (d+1)^{2\zeta(\alpha\tau)\frac{1}{1-\tau}\frac{\sum_{j=1}^{d}\gamma_{d,j}^{\tau}}{\log(d+1)}}, \tag{31}$$

and

$$\mathcal{C}_{3,d,\tau,\delta} := \left(\frac{c_{3,d,\tau,\tau,\delta}}{c_{2,d,\tau}} \right)^{\tau} + 1$$

$$\leqslant \left(\tfrac{1-\tau}{\tau}\right)^{\tau} \left(\tfrac{1}{\delta}\right)^{\tau} \exp\left(2\zeta(\alpha\tau)\left(1+\delta^{\tau}\right) \sum_{j=1}^{d} \gamma_{d,j}^{\tau} \right) + 1 \tag{32}$$

$$= \left(\tfrac{1-\tau}{\tau}\right)^{\tau} \left(\tfrac{1}{\delta}\right)^{\tau} (d+1)^{2\zeta(\alpha\tau)(1+\delta^{\tau})\frac{\sum_{j=1}^{d}\gamma_{d,j}^{\tau}}{\log(d+1)}} + 1. \tag{33}$$

Suppose first that $s_\gamma < 1$ which, as we know, is needed for strong tractability. We take $\tau \in (\max(1/\alpha, s_\gamma), 1)$ and $\delta \in (0, (1-1/\kappa)^{\alpha}]$. Since $\alpha\tau > 1$ and $\tau > s_\gamma$, we have $\zeta(\alpha\tau) < \infty$ and $\sup_{d\geqslant 1} \sum_{j=1}^{d} \gamma_{d,j}^{\tau} < \infty$. Thus we see from (28), (30), and (32) that

$$\sup_{d\geqslant 1} \mathcal{C}_{1,d,\tau,\delta} < \infty, \quad \sup_{d\geqslant 1} \mathcal{C}_{2,d,\tau} < \infty \quad \text{and} \quad \sup_{d\geqslant 1} \mathcal{C}_{3,d,\tau,\delta} < \infty.$$

Assume next that $t_\gamma < 1$ which is needed for tractability. We choose $\tau \in (\max(1/\alpha, t_\gamma), 1)$ and $\delta \in (0, (1-1/\kappa)^{\alpha}]$. Then $\zeta(\alpha\tau) < \infty$ and $R_\tau < \infty$. Hence it follows from (29), (31), and (33) that

$$\mathcal{C}_{1,d,\tau,\delta} = O(d^a), \quad \mathcal{C}_{2,d,\tau} = O(d^b) \quad \text{and} \quad \mathcal{C}_{3,d,\tau,\delta} = O(d^c),$$

where $a > b$, $a > c$, and a is arbitrarily close to $2\zeta(\alpha\tau)[2 + \alpha\tau/(1-\tau) + \delta^{\tau}]R_\tau$.

We summarize the analysis of this section in the following theorem.

Theorem 9. *Consider multivariate approximation in the average case setting defined as in Theorem* 1.

(a) *Suppose that* $s_\gamma < 1$. *Given* $d \geqslant 1$, $\varepsilon \in (0,1)$, $\kappa > 1$, $\tau \in (\max(1/\alpha, s_\gamma), 1)$, *and* $\delta \in (0, (1-1/\kappa)^{\alpha}]$, *for each* $i = 1, 2$, *the approximation algorithm* $A_{n,d,M}^{(i)}$, *with* M *and* n *given by* (26) *and* (27), *and with generating vector* $\boldsymbol{z}^{(i)}$ *constructed by Algorithm* 7(i), *achieves the error bound* $e_{n,d,M}^{\mathrm{avg}\,(i)}(\boldsymbol{z}^{(i)}) \leqslant \varepsilon e_{0,d}^{\mathrm{avg}}$ *using*

$$n = O\left(\varepsilon^{-\frac{2\tau}{1-\tau}[1+\tau(\alpha-1)]} \right)$$

function values. The implied factor in the big O notation is independent of ε *and* d *but depends on* τ *and* δ. *The exponent of* ε^{-1} *can be arbitrarily close to*

$$p^{\mathrm{avg}}(\Lambda^{\mathrm{std}})\,[1 + \max(1/\alpha, s_\gamma)(\alpha - 1)]\,.$$

(b) *Suppose that* $t_\gamma < 1$. *Given* $d \geqslant 1$, $\varepsilon \in (0,1)$, $\kappa > 1$, $\tau \in (\max(1/\alpha, t_\gamma), 1)$, *and* $\delta \in (0, (1-1/\kappa)^\alpha]$, *for each* $i = 1, 2$, *the approximation algorithm* $A^{(i)}_{n,d,M}$, *with* M *and* n *given by* (26) *and* (27), *and with generating vector* $\boldsymbol{z}^{(i)}$ *constructed by Algorithm* 7(i), *achieves the error bound* $e^{\mathrm{avg}(i)}_{n,d,M}(\boldsymbol{z}^{(i)}) \leqslant \varepsilon e^{\mathrm{avg}}_{0,d}$ *using*

$$n = O\left(d^a \varepsilon^{-\frac{2\tau}{1-\tau}[1+\tau(\alpha-1)]}\right)$$

function values, *where* a *is arbitrarily close to*

$$2\zeta(\alpha\tau)\left(2 + \tfrac{\alpha\tau}{1-\tau} + \delta^\tau\right) R_\tau.$$

The implied factor in the big O notation is independent of ε *and* d *but depends on* τ, δ, *and* a.

In this section we have sought the optimal relationship between M and n in order to explore the issues relating to tractability and the corresponding exponents. We mention, however, that in practice there is no need to choose M and n in this way, since the average case error expression in Corollary 4 as well as its upper bound in Lemma 8 are both valid regardless of how M and n are chosen. It is natural to think of M as determining the linear space

$$\{e^{2\pi i \boldsymbol{h}\cdot\boldsymbol{x}} : \boldsymbol{h} \in \mathcal{A}_d(M)\} \tag{34}$$

from which the algorithms $A^{(i)}_{n,d,M}(f)$ in (21) and (22) are to be chosen, while n is the number of points in the quadrature scheme used to approximate the corresponding Fourier coefficients. As a final remark about the choice of M and n, we may note that a different approximating linear space to (34) was considered in [5], with the requirement that the dimension of the approximating space is equal to n. There is no such requirement here.

5.3. Improved error bounds

We recall from Theorem 1 that the exponents of ε^{-1} for strong tractability and tractability depend on

$$\max(1/\alpha, s_\gamma) \quad \text{and} \quad \max(1/\alpha, t_\gamma),$$

respectively. Thus if s_γ or t_γ is greater than $1/\alpha$, then the smoothness parameter α has no effect on the exponent of ε^{-1} whatsoever. Hence, even if we have infinite smoothness, the problem can be extremely hard if the weights do not decay fast enough.

Assume for the moment that $s_\gamma < 1$, which is needed for strong tractability. Theorem 9 presents us with an undesirable dilemma: in the case of $s_\gamma > 1/\alpha$, the exponent of ε^{-1} is arbitrarily close to

$$p^{\mathrm{avg}}(\Lambda^{\mathrm{std}})\,[1 + s_\gamma(\alpha - 1)]\,,$$

which increases as the smoothness parameter α increases, and it can grow to infinity if α is allowed to approach infinity. In other words, higher smoothness actually hurts!

This counter intuitive observation can be partially explained from the bounds on the cardinality of the set $\mathcal{A}_d(M)$. We recall from Lemma 5 that

$$(\gamma_{d,1} M)^{1/\alpha} \leqslant |\mathcal{A}_d(M)| \leqslant M^q \prod_{j=1}^{d}\left(1 + 2\zeta(\alpha q)\gamma^q_{d,j}\right) \quad \forall q > \tfrac{1}{\alpha}.$$

To ensure that the upper bound on $|\mathcal{A}_d(M)|$ is uniformly bounded in d, we need to impose an additional condition that $q > s_\gamma$. If $s_\gamma \leqslant 1/\alpha$, then we can take q arbitrarily close to $1/\alpha$ and the bounds on $|\mathcal{A}_d(M)|$ have essentially the same order $M^{1/\alpha}$. On the other hand, when $s_\gamma > 1/\alpha$, the value of q can be no smaller than s_γ and there is a gap between the upper and lower bounds with respect to the exponent of M. This gap can be huge for large α.

Put differently, if $1/\alpha < q < s_\gamma < 1$, then the upper bound on $|\mathcal{A}_d(M)|/M^q$ depends on d and may go to infinity with d faster than any polynomial in d. To illustrate this point, assume for the moment that $\gamma_{d,j} = d^{-1/\beta}$ for some $\beta \in (1/\alpha, 1)$. Then $s_\gamma = \beta$, and for $q \in (1/\alpha, s_\gamma)$ we have

$$\prod_{j=1}^{d} \left(1 + 2\zeta(\alpha q)\gamma_{d,j}^q\right) = \exp\left(2\zeta(\alpha q) d^{1-q/\beta}\right)(1 + o(1)) \quad \text{as } d \to \infty,$$

which goes faster with d to infinity than any polynomial in d. To see that $|\mathcal{A}_d(M)|/M^q$ indeed grows with d and not just its upper bound, take $M = 2^{k\alpha} d^{k/\beta}$ for some integer $k < d$ and keep $\gamma_{d,j} = d^{-1/\beta}$. Then all vectors $\boldsymbol{h}$ with at least $d - k$ components equal to 0, and at most k components from the set $\{-2, -1, 1, 2\}$ belong to the set $\mathcal{A}_d(M)$. Since we have $4^k \binom{d}{k}$ such vectors, the cardinality of $\mathcal{A}_d(M)$ is at least of order d^k. Hence,

$$\frac{|\mathcal{A}_d(M)|}{M^q} = \Omega\left(d^{k(1-q/\beta)}\right) \quad \forall q \in \left(\tfrac{1}{\alpha}, s_\gamma\right).$$

We stress that k can be arbitrarily large. This means that for $q < s_\gamma = \beta$, $|\mathcal{A}_d(M)|/M^q$ may depend on an arbitrarily large power of d. This proves that for slowly decaying weights, the cardinality of $\mathcal{A}_d(M)$ does not depend on α alone but on the sum exponent of weights.

To remove this undesirable dilemma, we will use an artificial smoothness parameter $\tilde{\alpha} \leqslant \alpha$ defined by

$$\tilde{\alpha} := \min(\alpha, 1/s_\gamma),$$

when $s_\gamma < 1$, and

$$\tilde{\alpha} := \min(\alpha, 1/t_\gamma),$$

when $t_\gamma < 1$. We define a new set

$$\tilde{\mathcal{A}}_d(M) := \{\boldsymbol{h} \in \mathbb{Z}^d : r_d(\tilde{\alpha}, \boldsymbol{\gamma}_d, \boldsymbol{h}) \leqslant M\},$$

which is essentially the set $\mathcal{A}_d(M)$ defined in (18), but with the α in $r_d(\boldsymbol{h}) = r_d(\alpha, \boldsymbol{\gamma}_d, \boldsymbol{h})$ replaced by $\tilde{\alpha}$. Then we define two new algorithms $A^{(i)}_{n,d,M,\tilde{\alpha}}$, $i = 1, 2$, by replacing the set $\mathcal{A}_d(M)$ in (21) and (22) with the new set $\tilde{\mathcal{A}}_d(M)$. Note that all other occurrences of α in $r_d(\boldsymbol{h})$ and $F_d(\boldsymbol{h}, \boldsymbol{z})$ in the formula (21) remain unchanged.

The average case errors $e^{\mathrm{avg}(i)}_{n,d,M,\tilde{\alpha}}(\boldsymbol{z})$, $i = 1, 2$, for the new algorithms can be obtained by renaming $E^{(i)}_d(\boldsymbol{z})$ to $\tilde{E}^{(i)}_d(\boldsymbol{z})$ and simply replacing all occurrences of $\mathcal{A}_d(M)$ with $\tilde{\mathcal{A}}_d(M)$ in Corollary 4, while keeping all other occurrences of α unchanged. The component-by-component constructions given by Algorithm 7(i), $i = 1, 2$, can be modified accordingly. When it comes to analyzing the average case errors achieved by the new algorithms, we can make use of all our existing analysis. Since the value of $r_d(\alpha, \boldsymbol{\gamma}_d, \boldsymbol{h})$ increases as α decreases, all the error expressions can be bounded from above by replacing *all* occurrences of α with $\tilde{\alpha}$. (Note that we are not changing

the smoothness parameter; all we are doing is changing the upper bounds on the average case errors.) Hence we obtain the following Theorem analogous to Theorem 9.

Theorem 10. *Consider multivariate approximation in the average case setting defined as in Theorem* 1.

(a) *Suppose that* $s_\gamma < 1$ *and define*

$$\tilde{\alpha} := \min(\alpha, 1/s_\gamma).$$

Given $d \geqslant 1$, $\varepsilon \in (0,1)$, $\kappa > 1$, $\tau \in (1/\tilde{\alpha}, 1)$, *and* $\delta \in (0, (1-1/\kappa)^{\tilde{\alpha}}]$, *for each* $i = 1, 2$, *the approximation algorithm* $A^{(i)}_{n,d,M,\tilde{\alpha}}$, *with* M *and* n *given by* (26) *and* (27) *where every occurrence of* α *is replaced by* $\tilde{\alpha}$, *and with generating vector* $\boldsymbol{z}^{(i)}$ *constructed by the modified Algorithm* 7(i), *achieves the error bound* $e^{\mathrm{avg}(i)}_{n,d,M,\tilde{\alpha}}(\boldsymbol{z}^{(i)}) \leqslant \varepsilon e^{\mathrm{avg}}_{0,d}$ *using*

$$n = O\left(\varepsilon^{-\frac{2\tau}{1-\tau}[1+\tau(\tilde{\alpha}-1)]}\right)$$

function values. The implied factor in the big O notation is independent of ε *and* d *but depends on* τ *and* δ. *The exponent of* ε^{-1} *can be arbitrarily close to*

$$\frac{2}{\tilde{\alpha}-1} + \frac{2}{\tilde{\alpha}} = p^{\mathrm{avg}}(\Lambda^{\mathrm{std}}) + \frac{2p^{\mathrm{avg}}(\Lambda^{\mathrm{std}})}{2 + p^{\mathrm{avg}}(\Lambda^{\mathrm{std}})}.$$

(b) *Suppose that* $t_\gamma < 1$ *and define*

$$\tilde{\alpha} := \min(\alpha, 1/t_\gamma).$$

Given $d \geqslant 1$, $\varepsilon \in (0,1)$, $\kappa > 1$, $\tau \in (1/\tilde{\alpha}, 1)$, *and* $\delta \in (0, (1-1/\kappa)^{\tilde{\alpha}}]$, *for each* $i = 1, 2$, *the approximation algorithm* $A^{(i)}_{n,d,M,\tilde{\alpha}}$, *with* M *and* n *given by* (26) *and* (27) *where every occurrence of* α *is replaced by* $\tilde{\alpha}$, *and with generating vector* $\boldsymbol{z}^{(i)}$ *constructed by the modified Algorithm* 7(i), *achieves the error bound* $e^{\mathrm{avg}(i)}_{n,d,M,\tilde{\alpha}}(\boldsymbol{z}^{(i)}) \leqslant \varepsilon e^{\mathrm{avg}}_{0,d}$ *using*

$$n = O\left(d^a \varepsilon^{-\frac{2\tau}{1-\tau}[1+\tau(\tilde{\alpha}-1)]}\right)$$

function values, *where* a *is arbitrarily close to*

$$2\zeta(\tilde{\alpha}\tau)\left(2 + \tfrac{\tilde{\alpha}\tau}{1-\tau} + \delta^\tau\right) R_\tau.$$

The implied factor in the big O notation is independent of ε *and* d *but depends on* τ, δ, *and* a.

Note that

$$p^{\mathrm{avg}}(\Lambda^{\mathrm{std}}) + \frac{2p^{\mathrm{avg}}(\Lambda^{\mathrm{std}})}{2 + p^{\mathrm{avg}}(\Lambda^{\mathrm{std}})} \leqslant p^{\mathrm{avg}}(\Lambda^{\mathrm{std}})\,[1 + \max(1/\alpha, s_\gamma)(\alpha - 1)]\,.$$

Thus, Theorem 10 is indeed an improvement over Theorem 9.

6. Generating vectors constructed for integration

For the moment let us forget about the artificial smoothness parameter $\tilde{\alpha}$. The lattice rule algorithms $A^{(1)}_{n,d,M}$ and $A^{(2)}_{n,d,M}$ approximate the Fourier coefficients of f, which are integrals, by lattice rules. Thus it makes sense to ask if a generating vector $\boldsymbol{z}$ constructed for the integration problem can also be efficiently used for the approximation problem.

For n prime and $\boldsymbol{z} \in \mathcal{Z}_n^d$, we approximate the integral of f

$$I_d(f) = \int_{[0,1]^d} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$$

by the optimal integration rule $Q^{(\mathrm{opt})}_{n,d}$ using lattice points $\{k\boldsymbol{z}/n\}$ for $k = 1, 2, \ldots, n$. That is,

$$Q^{(\mathrm{opt})}_{n,d}(f) = \sum_{k=1}^{n} w_k f\left(\left\{\tfrac{k\boldsymbol{z}}{n}\right\}\right),$$

where $w_1, w_2, \ldots, w_n$ are integration weights chosen to minimize the average case integration error:

$$e^{\text{avg-INT (opt)}}_{n,d}(\boldsymbol{z}) := \left(\int_{\mathcal{F}} |I_d(f) - Q^{(\mathrm{opt})}_{n,d}(f)|^2 \mu_d(\mathrm{d}f)\right)^{1/2}.$$

Since $I_d(f) = \hat{f}(\boldsymbol{0})$, we see from Lemma 3 and its proof that the optimal integration weights are given by

$$w_k = \hat{a}_k(\boldsymbol{0}) = \frac{1}{n}\frac{1}{1 + F_d(\boldsymbol{0}, \boldsymbol{z})}, \quad k = 1, 2, \ldots, n,$$

and the average case error of $Q^{(\mathrm{opt})}_{n,d}$ is

$$e^{\text{avg-INT (opt)}}_{n,d}(\boldsymbol{z}) = e^{1/2}_{\boldsymbol{0}} = \left(\frac{F_d(\boldsymbol{0}, \boldsymbol{z})}{1 + F_d(\boldsymbol{0}, \boldsymbol{z})}\right)^{1/2} = \left(1 - \frac{1}{1 + F_d(\boldsymbol{0}, \boldsymbol{z})}\right)^{1/2}.$$

If, instead of using the optimal integration weights given above, we use the typical equal weights $1/n$, then the average case integration error is $[F_d(\boldsymbol{0}, \boldsymbol{z})]^{1/2}$. In both cases, it is sufficient that we find a vector $\boldsymbol{z}$ which leads to a small value of $F_d(\boldsymbol{0}, \boldsymbol{z})$. We construct $\boldsymbol{z}$ component-by-component as follows, keeping in mind that we must use the same weight vector $\boldsymbol{\gamma}_d = (\gamma_{d,1}, \gamma_{d,2}, \ldots, \gamma_{d,d})$ throughout the entire construction process.

Algorithm 11. *Let n be a prime number.*

1. *Set* $z_1 = 1$.
2. *For* $s = 2, 3, \ldots, d$, *with* $z_1, \ldots, z_{s-1}$ *fixed, find* z_s in $\{1, 2, \ldots, n-1\}$ *to minimize*

$$F_{d,s}(\boldsymbol{0}, (z_1, \ldots, z_{s-1}, z_s)).$$

Lemma 12. *Let n be prime. The average case errors for the algorithms $A^{(1)}_{n,d,M}$ and $A^{(2)}_{n,d,M}$ with $\boldsymbol{z}^{(0)} \in \mathcal{Z}_n^d$ constructed by Algorithm* 11 *satisfy*

$$e^{\text{avg (1)}}_{n,d,M}(\boldsymbol{z}^{(0)}) \leqslant e^{\text{avg (2)}}_{n,d,M}(\boldsymbol{z}^{(0)}) \leqslant \left(\frac{c_{0,d,\tau}}{|\mathcal{A}_d(M)|^{1/\tau - 1}} + \frac{c_{4,d,\lambda}\, M\, |\mathcal{A}_d(M)|}{(n-1)^{1/\lambda}}\right)^{1/2}$$

for all $\tau \in (1/\alpha, 1)$ *and* $\lambda \in (1/\alpha, 1]$, *where* $c_{0,d,\tau}$ *is defined as in* (23) *and*

$$c_{4,d,\lambda} := 2^{\alpha \hat{k}_{\gamma_d}} \left(\prod_{j=1}^{\hat{k}_{\gamma_d}} \gamma_{d,j} \right) \prod_{j=1}^{d} \left(1 + 2\zeta(\alpha\lambda)\gamma_{d,j}^{\lambda} \right)^{1/\lambda},$$

with $\hat{k}_{\gamma_d}$ *denoting the largest index* $1 \leqslant k \leqslant d$ *such that* $\gamma_{d,k} > 2^{-\alpha}$, *and* $\hat{k}_{\gamma_d} = 0$ *if no such index exists*.

Proof. Using the following estimate proved in [6], also used in [4, Section 2],

$$\frac{1}{r_d(\boldsymbol{h}+\boldsymbol{\ell})} \leqslant \frac{r_d(\boldsymbol{h})}{r_d(\boldsymbol{\ell})} \prod_{j=1}^{d} \max(1, 2^{\alpha}\gamma_{d,j}),$$

we have from (16) that $F_d(\boldsymbol{h}, \boldsymbol{z}) \leqslant r_d(\boldsymbol{h}) F_d(\boldsymbol{0}, \boldsymbol{z}) \prod_{j=1}^{d} \max(1, 2^{\alpha}\gamma_{d,j})$, which leads to

$$\begin{aligned} E_d^{(1)}(\boldsymbol{z}) \leqslant E_d^{(2)}(\boldsymbol{z}) &\leqslant \sum_{\boldsymbol{h} \in \mathcal{A}_d(M)} r_d(\boldsymbol{h}) F_d(\boldsymbol{0}, \boldsymbol{z}) \prod_{j=1}^{d} \max(1, 2^{\alpha}\gamma_{d,j}) \\ &\leqslant M |\mathcal{A}_d(M)|\, F_d(\boldsymbol{0}, \boldsymbol{z}) 2^{\alpha \hat{k}_{\gamma_d}} \prod_{j=1}^{\hat{k}_{\gamma_d}} \gamma_{d,j}. \end{aligned}$$

We know from existing results for the integration problem in the worst case setting, see also (12) of [4], that the generating vector $\boldsymbol{z}^{(0)}$ constructed by Algorithm 11 satisfies

$$F_d(\boldsymbol{0}, \boldsymbol{z}^{(0)}) \leqslant (n-1)^{-1/\lambda} \prod_{j=1}^{d} \left(1 + 2\zeta(\alpha\lambda)\gamma_{d,j}^{\lambda} \right)^{1/\lambda} \quad \forall \lambda \in \left(\tfrac{1}{\alpha}, 1 \right].$$

The inequality in Lemma 12 now follows from Corollary 4 by combining the bounds above and making use of Lemma 6. □

Before we proceed to choose M and n, we first prove a simple result concerning $\hat{k}_{\gamma_d}$ defined in Lemma 12. Note that this result holds when $2^{-\alpha}$ is replaced by any fixed number $w > 0$.

Lemma 13. *If* $s_{\boldsymbol{\gamma}} < 1$ *then*

$$\sup_{d \geqslant 1} \hat{k}_{\gamma_d} < \infty.$$

If $t_{\boldsymbol{\gamma}} < 1$ *then*

$$\sup_{d \geqslant 1} \frac{\hat{k}_{\gamma_d}}{\log(d+1)} < \infty.$$

Proof. Let $w = 2^{-\alpha}$. First we assume that $s_{\boldsymbol{\gamma}} < 1$. Then for all $d \geqslant 1$ we have

$$\sum_{j=1}^{d} \gamma_{d,j} \leqslant Y \quad \text{where } Y := \sup_{d \geqslant 1} \sum_{j=1}^{d} \gamma_{d,j} < \infty.$$

Let $d_0 := \lceil Y/w \rceil$. Since the $\gamma_{d,j}$ are non-increasing in j, for $d \geqslant d_0$ we have

$$\gamma_{d,d_0} \leqslant \frac{1}{d_0} \sum_{j=1}^{d_0} \gamma_{d,j} \leqslant \frac{Y}{d_0} \leqslant w.$$

In other words, $\hat{k}_{\gamma_d} < d_0$ for all $d \geqslant d_0$, and hence $\sup_{d \geqslant 1} \hat{k}_{\gamma_d} < \infty$.

Now we assume that $t_\gamma < 1$. Then for all $d \geqslant 1$ we have

$$\frac{\sum_{j=1}^{d} \gamma_{d,j}}{\log(d+1)} \leqslant X \quad \text{where } X := \sup_{d \geqslant 1} \frac{\sum_{j=1}^{d} \gamma_{d,j}}{\log(d+1)} < \infty.$$

Let $d_1 := \lceil X/w \rceil$ and $d_2 := \max\{d \geqslant 1 : d/\log(d+1) < d_1\}$. Then $d_2 \geqslant d_1$, and for $d > d_2$ we have $d_1 \log(d+1) \leqslant d$, and hence

$$\gamma_{d,\lceil d_1 \log(d+1) \rceil} \leqslant \frac{1}{\lceil d_1 \log(d+1) \rceil} \sum_{j=1}^{\lceil d_1 \log(d+1) \rceil} \gamma_{d,j} \leqslant \frac{\sum_{j=1}^{d} \gamma_{d,j}}{d_1 \log(d+1)} \leqslant \frac{X}{d_1} \leqslant w.$$

It follows that $\hat{k}_{\gamma_d} < d_1 \log(d+1)$ for all $d > d_2$. Hence

$$\sup_{d \geqslant 1} \frac{\hat{k}_{\gamma_d}}{\log(d+1)} \leqslant \max\left(\frac{d_2}{\log(d_2+1)}, d_1 \right) = d_1 < \infty.$$

This completes the proof. □

We want to find small M and n for which the upper bound of the errors $e^{\text{avg}\,(i)}_{n,d,M}(\boldsymbol{z}^{(0)})$, $i = 1, 2$, given in Lemma 12 is at most $\varepsilon e^{\text{avg}}_{0,d}$. Using the same line of argument as in Section 5.2, for $\tau \in (1/\alpha, 1)$ we choose M as in (26) and

$$n = \text{pr}\left(\left(\frac{c_{5,d,\tau} M^{\frac{1-\tau+\alpha\tau+\alpha\tau^2}{\alpha\tau}}}{c_{2,d,\tau}} \right)^{\tau} + 1 \right), \tag{35}$$

where $c_{2,d,\tau}$ is defined as in (25) and

$$c_{5,d,\tau} := 2^{\alpha \hat{k}_{\gamma_d}} \left(\prod_{j=1}^{\hat{k}_{\gamma_d}} \gamma_{d,j} \right) \prod_{j=1}^{d} \left(1 + 2\zeta(\alpha\tau) \gamma_{d,j}^{\tau} \right)^{1+\frac{1}{\tau}}.$$

Substituting (26) into (35), we obtain

$$n = \text{pr}\left(\max\left(\mathcal{C}_{4,d,\tau} \varepsilon^{-\frac{2\tau}{1-\tau}[1+\tau(\alpha\tau+\alpha-1)]} + 1, \mathcal{C}_{5,d,\tau} \right) \right),$$

where

$$\begin{aligned}
\mathcal{C}_{4,d,\tau} &:= \left(\frac{c_{5,d,\tau}}{c_{2,d,\tau}}\left(\frac{2c_{2,d,\tau}}{[e_{0,d}^{\mathrm{avg}}]^2}\right)^{\frac{1-\tau+\alpha\tau+\alpha\tau^2}{1-\tau}}\right)^{\tau} \\
&\leqslant 2^{\tau+\frac{\alpha\tau^2(1+\tau)}{1-\tau}}\left(\frac{\tau}{1-\tau}\right)^{\frac{\alpha\tau^2(1+\tau)}{1-\tau}}\left(\frac{1}{\gamma_0}\right)^{\tau(1+\tau)} \\
&\quad\times\exp\left(\alpha\tau(\log 2)\hat{k}_{\gamma_d}+2\zeta(\alpha\tau)\left(1+\tau+\tfrac{\alpha\tau(1+\tau)}{1-\tau}\right)\sum_{j=1}^{d}\gamma_{d,j}^{\tau}\right) \\
&= 2^{\tau+\frac{\alpha\tau^2(1+\tau)}{1-\tau}}\left(\frac{\tau}{1-\tau}\right)^{\frac{\alpha\tau^2(1+\tau)}{1-\tau}}\left(\frac{1}{\gamma_0}\right)^{\tau(1+\tau)} \\
&\quad\times(d+1)^{\alpha\tau(\log 2)\frac{\hat{k}_{\gamma_d}}{\log(d+1)}+2\zeta(\alpha\tau)\left(1+\tau+\frac{\alpha\tau(1+\tau)}{1-\tau}\right)\frac{\sum_{j=1}^{d}\gamma_{d,j}^{\tau}}{\log(d+1)}},
\end{aligned}$$

and

$$\begin{aligned}
\mathcal{C}_{5,d,\tau} &:= \left(\frac{c_{5,d,\tau}}{c_{2,d,\tau}}\right)^{\tau}+1 \leqslant \left(\frac{1-\tau}{\tau}\right)^{\tau}\exp\left(\alpha\tau(\log 2)\hat{k}_{\gamma_d}+2\zeta(\alpha\tau)\tau\sum_{j=1}^{d}\gamma_{d,j}^{\tau}\right)+1 \\
&= \left(\frac{1-\tau}{\tau}\right)^{\tau}(d+1)^{\alpha\tau(\log 2)\frac{\hat{k}_{\gamma_d}}{\log(d+1)}+2\zeta(\alpha\tau)\tau\frac{\sum_{j=1}^{d}\gamma_{d,j}^{\tau}}{\log(d+1)}}+1.
\end{aligned}$$

If $s_\gamma < 1$, then by choosing $\tau \in (\max(1/\alpha, s_\gamma), 1)$ we have $\sup_{d\geqslant 1}\mathcal{C}_{4,d,\tau} < \infty$ and $\sup_{d\geqslant 1}\mathcal{C}_{5,d,\tau} < \infty$. On the other hand if $t_\gamma < 1$, then we choose $\tau \in (\max(1/\alpha, t_\gamma), 1)$ and this leads to $\mathcal{C}_{4,d,\tau} = O(d^a)$ and $\mathcal{C}_{5,d,\tau} = O(d^c)$, where $a > c$ and a can be arbitrarily close to $\alpha\tau(\log 2)\hat{K}+2\zeta(\alpha\tau)[1+\tau+\alpha\tau(1+\tau)/(1-\tau)]R_\tau$, with $\hat{K} = \limsup_{d\to\infty}\hat{k}_{\gamma_d}/\log(d+1)$, and R_τ given in Theorem 1.

Now we switch to the artificial smoothness parameter $\tilde{\alpha}$ introduced in Section 5.3 and consider the average case errors of the modified algorithms $A^{(i)}_{n,d,M,\tilde{\alpha}}$, $i = 1, 2$. As in Section 5.3, all our analysis above can be carried forward by simply replacing every occurrence of α with $\tilde{\alpha}$. We obtain the following theorem.

Theorem 14. *Consider multivariate approximation in the average case setting defined as in Theorem* 1.

(a) *Suppose that* $s_\gamma < 1$ *and define*

$$\tilde{\alpha} := \min(\alpha, 1/s_\gamma).$$

Given $d\geqslant 1$, $\varepsilon \in (0,1)$ *and* $\tau \in (1/\tilde{\alpha}, 1)$, *for each* $i = 1, 2$, *the approximation algorithm* $A^{(i)}_{n,d,M,\tilde{\alpha}}$, *with* M *and* n *given by* (26) *and* (35) *where every occurrence of* α *is replaced by* $\tilde{\alpha}$, *and with generating vector* $\boldsymbol{z}^{(0)}$ *constructed by the modified Algorithm* 11, *achieves the error bound* $e^{\mathrm{avg}\,(i)}_{n,d,M,\tilde{\alpha}}(\boldsymbol{z}^{(0)}) \leqslant \varepsilon\, e^{\mathrm{avg}}_{0,d}$ *using*

$$n = O\left(\varepsilon^{-\frac{2\tau}{1-\tau}[1+\tau(\tilde{\alpha}\tau+\tilde{\alpha}-1)]}\right)$$

function values. The implied factor in the big O notation is independent of ε *and* d *but depends on* τ. *The exponent of* ε^{-1} *can be arbitrarily close to*

$$2\, p^{\mathrm{avg}}(\Lambda^{\mathrm{std}}).$$

(b) *Suppose that* $t_\gamma < 1$ *and define*

$$\tilde{\alpha} := \min(\alpha, 1/t_\gamma).$$

Given $d \geqslant 1$, $\varepsilon \in (0, 1)$, *and* $\tau \in (1/\tilde{\alpha}, 1)$, *for each* $i = 1, 2$, *the approximation algorithm* $A^{(i)}_{n,d,M,\tilde{\alpha}}$, *with* M *and* n *given by* (26) *and* (35) *where every occurrence of* α *is replaced by* $\tilde{\alpha}$, *and with generating vector* $\boldsymbol{z}^{(0)}$ *constructed by the modified Algorithm* 11, *achieves the error bound* $e^{\mathrm{avg}\,(i)}_{n,d,M,\tilde{\alpha}}(\boldsymbol{z}^{(0)}) \leqslant \varepsilon\, e^{\mathrm{avg}}_{0,d}$ *using*

$$n = O\left(d^a\, \varepsilon^{-\frac{2\tau}{1-\tau}[1+\tau(\tilde{\alpha}\tau+\tilde{\alpha}-1)]}\right)$$

function values, *where* a *is arbitrarily close to*

$$\tilde{\alpha}\tau(\log 2)\hat{K} + 2\zeta(\tilde{\alpha}\tau)\left(1 + \tau + \tfrac{\tilde{\alpha}\tau(1+\tau)}{1-\tau}\right) R_\tau,$$

with $\hat{K} = \limsup_{d\to\infty} \hat{k}_{\gamma_d}/\log(d+1) < \infty$, *and* R_τ *given in Theorem* 1. *The implied factor in the big O notation is independent of* ε *and* d *but depends on* τ *and* a.

7. Comparison with the worst case setting

In this section we briefly compare the results obtained in this paper for the approximation problem in the average case setting with the results obtained in [4] for the approximation problem in the worst case setting for weighted Korobov spaces.

The weighted Korobov space $H_{d,\beta,\boldsymbol{\eta}_d}$ with $\beta > 1$ consists of 1-periodic real functions defined on $[0, 1]^d$ for which

$$\|f\|_{H_{d,\beta,\boldsymbol{\eta}_d}} = \left(\sum_{\boldsymbol{h}\in\mathbb{Z}^d} r_d(\beta, \boldsymbol{\eta}_d, \boldsymbol{h})|\hat{f}(\boldsymbol{h})|^2\right)^{1/2} < \infty, \tag{36}$$

with $r_d(\beta, \boldsymbol{\eta}_d, \boldsymbol{h})$ given by (5) when we replace α by β, and the weight vector $\boldsymbol{\gamma}_d$ by $\boldsymbol{\eta}_d = (\eta_{d,1}, \eta_{d,2}, \ldots, \eta_{d,d})$ with $1 \geqslant \eta_{d,1} \geqslant \eta_{d,2} \geqslant \cdots \geqslant \eta_{d,d} > 0$. The reproducing kernel of $H_{d,\beta,\boldsymbol{\eta}_d}$ is $K_{d,\beta,\boldsymbol{\eta}_d}$ given by (4) with the same change of α to β, and $\boldsymbol{\gamma}_d$ to $\boldsymbol{\eta}_d$. Note that the weights in [4] do not depend on d, but the results can be generalized in much the same way as in the previous sections.

The approximation problem has been studied in the worst case setting for the unit ball of the space $H_{d,\beta,\boldsymbol{\eta}_d}$. For technical reasons we consider this problem for the ball of radius q. Clearly, all tractability results obtained for the unit ball are also valid for the ball of radius q. The reason is that the initial worst case error as well as the worst case error of all linear algorithms are simply multiplied by q, and therefore the reduction of the initial error by a factor ε is independent of q.

The *worst case error* of a linear algorithm (8) is now defined as

$$e^{\mathrm{wor}}(A_{n,d}) := \sup_{\substack{f\in H_{d,\beta,\boldsymbol{\eta}_d} \\ \|f\|_{H_{d,\beta,\boldsymbol{\eta}_d}} \leqslant q}} \|f - A_{n,d}(f)\|_{\mathcal{G}} = q\, \|\mathrm{EMB}_d - A_{n,d}\|.$$

For $n = 0$ and $A_{0,d} = 0$, the initial error is

$$e^{\mathrm{wor}}_{0,d} = q\,\|\mathrm{EMB}_d\|.$$

Let

$$n^{\mathrm{wor}}(\varepsilon, d, \Lambda) := \min\{n : \exists A_{n,d} \text{ with } L_k \in \Lambda \text{ such that } e^{\mathrm{wor}}(A_{n,d}) \leqslant \varepsilon\, e^{\mathrm{wor}}_{0,d}\}$$

be the minimal number of evaluations from the class Λ which is needed to reduce the initial error in the worst case setting by a factor ε for the approximation problem in the worst case setting for the ball of $H_{d,\beta,\boldsymbol{\eta}_d}$ of radius q.

The approximation problem in the worst case setting is *tractable* in the class Λ iff

$$n^{\mathrm{wor}}(\varepsilon, d, \Lambda) \leqslant C\varepsilon^{-p}\, d^a \quad \forall d = 1, 2, \ldots, \forall \varepsilon \in (0, 1), \tag{37}$$

where C, p and a are non-negative numbers independent of ε and d. The approximation problem is *strongly tractable* if (37) holds with $a = 0$. In this case, the infimum of p from (37) is called the *exponent of strong tractability*, and is denoted by $p^{\mathrm{wor}}(\Lambda)$.

It is known, see [6,18], that the approximation problem is strongly tractable in either class Λ^{all} or Λ^{std} iff

$$\sup_{d \geqslant 1} \sum_{j=1}^{d} \eta_{d,j} < \infty.$$

When this holds then

$$p^{\mathrm{wor}}(\Lambda^{\mathrm{all}}) = 2\max(1/\beta, s_{\boldsymbol{\eta}}) \quad \text{and} \quad p^{\mathrm{wor}}(\Lambda^{\mathrm{std}}) \in \Big[p^{\mathrm{wor}}(\Lambda^{\mathrm{all}}),\, p^{\mathrm{wor}}(\Lambda^{\mathrm{all}}) + 2\Big],$$

where $s_{\boldsymbol{\eta}}$ is given in Theorem 1. The approximation problem is tractable in either class Λ^{all} or Λ^{std} iff

$$L_{\boldsymbol{\eta}} := \limsup_{d\to\infty} \frac{\sum_{j=1}^{d} \eta_{d,j}}{\log(d+1)} < \infty. \tag{38}$$

In [4], we constructed lattice rule algorithms whose worst case errors achieve strong tractability or tractability error bounds. More precisely, for strong tractability we need $O(\varepsilon^{-p})$ function values whereas for tractability we need $O(d^a\varepsilon^{-4})$ function values, with the factors in the big O notation independent of ε and d but dependent on p and a, respectively, where p is arbitrarily close to $2p^{\mathrm{wor}}(\Lambda^{\mathrm{all}})$ and a is arbitrarily close to $4\zeta(\beta)L_{\boldsymbol{\eta}}$.

We now discuss the average case setting of the approximation problem for the weighted Korobov space $H_{d,\beta,\boldsymbol{\eta}_d}$. Since $H_{d,\beta,\boldsymbol{\eta}_d}$ is a (Borel) subset of $\mathcal{F} = C([0,1]^d)$, we equip the linear space $H_{d,\beta,\boldsymbol{\eta}_d}$ with the Gaussian measure μ_d with mean element zero and covariance function K_{d,α,γ_d} defined in Section 2. We have

$$\mu_d(H_{d,\beta,\boldsymbol{\eta}_d}) = \nu_d(H_{d,\beta,\boldsymbol{\eta}_d}),$$

where, as before, $\nu_d = \mu_d\mathrm{EMB}_d^{-1}$ is a Gaussian probability measure on $\mathcal{G} = L_2([0,1]^d)$.

We now show that ν_d is a probability measure on $H_{d,\beta,\boldsymbol{\eta}_d}$ iff $\alpha > \beta + 1$. To do this, we apply the Kolmogorov (or zero-one) principle, see [9] as well as [15, p. 308], which says that for a linear subspace B of the separable Hilbert space $\mathcal{G}$ we have $\nu_d(B) = 0$ or $\nu_d(B) = 1$. Furthermore, we

know when the measure of B is zero and when it is one: namely, if we denote the orthonormal system of eigenelements $\{z_{\boldsymbol{h}}\}$ of C_{ν_d}, with $C_{\nu_d} z_{\boldsymbol{h}} = \lambda_{\boldsymbol{h}} z_{\boldsymbol{h}}$, and if

$$B = \left\{ g \in \mathcal{G} : \sum_{\boldsymbol{h} \in \mathbb{Z}^d} |\langle g, z_{\boldsymbol{h}} \rangle|^2 a_{\boldsymbol{h}} < \infty \right\}$$

for some non-negative $a_{\boldsymbol{h}}$, then

$$\nu_d(B) = 1 \quad \text{iff} \sum_{\boldsymbol{h} \in \mathbb{Z}^d} a_{\boldsymbol{h}} \lambda_{\boldsymbol{h}} < \infty.$$

In our case, $B = H_{d,\beta,\boldsymbol{\eta}_d}$ and we have $\{a_{\boldsymbol{h}}\} = \{r_d(\beta, \boldsymbol{\eta}_d, \boldsymbol{h})\}$ and $\{\lambda_{\boldsymbol{h}}\} = \{1/r_d(\alpha, \gamma_d, \boldsymbol{h})\}$. Furthermore,

$$\sum_{\boldsymbol{h} \in \mathbb{Z}^d} \frac{r_d(\beta, \boldsymbol{\eta}_d, \boldsymbol{h})}{r_d(\alpha, \gamma_d, \boldsymbol{h})} = \prod_{j=1}^{d} \left(1 + 2\zeta(\alpha - \beta) \frac{\gamma_{d,j}}{\eta_{d,j}} \right) < \infty \quad \text{iff } \alpha > \beta + 1.$$

That is why from now on we assume that $\alpha > \beta + 1$. For such α, it follows that ν_d restricted to $H_{d,\beta,\boldsymbol{\eta}_d}$ is a Gaussian probability measure with mean element zero and covariance function K_{d,α,γ_d}. From this it follows that the average case setting for the approximation problem over the space $H_{d,\beta,\boldsymbol{\eta}_d}$ is the same as over the space $\mathcal{F} = C([0,1]^d)$.

We want to compare the results for approximation in the worst case and average case settings defined over the ball of $H_{d,\beta,\boldsymbol{\eta}_d}$ of radius q,

$$\text{Ball}_q = \{ f \in H_{d,\beta,\boldsymbol{\eta}_d} : \|f\|_{H_{d,\beta,\boldsymbol{\eta}_d}} \leqslant q \}.$$

Thus we need to equip Ball_q with a probability measure. To achieve this, we normalize the measure ν_d and take

$$\nu_{d,q}(A) = \frac{\nu_d\left(A \cap \text{Ball}_q\right)}{\nu_d\left(\text{Ball}_q\right)} \tag{39}$$

for any Borel set A of $\mathcal{G}$.

Let $n_q^{\text{avg}}(\varepsilon, d, \Lambda)$ denote the minimal number of evaluations needed from the class $\Lambda \subset \{\Lambda^{\text{std}}, \Lambda^{\text{all}}\}$ to reduce the initial error by a factor ε for the approximation problem in the average case setting for the ball of $H_{d,\beta,\boldsymbol{\eta}_d}$ of radius q which is equipped with the measure $\nu_{d,q}$. Let $n^{\text{avg}}(\varepsilon, d, \Lambda) = n_\infty^{\text{avg}}(\varepsilon, d, \Lambda)$ denote the minimal number of evaluation when we consider the approximation problem for the whole space $H_{d,\beta,\boldsymbol{\eta}_d}$, i.e., with $q = \infty$ and $\nu_{d,\infty} = \nu_d$. Note that for $q = \infty$ we use the same notation as in the previous sections since the average case setting for the approximation problems for $\mathcal{F} = C([0,1]^d)$ and for the whole space $H_{d,\beta,\boldsymbol{\eta}_d}$ are the same.

We are ready to prove the following theorem.

Theorem 15. *Let $\beta > 1$, $\alpha > \beta + 1$, and let*

$$P := \sup_{d \geqslant 1} \sum_{j=1}^{d} \frac{\gamma_{d,j}}{\eta_{d,j}} < \infty. \tag{40}$$

Then there exists a positive number q_0 depending only on $\alpha - \beta$ and P and independent of ε and d such that the approximation problem in the average case setting for the whole space $H_{d,\beta,\boldsymbol{\eta}_d}$

equipped with the measure ν_d and for the ball of $H_{d,\beta,\boldsymbol{\eta}_d}$ of radius $q \geqslant q_0$ equipped with the measure $\nu_{d,q}$ are essentially the same, in that

$$c_1\, n^{\mathrm{avg}}(c_2\varepsilon, d, \Lambda) \leqslant n_q^{\mathrm{avg}}(\varepsilon, d, \Lambda) \leqslant c_3\, n^{\mathrm{avg}}(c_4\varepsilon, d, \Lambda),$$

where

$$\begin{aligned}
c_1 &= 1 - \sqrt{10/27} = 0.391419\ldots,\\
c_2 &= \left(1 + \sqrt{3/10}\right)/\sqrt{1 - \sqrt{10/27}} = 2.473841\ldots,\\
c_3 &= 10/9 = 1.111111\ldots,\\
c_4 &= \sqrt{9/10}\sqrt{1 - \sqrt{3/10}} = 0.638004\ldots.
\end{aligned}$$

Proof. We use the known fact that the average case setting for the ball of radius q is essentially the same as the average case setting for the whole space if q is sufficiently large compared to the trace of the covariance operator, see [15, Chapter 6, Section 5.8]. In particular, we need to choose q such that $1 - \nu_d(\mathrm{Ball}_q) \leqslant x := 0.1$. Arguing as in [15, p. 258], we observe that for any real a,

$$\begin{aligned}
1 - \nu_d(\mathrm{Ball}_q) &\leqslant \int_{\|f\|_{H_{d,\beta,\boldsymbol{\eta}_d}} > q} \exp\left(a\left(\|f\|^2_{H_{d,\beta,\boldsymbol{\eta}_d}} - q^2\right)\right) \nu_d(\mathrm{d}f)\\
&\leqslant \exp\left(-q^2 a\right) \int_{\mathcal{G}} \exp\left(a\|f\|^2_{H_{d,\beta,\boldsymbol{\eta}_d}}\right) \nu_d(\mathrm{d}f).
\end{aligned}$$

To evaluate the last integral we first use [15, p. 258, Remark 5.8.1], which states that for a Gaussian measure ν_d on a separable Hilbert space $\mathcal{G}$

$$\int_{\mathcal{G}} \exp\left(a\|f\|^2_{\mathcal{G}}\right) \nu_d(\mathrm{d}f) = \prod_{i=1}^{\infty} \frac{1}{\sqrt{1 - 2a\lambda_{d,i}}},$$

if $\{\lambda_{d,i}\}$ are the eigenvalues of the covariance operator C_{ν_d}, and $a < (2\max_{i \geqslant 1} \lambda_{d,i})^{-1}$. In our particular case the eigenvalues of the covariance operator are given by (7), but the result needs to be modified because the norm of f in the exponent is the norm in $H_{d,\beta,\boldsymbol{\eta}_d}$ given by (36), not the norm in $\mathcal{G} = L_2([0,1]^d)$. With the appropriate modification, the result becomes

$$\int_{\mathcal{G}} \exp\left(a\|f\|^2_{H_{d,\beta,\boldsymbol{\eta}_d}}\right) \nu_d(\mathrm{d}f) = \prod_{\boldsymbol{h} \in \mathbb{Z}^d} \left(1 - 2a\frac{r_d(\beta, \boldsymbol{\eta}_d, \boldsymbol{h})}{r_d(\alpha, \boldsymbol{\gamma}_d, \boldsymbol{h})}\right)^{-1/2},$$

if $a < [2\max_{\boldsymbol{h} \in \mathbb{Z}^d} r_d(\beta, \boldsymbol{\eta}_d, \boldsymbol{h})/r_d(\alpha, \boldsymbol{\gamma}_d, \boldsymbol{h})]^{-1}$. It follows that, under the tighter condition $a < [2\sum_{\boldsymbol{h} \in \mathbb{Z}^d} r_d(\beta, \boldsymbol{\eta}_d, \boldsymbol{h})/r_d(\alpha, \boldsymbol{\gamma}_d, \boldsymbol{h})]^{-1}$, we have

$$\begin{aligned}
\int_{\mathcal{G}} \exp\left(a\|f\|^2_{H_{d,\beta,\boldsymbol{\eta}_d}}\right) \nu_d(\mathrm{d}f) &\leqslant \left(1 - 2a\sum_{\boldsymbol{h} \in \mathbb{Z}^d} \frac{r_d(\beta, \boldsymbol{\eta}_d, \boldsymbol{h})}{r_d(\alpha, \boldsymbol{\gamma}_d, \boldsymbol{h})}\right)^{-1/2}\\
&= \left(1 - 2a\prod_{j=1}^{d}\left(1 + 2\zeta(\alpha - \beta)\frac{\gamma_{d,j}}{\eta_{d,j}}\right)\right)^{-1/2}.
\end{aligned}$$

For $a = [4\prod_{j=1}^{d}(1+2\zeta(\alpha-\beta)\gamma_{d,j}/\eta_{d,j})]^{-1}$ we therefore have

$$\int_{\mathcal{G}} \exp\left(a\|f\|^2_{H_{d,\beta,\boldsymbol{\eta}_d}}\right) v_d(\mathrm{d}f) \leqslant \sqrt{2}.$$

Hence

$$1 - v_d(\mathrm{Ball}_q) \leqslant \sqrt{2}\,\exp\left(-\frac{q^2}{4\prod_{j=1}^{d}\left(1+2\zeta(\alpha-\beta)\gamma_{d,j}/\eta_{d,j}\right)}\right).$$

We need to choose q such that $1 - v_d(\mathrm{Ball}_q) \leqslant x$. Note that $P < \infty$ implies that $\limsup_{d\to\infty}\sum_{j=1}^{d}\gamma_{d,j}/\eta_{d,j}$ and $\sup_{d\geqslant 1}\prod_{j=1}^{d}(1+2\zeta(\alpha-\beta)\gamma_{d,j}/\eta_{d,j})$ are finite. Hence, there is a number q_0 depending only on how large P and $\zeta(\alpha-\beta)$ are such that for all $q \geqslant q_0$ we have $1 - v_d(\mathrm{Ball}_q) \leqslant x$. Observe that the value of $\zeta(\alpha-\beta)$ depends only on how close $\alpha-\beta$ is to 1. Hence, q_0 depends on $\alpha-\beta$ and P, but is independent of d and, of course, of ε.

Let $e^{\mathrm{avg}}_{0,d,q}$ denote the initial error for the approximation problem over the ball of $H_{d,\beta,\boldsymbol{\eta}_d}$ of radius $q \geqslant q_0$ equipped with the measure $v_{d,q}$, and let $e^{\mathrm{avg}}_{0,d} = e^{\mathrm{avg}}_{0,d,\infty}$. From [15, p. 261, Theorem 5.8.2] we know that

$$\sqrt{1-\sqrt{3x}} \leqslant \frac{e^{\mathrm{avg}}_{0,d,q}}{e^{\mathrm{avg}}_{0,d,\infty}} \leqslant 1+\sqrt{3x}.$$

From [15, p. 259, Theorem 5.8.1] we know that

$$c_1 n^{\mathrm{avg}}\left(\bar{c}_2\frac{e^{\mathrm{avg}}_{0,d,q}}{e^{\mathrm{avg}}_{0,d,\infty}}\varepsilon, d, \Lambda\right) \leqslant n^{\mathrm{avg}}_q(\varepsilon, d, \Lambda) \leqslant c_3 n^{\mathrm{avg}}\left(\bar{c}_4\frac{e^{\mathrm{avg}}_{0,d,q}}{e^{\mathrm{avg}}_{0,d,\infty}}\varepsilon, d, \Lambda\right),$$

with $c_1 = 1-\sqrt{3x}/(1-x)$, $c_3 = 1/(1-x)$, $\bar{c}_2 = 1/\sqrt{c_1}$, and $\bar{c}_4 = 1/\sqrt{c_3}$. Since

$$n^{\mathrm{avg}}\left(\bar{c}_2\frac{e^{\mathrm{avg}}_{0,d,q}}{e^{\mathrm{avg}}_{0,d,\infty}}\varepsilon, d, \Lambda\right) \geqslant n^{\mathrm{avg}}\left(\bar{c}_2\,(1+\sqrt{3x})\varepsilon, d, \Lambda\right) = n^{\mathrm{avg}}\left(c_2\,\varepsilon, d, \Lambda\right),$$

$$n^{\mathrm{avg}}\left(\bar{c}_4\frac{e^{\mathrm{avg}}_{0,d,q}}{e^{\mathrm{avg}}_{0,d,\infty}}\varepsilon, d, \Lambda\right) \leqslant n^{\mathrm{avg}}\left(\bar{c}_4\sqrt{1-\sqrt{3x}}\varepsilon, d, \Lambda\right) = n^{\mathrm{avg}}(c_4\,\varepsilon, d, \Lambda),$$

the proof is completed. □

Theorem 15 allows us to compare the tractability results for the approximation problem in the worst case and average case settings. Indeed, under the assumptions of Theorem 15, we know that $n^{\mathrm{avg}}_q(\varepsilon, d, \Lambda)$ behaves essentially as $n^{\mathrm{avg}}(\varepsilon, d, \Lambda)$ which was studied in the previous sections.

We take $\gamma_{d,j} = \eta^u_{d,j}$ for $u > 1$. Assume first that strong tractability holds in the worst case setting. Then we have $\sup_{d\geqslant 1}\sum_{j=1}^{d}\eta_{d,j} < \infty$ which implies that $s_{\boldsymbol{\eta}} \leqslant 1$. To guarantee that (40) holds we assume that $u > 1 + s_{\boldsymbol{\eta}}$. Then

$$s_\gamma = s_{\boldsymbol{\eta}}/u < s_{\boldsymbol{\eta}}/(1+s_{\boldsymbol{\eta}}) < 1.$$

This implies that we have also strong tractability in the average case setting. Since $\alpha > \beta + 1$, it is easy to check that the exponent of strong tractability in the average case setting is smaller that the exponent of strong tractability in the worst case setting:

$$p^{\mathrm{avg}}(\Lambda^{\mathrm{all}}) = \frac{2\max(1/\alpha, s_\gamma)}{1 - \max(1/\alpha, s_\gamma)} < p^{\mathrm{wor}}(\Lambda^{\mathrm{all}}) = 2\max(1/\beta, s_{\boldsymbol{\eta}}).$$

We now assume that tractability holds in the worst case setting. Then $L_{\boldsymbol{\eta}}$ defined in (38) is finite which implies that $t_{\boldsymbol{\eta}} \leqslant 1$. Furthermore, (40) holds if $u > 1+t_{\boldsymbol{\eta}}$. This implies that $t_\gamma = t_{\boldsymbol{\eta}}/(1+t_{\boldsymbol{\eta}}) < 1$, and we have also tractability in the average case setting.

Obviously, it can happen that there is no tractability in the worst case setting but we have even strong tractability in the average case setting. For example, take $\eta_{d,j} = j^{-1/2}$. Then $s_{\boldsymbol{\eta}} = 2$ and $L_{\boldsymbol{\eta}} = \infty$, and we have no tractability in the worst case setting, whereas for $u > 1 + s_{\boldsymbol{\eta}} = 3$ we have $s_\gamma = s_{\boldsymbol{\eta}}/u < 1$ which yields strong tractability in the average case setting.

8. Numerical experiments

8.1. Computational issues

Here we discuss the computational cost of Algorithm 7 and tricks for speeding up the calculations. Throughout this section, let $\boldsymbol{h} = (\boldsymbol{h}', h_s)$ and $\boldsymbol{z} = (\boldsymbol{z}', z_s)$ with $\boldsymbol{h}', \boldsymbol{z}' \in \mathbb{Z}^{s-1}$. Clearly $(\boldsymbol{h}', h_s) \in \mathcal{A}_{d,s}(M)$ iff $(\boldsymbol{h}', -h_s) \in \mathcal{A}_{d,s}(M)$. Furthermore, if $\boldsymbol{h}' \in \mathcal{A}_{d,s-1}(M)$ and we want $\boldsymbol{h} \in \mathcal{A}_{d,s}(M)$, then the magnitude of h_s can be no greater than

$$L(\boldsymbol{h}') := \left\lfloor \left(\frac{\gamma_{d,s} M}{r_{d,s-1}(\boldsymbol{h}')} \right)^{1/\alpha} \right\rfloor.$$

For computational efficiency, we write

$$E^{(1)}_{d,s}(\boldsymbol{z}) = \sum_{\boldsymbol{h}' \in \mathcal{A}_{d,s-1}(M)} \sum_{h_s = -L(\boldsymbol{h}')}^{L(\boldsymbol{h}')} \frac{F_{d,s}((\boldsymbol{h}', h_s), \boldsymbol{z})}{1 + r_{d,s}(\boldsymbol{h}', h_s) F_{d,s}((\boldsymbol{h}', h_s), \boldsymbol{z})},$$

$$E^{(2)}_{d,s}(\boldsymbol{z}) = \sum_{\boldsymbol{h}' \in \mathcal{A}_{d,s-1}(M)} \sum_{h_s = -L(\boldsymbol{h}')}^{L(\boldsymbol{h}')} F_{d,s}((\boldsymbol{h}', h_s), \boldsymbol{z}),$$

where (see [4])

$$F_{d,s}(\boldsymbol{h}, \boldsymbol{z}) = -\frac{1}{r_{d,s}(\boldsymbol{h})} + \frac{1}{n} \prod_{j=1}^{s} \left(1 + 2\zeta(\alpha)\gamma_{d,j}\right) + \frac{\Theta_{\boldsymbol{h}}(z_s)}{n},$$

with (suppressing the dependence on $\boldsymbol{z}'$)

$$\Theta_{\boldsymbol{h}}(z_s) := \sum_{k=1}^{n-1} \left(e^{-2\pi \mathrm{i} k \boldsymbol{h} \cdot \boldsymbol{z}/n} \prod_{j=1}^{s} \left(1 + \gamma_{d,j}\, \omega\left(\tfrac{k z_j}{n}\right)\right)\right) \quad \text{and} \quad \omega(x) := \sum_{\substack{\ell=-\infty \\ \ell \neq 0}}^{\infty} \frac{e^{2\pi \mathrm{i} \ell x}}{|\ell|^\alpha}.$$

To avoid working with complex numbers, we use basic properties of sine and cosine and the symmetry of $\omega(x)$ about $x = 1/2$ to write

$$\Theta_{\boldsymbol{h}}(z_s) = \sum_{k=1}^{n-1} \left(\left[\cos\left(\tfrac{2\pi k \boldsymbol{h}'\cdot \boldsymbol{z}'}{n}\right) \cos\left(\tfrac{2\pi k h_s z_s}{n}\right) - \sin\left(\tfrac{2\pi k \boldsymbol{h}'\cdot \boldsymbol{z}'}{n}\right) \sin\left(\tfrac{2\pi k h_s z_s}{n}\right) \right] \times \prod_{j=1}^{s} \left(1 + \gamma_{d,j}\,\omega\left(\tfrac{k z_j}{n}\right)\right) \right). \tag{41}$$

Note that for fixed z_s, the k and $n-k$ terms in the sum (41) are equal. Furthermore, we observe that

$$F_{d,s}((\boldsymbol{h}', h_s), (\boldsymbol{z}', n - z_s)) = F_{d,s}((\boldsymbol{h}', -h_s), (\boldsymbol{z}', z_s)),$$

which implies

$$E^{(i)}_{d,s}(\boldsymbol{z}', z_s) = E^{(i)}_{d,s}(\boldsymbol{z}', n - z_s), \quad i = 1, 2.$$

Thus it suffices to search through $z_s \in \{1, 2, \ldots, (n-1)/2\}$.

When α is an even integer, $\omega(x)$ can be easily computed, since

$$\omega(x) = \frac{(2\pi)^\alpha}{(-1)^{\frac{\alpha}{2}+1}\alpha!}\mathrm{Ber}_\alpha(x),$$

where Ber_α is the Bernoulli polynomial of degree α. If we store $r_{d,s-1}(\boldsymbol{h}')$ and $\boldsymbol{h}'\cdot\boldsymbol{z}'$ for all $\boldsymbol{h}' \in \mathcal{A}_{d,s-1}(M)$ as well as the products $\prod_{j=1}^{s-1}(1 + \gamma_{d,j}\omega(kz_j/n))$ for all $k \in \{1, 2, \ldots, n-1\}$, then the cost to evaluate $E^{(i)}_{d,s}(\boldsymbol{z})$ for one z_s is $O(|\mathcal{A}_{d,s}(M)|\, n)$ operations. (There is no need to store the vectors $\boldsymbol{h}'$.) Thus the total cost for the CBC construction up to dimension d is $O(|\mathcal{A}_d(M)|\, n^2 d)$ operations, at the expense of $O(|\mathcal{A}_d(M)| + n)$ storage.

To further speed up the computation, we need to write the error calculation for all z_s as some matrix–vector products with *circulant* matrices. The precise detail can be found in [8]. It is important to note that $\boldsymbol{\Theta}_{\boldsymbol{h}}(z_s) \neq \boldsymbol{\Theta}_{\boldsymbol{h}}(n - z_s)$ because $\sin(2\pi x)$ is antisymmetric about $x = 1/2$. As a result, if we consider only $z_s \in \{1, 2, \ldots, (n-1)/2\}$, then the error calculation cannot be expressed in terms of circulant matrix–vector products. Therefore, we must consider all $z_s \in \{1, 2, \ldots, n-1\}$ even though z_s and $n - z_s$ give the same error. Expressing (41) as a vector for $z_s \in \{1, 2, \ldots, n-1\}$, we have

$$\vec{\boldsymbol{\Theta}}_{\boldsymbol{h}} = \boldsymbol{\Omega}^{\mathrm{c}}_{h_s} \boldsymbol{p}^{\mathrm{c}}_{\boldsymbol{h}'} - \boldsymbol{\Omega}^{\mathrm{s}}_{h_s} \boldsymbol{p}^{\mathrm{s}}_{\boldsymbol{h}'},$$

where the entries of the $(n-1)$ by $(n-1)$ matrices and the vectors are given by

$$\Omega^{\mathrm{c}}_{h_s}(z_s, k) = \cos\left(2\pi h_s \tfrac{k z_s}{n}\right)\left(1 + \gamma_{d,s}\,\omega\left(\tfrac{k z_s}{n}\right)\right),$$

$$\Omega^{\mathrm{s}}_{h_s}(z_s, k) = \sin\left(2\pi h_s \tfrac{k z_s}{n}\right)\left(1 + \gamma_{d,s}\,\omega\left(\tfrac{k z_s}{n}\right)\right),$$

$$p^{\mathrm{c}}_{\boldsymbol{h}'}(k) = \cos\left(\tfrac{2\pi k \boldsymbol{h}'\cdot\boldsymbol{z}'}{n}\right) \prod_{j=1}^{s-1}\left(1 + \gamma_{d,j}\,\omega\left(\tfrac{k z_j}{n}\right)\right),$$

$$p^{\mathrm{s}}_{\boldsymbol{h}'}(k) = \sin\left(\tfrac{2\pi k \boldsymbol{h}'\cdot\boldsymbol{z}'}{n}\right) \prod_{j=1}^{s-1}\left(1 + \gamma_{d,j}\,\omega\left(\tfrac{k z_j}{n}\right)\right).$$

Clearly the entries of the matrices depend only on $kz_s \bmod n$. By a reordering of the z_s and k indices, both matrices can be brought into circulant form. More precisely, since n is prime, there is a generator g which generates the multiplicative group modulo n, that is, $\{g^i \bmod n : 0 \leqslant i \leqslant n-2\} = \{1, 2, \ldots, n-1\}$. To get a circulant matrix, we order z_s and k according to g^i and $g^{-i'}$, respectively, for $0 \leqslant i, i' \leqslant n-2$. The matrix–vector products with circulant matrices can be done in $O(n \log n)$ operations, thus reducing the total cost for the CBC construction to $O(|\mathcal{A}_d(M)| n \log n\, d)$ operations.

In particular, the CBC construction based on $E^{(2)}_{d,s}(\boldsymbol{z})$ can be simplified due to the linearity in the error expression. We can actually use $\sum_{\boldsymbol{h} \in \mathcal{A}_{d,s}(M)} \Theta_{\boldsymbol{h}}(z_s)$ as our search criterion, and the search can also be restricted to $z_s \in \{1, 2, \ldots, (n-1)/2\}$. Expressed as a vector of $z_s \in \{1, 2, \ldots, (n-1)/2\}$, we have

$$\sum_{\boldsymbol{h} \in \mathcal{A}_{d,s}(M)} \vec{\Theta}_{\boldsymbol{h}} = \sum_{\boldsymbol{h}' \in \mathcal{A}_{d,s-1}(M)} \boldsymbol{\Omega}_{\boldsymbol{h}'} \boldsymbol{p}_{\boldsymbol{h}'},$$

where the entries of the $(n-1)/2$ by $(n-1)/2$ matrix and the vector are given by

$$\Omega_{\boldsymbol{h}'}(z_s, k) = \left(1 + 2 \sum_{h_s=1}^{L(\boldsymbol{h}')} \cos\left(2\pi h_s \tfrac{kz_s}{n}\right)\right)\left(1 + \gamma_{d,s}\, \omega\left(\tfrac{kz_s}{n}\right)\right),$$

$$p_{\boldsymbol{h}'}(k) = \cos\left(\tfrac{2\pi k \boldsymbol{h}' \cdot \boldsymbol{z}'}{n}\right) \prod_{j=1}^{s-1} \left(1 + \gamma_{d,j}\, \omega\left(\tfrac{kz_j}{n}\right)\right).$$

Note that $1 + 2\sum_{h=1}^{L} \cos(2\pi h t) = 2\cos(\pi h t)\sin(\pi(h+1)t)/\sin(\pi t) - 1$ for $t \in (0, 1)$. The total computational cost is again $O(|\mathcal{A}_d(M)| n \log n\, d)$ operations, but the implied factor is smaller.

8.2. Numerical results

Following from our theory, M and n should be chosen according to (26) and (27), with respect to suitable parameters ε, κ, τ, and δ. Unfortunately, this leads to enormous M and n, and the computation becomes infeasible. This is because our error bounds are not sharp with respect to these parameters.

We carried out some preliminary calculations with

$$d = 20,\ \gamma_{d,j} = 0.1 \times 0.75^{j-1},\ \alpha \in \{2, 4, 6\},\ M = 5000,\ n \in \{1009, 2003, 4001\},$$

just to get a feel of how the CBC algorithms work.

In Table 1 we present the results from Algorithm 7(1), i.e. based on the criterion $E^{(1)}_{d,s}(\boldsymbol{z})$, for $\alpha = 2$ and $n = 4001$. The first term of the average case error is independent of n and $\boldsymbol{z}$, and it can be computed using

$$\sum_{\boldsymbol{h} \notin \mathcal{A}_{d,s}(M)} \frac{1}{r_{d,s}(\boldsymbol{h})} = [e^{\mathrm{avg}}_{0,d,s}]^2 - \sum_{\boldsymbol{h} \in \mathcal{A}_{d,s}(M)} \frac{1}{r_{d,s}(\boldsymbol{h})},$$

where $e^{\mathrm{avg}}_{0,d,s} := \prod_{j=1}^{s} (1 + 2\zeta(\alpha)\gamma_{d,j})^{1/2}$ denotes the s-dimensional truncated variant of the initial error $e^{\mathrm{avg}}_{0,d}$. We see from the numbers that the growth of $|\mathcal{A}_{d,s}(5000)|$ slows down as s increases. This is consistent with decaying weights. The first term of the average case error indicates that the choice of $M = 5000$ corresponds roughly to $\varepsilon = 0.64$. The second term of the average

Table 1
Algorithm 7(1) with $n = 4001$ and $\alpha = 2$

s	$\lvert\mathcal{A}_{20,s}(5000)\rvert$	$\sum_{\boldsymbol{h}\notin\mathcal{A}_{20,s}(5000)} \frac{1}{r_{20,s}(2,\boldsymbol{\gamma}_{20},\boldsymbol{h})}/[e^{\mathrm{avg}}_{0,20,s}]^2$	z_s	$E^{(1)}_{20,s}(\boldsymbol{z})/[e^{\mathrm{avg}}_{0,20,s}]^2$
1	45	6.68737e−03	1	6.95869e−07
2	139	2.09592e−02	1527	5.34309e−05
3	251	4.21366e−02	1074	4.71954e−04
4	367	6.59559e−02	898	1.79457e−03
5	463	9.01138e−02	689	4.09777e−03
6	543	1.11820e−01	1670	6.76412e−03
7	609	1.30264e−01	638	9.45634e−03
8	669	1.45106e−01	751	1.22745e−02
9	715	1.57408e−01	610	1.46260e−02
10	747	1.67582e−01	524	1.64046e−02
11	773	1.75582e−01	1491	1.77833e−02
12	793	1.81924e−01	355	1.88858e−02
13	807	1.87024e−01	1603	1.97265e−02
14	817	1.91063e−01	1128	2.03270e−02
15	821	1.94424e−01	1293	2.07056e−02
16	825	1.96940e−01	1443	2.10187e−02
17	829	1.98824e−01	1472	2.12800e−02
18	831	2.00349e−01	332	2.14550e−02
19	833	2.01491e−01	1177	2.15996e−02
20	835	2.02347e−01	1689	2.17181e−02

Table 2
Comparison of results for $\alpha = 2$

$\sum_{\boldsymbol{h}\notin\mathcal{A}_{20}(5000)} \frac{1}{r_{20}(2,\boldsymbol{\gamma}_{20},\boldsymbol{h})}/[e^{\mathrm{avg}}_{0,20}]^2 = 2.02347\mathrm{e}-01$			
n	$\boldsymbol{z}$	$E^{(1)}_{20}(\boldsymbol{z})/[e^{\mathrm{avg}}_{0,20}]^2$	$E^{(2)}_{20}(\boldsymbol{z})/[e^{\mathrm{avg}}_{0,20}]^2$
1009	$\boldsymbol{z}^{(1)}$	7.80515e−02	2.06908e−01
	$\boldsymbol{z}^{(2)}$	7.87842e−02	1.92599e−01
	$\boldsymbol{z}^{(0)}$	8.04402e−02	2.15769e−01
2003	$\boldsymbol{z}^{(1)}$	4.46415e−02	7.73733e−02
	$\boldsymbol{z}^{(2)}$	4.38831e−02	7.11386e−02
	$\boldsymbol{z}^{(0)}$	4.60306e−02	8.12095e−02
4001	$\boldsymbol{z}^{(1)}$	2.17181e−02	2.83850e−02
	$\boldsymbol{z}^{(2)}$	2.18408e−02	2.79887e−02
	$\boldsymbol{z}^{(0)}$	2.31070e−02	3.11909e−02

case error appears to be one magnitude smaller than the first term, suggesting that $n = 4001$ is unnecessarily large in relation to $M = 5000$. It took just over 1 min (on a PC with Pentium IV 2.8 GHz processor) to produce the results in Table 1. Note, however, that the computation time depends critically on the choice of weights, since they control the size of the set $\mathcal{A}_d(M)$.

In Tables 2–4 we present a comparison of the results from all three algorithms: Algorithm 7(1), Algorithm 7(2), and Algorithm 11. Recall that $\boldsymbol{z}^{(1)}$, $\boldsymbol{z}^{(2)}$, and $\boldsymbol{z}^{(0)}$ denote the vectors obtained by optimizing $E^{(1)}_d(\boldsymbol{z})$, $E^{(2)}_d(\boldsymbol{z})$, and $F_d(\mathbf{0},\boldsymbol{z})$, respectively. The underlined entries represent the

Table 3
Comparison of results for $\alpha = 4$

$\sum_{\boldsymbol{h}\notin\mathcal{A}_{20}(5000)} \frac{1}{r_{20}(4,\boldsymbol{\gamma}_{20},\boldsymbol{h})}/[e^{\mathrm{avg}}_{0,20}]^2 = 5.86045\mathrm{e}-02$			
n	$\boldsymbol{z}$	$E^{(1)}_{20}(\boldsymbol{z})/[e^{\mathrm{avg}}_{0,20}]^2$	$E^{(2)}_{20}(\boldsymbol{z})/[e^{\mathrm{avg}}_{0,20}]^2$
1009	$\boldsymbol{z}^{(1)}$	7.87141e−03	9.61386e−03
	$\boldsymbol{z}^{(2)}$	7.89395e−03	9.39313e−03
	$\boldsymbol{z}^{(0)}$	8.51012e−03	1.06525e−02
2003	$\boldsymbol{z}^{(1)}$	2.30353e−03	2.42050e−03
	$\boldsymbol{z}^{(2)}$	2.30364e−03	2.42037e−03
	$\boldsymbol{z}^{(0)}$	2.51769e−03	2.68799e−03
2001	$\boldsymbol{z}^{(1)}$	5.88871e−04	5.97576e−04
	$\boldsymbol{z}^{(2)}$	5.88871e−04	5.97576e−04
	$\boldsymbol{z}^{(0)}$	6.46782e−04	6.57251e−04

Table 4
Comparison of results for $\alpha = 6$

$\sum_{\boldsymbol{h}\notin\mathcal{A}_{20}(5000)} \frac{1}{r_{20}(6,\boldsymbol{\gamma}_{20},\boldsymbol{h})}/[e^{\mathrm{avg}}_{0,20}]^2 = 4.09159\mathrm{e}-02$			
n	$\boldsymbol{z}$	$E^{(1)}_{20}(\boldsymbol{z})/[e^{\mathrm{avg}}_{0,20}]^2$	$E^{(2)}_{20}(\boldsymbol{z})/[e^{\mathrm{avg}}_{0,20}]^2$
1009	$\boldsymbol{z}^{(1)}$	3.15063e−03	3.46769e−03
	$\boldsymbol{z}^{(2)}$	3.15063e−03	3.46769e−03
	$\boldsymbol{z}^{(0)}$	3.62148e−03	4.27081e−03
2003	$\boldsymbol{z}^{(1)}$	7.26309e−04	7.42414e−04
	$\boldsymbol{z}^{(2)}$	7.24054e−04	7.38721e−04
	$\boldsymbol{z}^{(0)}$	9.35084e−04	9.71324e−04
4001	$\boldsymbol{z}^{(1)}$	1.45628e−04	1.46191e−04
	$\boldsymbol{z}^{(2)}$	1.43076e−04	1.43627e−04
	$\boldsymbol{z}^{(0)}$	1.68453e−04	1.69649e−04

quantities which the algorithm attempted to optimize. In most cases, we see that our choices of n are unnecessarily large when compared to the choice of $M = 5000$. The vectors obtained from Algorithm 7(1) and Algorithm 7(2) appear to be of similar quality, however, the vectors obtained from Algorithm 11 are generally worse.

Acknowledgment

The authors thank Grzegorz Wasilkowski, Fred Hickernell, Xiaoyan Zeng, and Dirk Nuyens for valuable comments and suggestions. The support of the Australian Research Council under its Centres of Excellence program is gratefully acknowledged. The first author was supported by a University of New South Wales Vice-Chancellor's Postdoctoral Research Fellowship. The third author was also partially supported by the National Science Foundation under Grant DMS-0308713.

References

[1] F. J. Hickernell, G. W. Wasilkowski, H. Woźniakowski, Tractability of linear multivariate problems in the average case setting, in: A. Keller, S. Heinrich, H. Niederreiter (Eds.), Monte Carlo and Quasi-Monte Carlo Methods 2006, Springer, to appear.

[2] F.J. Hickernell, H. Woźniakowski, Integration and approximation in arbitrary dimensions, Adv. Comput. Math. 12 (2000) 25–58.

[3] F.Y. Kuo, Component-by-component constructions achieve the optimal rate of convergence for multivariate integration in weighted Korobov and Sobolev spaces, J. Complexity 19 (2003) 301–320.

[4] F.Y. Kuo, I.H. Sloan, H. Woźniakowski, Lattice rules for multivariate approximation in the worst case setting, in: H. Niederreiter, D. Talay (Eds.), Monte Carlo and Quasi-Monte Carlo Methods 2004, Springer, Berlin, 2006, pp. 289–330.

[5] D. Li, F.J. Hickernell, Trigonometric spectral collocation methods on lattices, in: S.Y. Cheng, C.-W. Shu, T. Tang (Eds.), Recent Advances in Scientific Computing and Partial Differential Equations, AMS Series in Contemporary Mathematics, vol. 330, American Mathematical Society, Providence, RI, 2003, pp. 121–132.

[6] E. Novak, I.H. Sloan, H. Woźniakowski, Tractability of approximation for weighted Korobov spaces on classical and quantum computers, Found. Comput. Math. 4 (2004) 121–156.

[7] E. Novak, H. Woźniakowski, When are integration and discrepancy tractable?, in: R.A. DeVore, A. Iserles, E. Süli (Eds.), Foundation of Computational Mathematics, Oxford 1999, Cambridge University Press, Cambridge, 2001, pp. 211–266.

[8] D. Nuyens, R. Cools, Fast algorithms for component-by-component construction of rank-1 lattice rules in shift-invariant reproducing kernel Hilbert spaces, Math. Comput. 75 (2006) 903–920.

[9] G.E. Shilov, F.D. Tin, Integral Measure and Derivative on Linear Spaces, Nauka, Moscow, 1967 (in Russian).

[10] I.H. Sloan, S. Joe, Lattice Methods for Multiple Integration, Oxford University Press, Oxford, 1994.

[11] I.H. Sloan, F.Y. Kuo, S. Joe, On the step-by-step construction of quasi-Monte Carlo integration rules that achieve strong tractability error bounds in weighted Sobolev spaces, Math. Comput. 71 (2002) 1609–1640.

[12] I.H. Sloan, F.Y. Kuo, S. Joe, Constructing randomly shifted lattice rules in weighted Sobolev spaces, SIAM J. Numer. Anal. 40 (2002) 1650–1665.

[13] I.H. Sloan, A.V. Reztsov, Component-by-component construction of good lattice rules, Math. Comput. 71 (2002) 263–273.

[14] I.H. Sloan, H. Woźniakowski, Tractability of multivariate integration for weighted Korobov classes, J. Complexity 17 (2001) 697–721.

[15] J.F. Traub, G.W. Wasilkowski, H. Woźniakowski, Information-Based Complexity, Academic Press, New York, 1988.

[16] J.F. Traub, A.G. Werschulz, Complexity and Information, Cambridge University Press, Cambridge, 1998.

[17] N.N. Vakhania, Probability Distributions on Linear Spaces, North-Holland, New York, 1981.

[18] G.W. Wasilkowski, H. Woźniakowski, On the power of standard information for weighted approximation, Found. Comput. Math. 1 (2001) 417–434.

[19] X.Y. Zeng, K.T. Leung, F.J. Hickernell, Error analysis of splines for periodic problems using lattice designs, in: H. Niederreiter, D. Talay (Eds.), Monte Carlo and Quasi-Monte Carlo Methods 2004, Springer, Berlin, 2006, pp. 501–514.