Randomized Smolyak algorithms based on digital sequences for multivariate integration

Josef Dick, Gunther Leobacher and Friedrich Pillichshammer *

Abstract

In this paper we consider Smolyak algorithms based on quasi-Monte Carlo rules for high-dimensional numerical integration. The quasi-Monte Carlo rules employed here use digital $(t, \alpha, \beta, \sigma, d)$-sequences as quadrature points. We consider the worst-case error for multivariate integration in certain Sobolev spaces and show that our quadrature rules achieve the optimal rate of convergence. By randomizing the underlying digital sequences we can also obtain a randomized Smolyak algorithm. The bound on the worst-case error holds also for the randomized algorithm in a statistical sense. Further we also show that the randomized algorithm is unbiased and that the integration error can be approximated as well. Numerical integration, quasi-Monte Carlo algorithm, Smolyak algorithm

1 Introduction

In this paper we consider numerical integration of functions over the unit cube, that is, we want to approximate an integral $\int_{[0,1]^s} f(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x}$ by a quadrature rule $\sum_{i=0}^{N-1} \omega_i f(\boldsymbol{x}_i)$, where $\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{N-1}$ are the deterministic quadrature points and $\omega_0, \ldots, \omega_{N-1}$ the weights. For high dimensions there are two main quadrature methods:

1. Quasi-Monte Carlo rules are quadrature rules where $\omega_i = 1/N$ for all $i = 0, \ldots, N-1$ and are hence equal-weight quadrature rules (see for example [3, 5, 6, 10, 11, 13] and the references therein). Here the properties of the quadrature rule depend entirely on the choice of the quadrature points.

2. Another method is based on Smolyak's construction principle [14] (see also [8, 12, 16] and the references therein) are also known by the name sparse grids. Here one normally uses a one-dimensional quadrature rule and out of this builds a high-dimensional

*The third author is supported by the Austrian Science Foundation (FWF), Project S9609, that is part of the Austrian Research Network "Analytic Combinatorics and Probabilistic Number Theory". The support of the Australian Research Council under its Centre of Excellence Program is gratefully acknowledged.

quadrature rule by adding up certain differences of tensor products of the one-dimensional quadrature rule. The quadrature rules obtained this way are generally not equal-weight quadrature rules.

Note that a straightforward tensor product of a one-dimensional quadrature rule is bound to fail, because if the one-dimensional quadrature rule uses n quadrature points, then the s-dimensional rule would use n^s quadrature points. In large dimensions, say $s = 100$, even a quadrature rule in one dimension which uses only 2 points would then use $2^{100} \approx 10^{30}$ points, making such a computation unfeasible.

Traditionally those two methods are viewed as competing, i.e. one either uses a quasi-Monte Carlo rule or a sparse grid. In this paper we consider a hybrid of those two methods and show that they can be combined in different ways. There are several interesting points to be made on such a combination. It allows us to compare those two methods in terms of their performance on integrals. The difference to other comparisons here is that in one dimension both methods start off with the same quadrature rule, but are generalized in different ways. Actually, one does not need to use a one-dimensional rule as starting point for the sparse grid, but can use a d-dimensional rule – in our case a d-dimensional quasi-Monte Carlo rule – and thereby obtain an $s = dl$ -dimensional quadrature rule. For example for $s = 4$ one can obtain various combinations of quasi-Monte Carlo and sparse grid (see the examples in the numerical section at the end of the paper).

Another point to be made is that this combination allows us to obtain randomized sparse grids. Randomization has been common practice for quasi-Monte Carlo rules, see for example [3, 11]. Using a randomized quasi-Monte Carlo rule as the underlying quadrature rule we obtain also randomized sparse grids. Randomization for quasi-Monte Carlo rules has some advantages, for example one obtains an unbiased estimator of the integral – a property which we show here also holds for our randomized sparse grids. Further one can also obtain an estimator of the integration error, which is also possible for sparse grids using a randomized quasi-Monte Carlo rule as underlying quadrature rule. A drawback is on the other hand that the upper bounds on the worst-case error only hold in a statistical sense. Here worst-case error means the supremum of the absolute value of the integration error over the unit ball of some Banach space (which will be made precise in Section 4). Such a setting has been considered frequently when analysing quadrature rules for numerical integration.

In this paper we also analyse the worst-case error for multivariate integration in certain Sobolev spaces. We show that the hybrid method achieves the optimal rate of convergence for every combination of an l-fold sparse grid based on a certain d-dimensional quasi-Monte Carlo rule (yielding an $s = dl$-dimensional quadrature rule). This result holds for deterministic and also randomized quadrature rules. For fixed s, the only possible difference in the error bounds lies in a possible difference in the unknown constant, which shows that their performance is of the same order.

The paper is organized as follows. In the following section we introduce digital sequences which are the building blocks for our algorithms. In Section 3 we introduce the deterministic and the randomized algorithm and in Section 4 we analyze the worst-case error for those algorithms. We conclude the paper with some numerical results in Section 5.

2 Digital $(t, \alpha, \beta, \sigma, d)$-sequences

The following definition gives the digital construction scheme, which was first considered in [5] and is a slight variation on the ideas of Niederreiter [10].

Definition 1 (Digital net) Let b be a prime and let $n, m, d \geq 1$ be integers. Let $C_1, \ldots, C_d$ be $n \times m$ matrices over the finite field $\mathbb{Z}_b = \{0, 1, \ldots, b-1\}$ of order b. Now we construct b^m points in $[0,1)^d$: for $0 \leq h \leq b^m - 1$ let $h = h_0 + h_1 b + \cdots + h_{m-1} b^{m-1}$ be the b-adic expansion of h. Identify h with the vector $\vec{h} = (h_0, \ldots, h_{m-1})^\top \in \mathbb{Z}_b^m$, where $\top$ means the transpose of the vector. For $1 \leq j \leq d$ multiply the matrix C_j by $\vec{h}$, i.e.,

$$C_j \vec{h} =: (y_{j,1}(h), \ldots, y_{j,n}(h))^\top \in \mathbb{Z}_b^n,$$

and set

$$x_{h,j} := \frac{y_{j,1}(h)}{b} + \cdots + \frac{y_{j,n}(h)}{b^n} \quad \text{and} \quad \boldsymbol{x}_h := (x_{h,1}, \ldots, x_{h,d}).$$

The point set $\{\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{b^m-1}\}$ is called a digital net (over $\mathbb{Z}_b$) (with generating matrices $C_1, \ldots, C_d$).

For $n, m = \infty$ we obtain a sequence $\{\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots\}$, which is called a digital sequence (over $\mathbb{Z}_b$) (with generating matrices $C_1, \ldots, C_d$).

In the following we define special digital nets and digital sequences, which were first introduced in [5].

Definition 2 Let $n, m, \alpha \geq 1$ be natural numbers, let $0 < \beta \leq \alpha m/n$ be a real number and let $0 \leq t \leq \beta n$ be a natural number. Let $\mathbb{Z}_b$ be the finite field of prime order b and let $C_1, \ldots, C_d \in \mathbb{Z}_b^{n \times m}$ with $C_j = (c_{j,1}, \ldots, c_{j,n})^\top$. If for all $1 \leq i_{j,\nu_j} < \cdots < i_{j,1} \leq m$, where $0 \leq \nu_j \leq n$ for all $j = 1, \ldots, d$, with

$$i_{1,1} + \cdots + i_{1,\min(\nu_1,\alpha)} + \cdots + i_{d,1} + \cdots + i_{d,\min(\nu_d,\alpha)} \leq \beta n - t$$

the vectors

$$c_{1,i_{1,\nu_1}}, \ldots, c_{1,i_{1,1}}, \ldots, c_{d,i_{d,\nu_d}}, \ldots, c_{d,i_{d,1}}$$

are linearly independent over $\mathbb{Z}_b$, then the digital net with generating matrices $C_1, \ldots, C_d$ is called a digital $(t, \alpha, \beta, n \times m, d)$-net over $\mathbb{Z}_b$. Further we call a digital $(t, \alpha, \beta, n \times m, d)$-net over $\mathbb{Z}_b$ with the largest possible value of β, i.e., $\beta = \alpha m/n$, a digital $(t, \alpha, n \times m, d)$-net over $\mathbb{Z}_b$.

We can also define sequences of points for which the first b^m points form a digital $(t, \alpha, \beta, \sigma, n \times m, d)$-net, see [5].

Definition 3 Let $\alpha, \sigma \geq 1$ and $t \geq 0$ be integers and let $0 < \beta \leq \alpha/\sigma$ be a real number. Let $\mathbb{Z}_b$ be the finite field of prime order b and let $C_1, \ldots, C_d \in \mathbb{Z}_b^{\infty \times \infty}$ with $C_j = (c_{j,1}, c_{j,2}, \ldots)^\top$. Further let $C_{j,\sigma m \times m}$ denote the left upper $\sigma m \times m$ submatrix of C_j. If for all $m > t/(\beta\sigma)$ the matrices $C_{1,\sigma m \times m}, \ldots, C_{d,\sigma m \times m}$ generate a digital $(t, \alpha, \beta, \sigma m \times m, d)$-net then the digital sequence with generating matrices $C_1, \ldots, C_d$ is called a digital $(t, \alpha, \beta, \sigma, d)$-sequence over $\mathbb{Z}_b$. Further we call a digital $(t, \alpha, 1, \alpha, d)$-sequence over $\mathbb{Z}_b$ a digital (t, α, d)-sequence over $\mathbb{Z}_b$.

Explicit constructions of such sequences were also given in [5]. There it was shown that one can, for any given integer $\sigma \geq 1$, construct a sequence $S_\sigma = \{\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots\}$ such that this sequence is a digital $(t, \alpha, \min(1, \alpha/\sigma), \sigma, d)$-sequence for all $\alpha \geq 1$. Note that the value of σ has to be chosen in advance and cannot be changed for a given sequence S_σ. The sequences S_σ can now be used to approximate the integral $\int_{[0,1]^d} f(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}$ of a function $f : [0,1)^d \to \mathbb{R}$ by the sum

$$A_{\sigma,m}(f) = b^{-m} \sum_{n=0}^{b^m - 1} f(\boldsymbol{x}_n).$$

Dick [5] showed that this approximation of the integral is of order $\mathcal{O}(b^{-\min(\sigma,\delta)m} m^{d\delta})$ for functions whose mixed partial derivatives of order δ in each variable are square integrable. This result holds in the worst-case sense and for the randomized sequence $\boldsymbol{y}_0, \boldsymbol{y}_1, \ldots$ in the root mean square sense (see [5, Corollary 5.5] and the discussion thereafter).

The randomized digital sequence is obtained in the following way: let $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_d) \in [0,1)^d$ with $\nu_i = \nu_{i,1} b^{-1} + \nu_{i,2} b^{-2} + \cdots$ be given. Let $\boldsymbol{x}_n = (x_{n,1}, \ldots, x_{n,d})$ and $x_{n,i} = x_{n,i,1} b^{-1} + x_{n,i,2} b^{-2} + \cdots$. Then we obtain the digitally shifted point $\boldsymbol{y}_n = (y_{n,1}, \ldots, y_{n,d})$ with $y_{n,i} = y_{n,i,1} b^{-1} + y_{n,i,2} b^{-2} + \cdots$ by setting $y_{n,i,j} = x_{n,i,j} + \nu_{i,j} \pmod{b}$ for all $n \geq 0$ and $j \geq 1$ and $1 \leq i \leq d$. By choosing $\boldsymbol{\nu} \in [0,1)^d$ i.i.d. we obtain a randomized digital $(t, \alpha, \beta, \sigma, d)$-sequence. In this case we approximate the integral $\int_{[0,1]^d} f(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}$ by

$$A_{\sigma,m,\boldsymbol{\nu}}(f) = b^{-m} \sum_{n=0}^{b^m - 1} f(\boldsymbol{y}_n).$$

These sequences form the basic building block for the construction of our quadrature rules which we introduce in the following section.

3 Smolyak algorithms and randomized Smolyak algorithms based on digital $(t, \alpha, \min(1, \alpha/\sigma), \sigma, d)$-sequences

In this section we introduce l fold sparse grids using d dimensional digital sequences.

3.1 The deterministic algorithm

We now consider functions $f : [0,1]^s \to \mathbb{R}$, where $s = dl$. For $1 \le j \le l$ let $A_{\sigma,m,j}$ denote the algorithm $A_{\sigma,m}$ applied to the coordinates $(j-1)d+1, \dots, jd$. Then we define $\Delta_{\sigma,0,j} = A_{\sigma,0,j}$ and for $m \ge 1$ we set $\Delta_{\sigma,m,j} = A_{\sigma,m,j} - A_{\sigma,m-1,j}$ for all $j = 1, \dots, l$ and $\sigma \ge 1$. The algorithm is now given by

$$\mathcal{A}_{\sigma,q,s} = \sum_{\substack{m_1,\dots,m_l \in \mathbb{N}_0 \\ m_1+\dots+m_l \le q}} \bigotimes_{j=1}^{l} \Delta_{\sigma,m_j,j}.$$

Here and throughout this paper let $\mathbb{N}_0$ denote the set of non-negative integers.

3.2 The randomized algorithm

We consider again functions $f : [0,1]^s \to \mathbb{R}$, where $s = dl$. Now for $1 \le j \le l$ let $A_{\sigma,m,\boldsymbol{\nu}_j,j}$ denote the algorithm $A_{\sigma,m,\boldsymbol{\nu}_j}$ applied to the coordinates $(j-1)d+1, \dots, jd$. Then we define $\Delta_{\sigma,0,\boldsymbol{\nu}_j,j} = A_{\sigma,0,\boldsymbol{\nu}_j,j}$ and for $m \ge 1$ we set $\Delta_{\sigma,m,\boldsymbol{\nu}_j,j} = A_{\sigma,m,\boldsymbol{\nu}_j,j} - A_{\sigma,m-1,\boldsymbol{\nu}_j,j}$ for all $j = 1, \dots, s$ and $\sigma \ge 1$. The algorithm is now given by

$$\mathcal{A}_{\sigma,q,\boldsymbol{\nu},s} = \sum_{\substack{m_1,\dots,m_l \in \mathbb{N}_0 \\ m_1+\dots+m_l \le q}} \bigotimes_{j=1}^{l} \Delta_{\sigma,m_j,\boldsymbol{\nu}_j,j}, \tag{3.1}$$

where $\boldsymbol{\nu} = (\boldsymbol{\nu}_1, \dots, \boldsymbol{\nu}_l) \in [0,1)^{dl}$. By choosing $\boldsymbol{\nu} \in [0,1)^{dl}$ uniformly i.i.d. we obtain a randomized Smolyak type algorithm.

4 Error analysis

For a given function f the error of approximating the integral $I_s(f) = \int_{[0,1]^s} f(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x}$, $s = dl$, using $\mathcal{A}_{\sigma,q,\boldsymbol{\nu},s}$ is given by

$$e_{\sigma,q,\boldsymbol{\nu},d,l}(f) = I_s(f) - \mathcal{A}_{\sigma,q,\boldsymbol{\nu},s}(f).$$

Specifically we consider functions $f : [0,1]^s \to \mathbb{R}$ from the Sobolev space $\mathcal{H}_{s,\delta}$, where $s \ge 1$ denotes the dimension and $\delta \ge 1$ refers to the smoothness requirements for functions in $\mathcal{H}_{s,\delta}$, i.e., the functions are required to have partial mixed derivatives of order up to δ in each variable which are square integrable. For the one-dimensional space the inner product in $\mathcal{H}_{s,\delta}$ is given by

$$\langle f, g \rangle_{\mathcal{H}_{1,\delta}} = \sum_{\tau=0}^{\delta-1} \int_0^1 f^{(\tau)}(x) \,\mathrm{d}x \int_0^1 g^{(\tau)}(x) \,\mathrm{d}x + \int_0^1 f^{(\delta)}(x) g^{(\delta)}(x) \,\mathrm{d}x,$$

where $f^{(\tau)}$ denotes the τ-th derivative of f and where $f^{(0)} = f$. The reproducing kernel (see [1] for more information about reproducing kernels) for this space is given by

$$\mathcal{K}_{1,\delta}(x,y) = \sum_{\tau=0}^{\delta} \frac{B_\tau(x)B_\tau(y)}{(\tau!)^2} + (-1)^{\delta+1}\frac{B_{2\delta}(|x-y|)}{(2\delta)!},$$

where B_τ denotes the Bernoulli polynomial of degree τ. For example we have $B_0(x) = 1$, $B_1(x) = x - 1/2$, $B_2(x) = x^2 - x + 1/6$ and so on.

The reproducing kernel for the s-dimensional Sobolev space $\mathcal{H}_{s,\delta}$ is now given by

$$\mathcal{K}_{s,\delta}(\boldsymbol{x},\boldsymbol{y}) = \prod_{j=1}^{s}\left(\sum_{\tau=0}^{\delta} \frac{B_\tau(x_j)B_\tau(y_j)}{(\tau!)^2} + (-1)^{\delta+1}\frac{B_{2\delta}(|x_j-y_j|)}{(2\delta)!}\right).$$

In the following we consider functions $f \in \mathcal{H}_{s,\delta}$. Note that in this case we have (see [1])

$$f(\boldsymbol{y}) = \langle f, \mathcal{K}_{s,\delta}(\cdot,\boldsymbol{y})\rangle_{\mathcal{H}_{s,\delta}}. \tag{4.1}$$

We are now ready to analyze the error for functions in the space $\mathcal{H}_{s,\delta}$.

4.1 Worst-case error analysis

Using the reproducing property (4.1) together with the linearity of the inner product it follows that

$$\begin{aligned}\bigotimes_{j=1}^{l} \Delta_{\sigma,m_j,\boldsymbol{\nu}_j,j}(f) &= \left\langle f, \bigotimes_{j=1}^{l} \Delta_{\sigma,m_j,\boldsymbol{\nu}_j,j}(\mathcal{K}_{s,\delta}(\cdot,\boldsymbol{y}))\right\rangle_{\mathcal{H}_{s,\delta}} \\ &= \left\langle f, \prod_{j=1}^{l} \Delta_{\sigma,m_j,\boldsymbol{\nu}_j,j}(\mathcal{K}_{d,\delta}(\cdot,(y_{(j-1)d+1},\ldots,y_{jd})))\right\rangle_{\mathcal{H}_{s,\delta}},\end{aligned}$$

where the inner product is with respect to the first variable of $\mathcal{K}_{s,\delta}$ and the operator $\Delta_{\sigma,m_j,\boldsymbol{\nu}_j,j}$ is applied to the second variable of $\mathcal{K}_{d,\delta}$.

Note that for any $\boldsymbol{x} \in [0,1]^s$ we have

$$\int_{[0,1]^s} \mathcal{K}_{s,\delta}(\boldsymbol{x},\boldsymbol{y})\,\mathrm{d}\boldsymbol{y} = 1.$$

Hence we have

$$e_{\sigma,q,\boldsymbol{\nu},s}(f) = \left\langle f, \int_{[0,1]^s} \mathcal{K}_{s,\delta}(\cdot,\boldsymbol{y})\,\mathrm{d}\boldsymbol{y} - \sum_{\substack{m_1,\ldots,m_l\in\mathbb{N}_0\\ m_1+\cdots+m_l\le q}} \prod_{j=1}^{l} \Delta_{\sigma,m_j,\boldsymbol{\nu}_j,j}(\mathcal{K}_{d,\delta}(\cdot,(y_{(j-1)d+1},\ldots,y_{jd})))\right\rangle_{\mathcal{H}_{s,\delta}}$$

$$= \left\langle f, 1 - \sum_{\substack{m_1,\ldots,m_l \in \mathbb{N}_0 \\ m_1+\cdots+m_l \le q}} \prod_{j=1}^{l} \Delta_{\sigma,m_j,\boldsymbol{\nu}_j,j}(\mathcal{K}_{d,\delta}(\cdot,(y_{(j-1)d+1},\ldots,y_{jd}))) \right\rangle_{\mathcal{H}_{s,\delta}} . \tag{4.2}$$

Consider now the unit ball $\mathcal{B}_{s,\delta} = \{f \in \mathcal{H}_{s,\delta} : \|f\|_{\mathcal{H}_{s,\delta}} \le 1\}$ in the space $\mathcal{H}_{s,\delta}$. Then it follows from (4.2) that the function $f \in \mathcal{B}_{s,\delta}$ for which $|e_{\sigma,q,\boldsymbol{\nu},s}(f)|$ is largest is given by the normalization of the function

$$f_{\mathrm{wce}}(\boldsymbol{x}) = 1 - \sum_{\substack{m_1,\ldots,m_l \in \mathbb{N}_0 \\ m_1+\cdots+m_l \le q}} \prod_{j=1}^{l} \Delta_{\sigma,m_j,\boldsymbol{\nu}_j,j}(\mathcal{K}_{d,\delta}((x_{(j-1)d+1},\ldots,x_{jd}),(y_{(j-1)d+1},\ldots,y_{jd}))),$$

where $\boldsymbol{x} = (x_1,\ldots,x_s)$, i.e., by the function $f_{\mathrm{wce}}\|f_{\mathrm{wce}}\|^{-1}_{\mathcal{H}_{s,\delta}}$.

For $\boldsymbol{y} = (y_1,\ldots,y_s)$ we write $\boldsymbol{y}_{d,j} = (y_{(j-1)d+1},\ldots,y_{jd})$ and similarly for $\boldsymbol{z}$.

Let the worst-case error in the space $\mathcal{H}_{s,\delta}$ now be given by

$$\mathrm{wce}(\mathcal{H}_{s,\delta}, \mathcal{A}_{\sigma,q,\boldsymbol{\nu},s}) = \sup_{f \in \mathcal{B}_{s,\delta}} |e_{\sigma,q,\boldsymbol{\nu},s}(f)|.$$

Then the above results imply that

$$\mathrm{wce}(\mathcal{H}_{s,\delta}, \mathcal{A}_{\sigma,q,\boldsymbol{\nu},s}) = \|f_{\mathrm{wce}}\|_{\mathcal{H}_{s,\delta}}$$

and hence

$$\begin{aligned} \mathrm{wce}^2(\mathcal{H}_{s,\delta}, \mathcal{A}_{\sigma,q,\boldsymbol{\nu},s}) &= \langle f_{\mathrm{wce}}, f_{\mathrm{wce}} \rangle_{\mathcal{H}_{s,\delta}} \\ &= -1 + \sum_{\substack{\boldsymbol{m} \in \mathbb{N}_0^l \\ \|\boldsymbol{m}\|_1 \le q}} \sum_{\substack{\boldsymbol{k} \in \mathbb{N}_0^l \\ \|\boldsymbol{k}\|_1 \le q}} \prod_{j=1}^{l} \Delta_{\sigma,m_j,\boldsymbol{\nu}_j,j} \Delta_{\sigma,k_j,\boldsymbol{\nu}_j,j} \left(\langle \mathcal{K}_{d,\delta}(\cdot,\boldsymbol{y}_{d,j}), \mathcal{K}_{d,\delta}(\cdot,\boldsymbol{z}_{d,j}) \rangle_{\mathcal{H}_{d,\delta}} \right) \\ &= -1 + \sum_{\substack{\boldsymbol{m} \in \mathbb{N}_0^l \\ \|\boldsymbol{m}\|_1 \le q}} \sum_{\substack{\boldsymbol{k} \in \mathbb{N}_0^l \\ \|\boldsymbol{k}\|_1 \le q}} \prod_{j=1}^{l} \Delta_{\sigma,m_j,\boldsymbol{\nu}_j,j} \Delta_{\sigma,k_j,\boldsymbol{\nu}_j,j} \left(\mathcal{K}_{d,\delta}(\boldsymbol{z}_{d,j}, \boldsymbol{y}_{d,j}) \right), \end{aligned} \tag{4.3}$$

where the operator $\Delta_{\sigma,m_j,\boldsymbol{\nu}_j,j}$ is applied to $\boldsymbol{y}_{d,j}$ and $\Delta_{\sigma,k_j,\boldsymbol{\nu}_j,j}$ is applied to $\boldsymbol{z}_{d,j}$ and where for $\boldsymbol{m} = (m_1,\ldots,m_l) \in \mathbb{N}_0^s$ we write $\|\boldsymbol{m}\|_1 = m_1 + \cdots + m_l$ and analogously for $\boldsymbol{k}$. Furthermore, here and in the sequel for two operators T and S we will write $ST(\cdot)$ instead of $S(T(\cdot))$.

In the following we show that

$$\sum_{m \in \mathbb{N}_0} \sum_{k \in \mathbb{N}_0} \Delta_{\sigma,m,\boldsymbol{\nu}} \Delta_{\sigma,k,\boldsymbol{\nu}} \left(\mathcal{K}_{d,\delta}(\boldsymbol{z},\boldsymbol{y}) \right) \tag{4.4}$$

with $\boldsymbol{y},\boldsymbol{z} \in [0,1)^d$, converges absolutely. To this end let $m,k \ge -1$ and let

$$Z_{m,k} = (I_d^{(\boldsymbol{y})} - A_{\sigma,m,\boldsymbol{\nu}}^{(\boldsymbol{y})})(I_d^{(\boldsymbol{z})} - A_{\sigma,k,\boldsymbol{\nu}}^{(\boldsymbol{z})}) \left(\mathcal{K}_{d,\delta}(\boldsymbol{z},\boldsymbol{y}) \right),$$

where $A^{(\boldsymbol{y})}_{\sigma,-1,\boldsymbol{\nu}}(f) = 0$ and where $I^{(\boldsymbol{y})}_d$ (resp. $A^{(\boldsymbol{y})}_{\sigma,m,\boldsymbol{\nu}})$) is the integral operator (resp. the randomized algorithm) with respect to $\boldsymbol{y}$ and similarly for $I^{(\boldsymbol{z})}_d$ and $A^{(\boldsymbol{z})}_{\sigma,k,\boldsymbol{\nu}}$. Note that $\Delta_{\sigma,m,\boldsymbol{\nu}}\Delta_{\sigma,k,\boldsymbol{\nu}}\left(\mathcal{K}_{d,\delta}(\boldsymbol{z},\boldsymbol{y})\right) = Z_{m,k} - Z_{m-1,k} - Z_{m,k-1} + Z_{m-1,k-1}$. Further observe that $Z_{m,k}$ also depends on δ and σ, but as those parameters are fixed, i.e. σ is fixed from the construction and δ is the smoothness of the function, we do not make this dependence explicit. We need some lemmas, but first we recall the definition of Walsh functions which will be used in the following.

Definition 4 (Walsh functions) *Let $b \geq 2$ be an integer. For a non-negative integer k with base b representation*

$$k = \kappa_0 + \kappa_1 b + \cdots + \kappa_a b^a,$$

with $\kappa_i \in \{0,\ldots,b-1\}$, we define the Walsh function ${}_b\mathrm{wal}_k : [0,1) \to \mathbb{C}$ by

$$_b\mathrm{wal}_k(x) := \mathrm{e}^{2\pi\mathtt{i}(x_1\kappa_0+\cdots+x_{a+1}\kappa_a)/b},$$

for $x \in [0,1)$ with base b representation $x = \frac{x_1}{b} + \frac{x_2}{b^2} + \cdots$ (unique in the sense that infinitely many of the x_i must be different from $b-1$). If it is clear which base b is chosen we will simply write wal_k.

Definition 5 (Multivariate Walsh functions) *Let $b \geq 2$ be an integer. For dimension $s \geq 2$, $x_1,\ldots,x_s \in [0,1)$ and $k_1,\ldots,k_s \in \mathbb{N}_0$ we define ${}_b\mathrm{wal}_{k_1,\ldots,k_s} : [0,1)^s \to \mathbb{C}$ by*

$$_b\mathrm{wal}_{k_1,\ldots,k_s}(x_1,\ldots,x_s) := \prod_{j=1}^{s} {}_b\mathrm{wal}_{k_j}(x_j).$$

For vectors $\boldsymbol{k} = (k_1,\ldots,k_s) \in \mathbb{N}_0^s$ and $\boldsymbol{x} = (x_1,\ldots,x_s) \in [0,1)^s$ we write

$$_b\mathrm{wal}_{\boldsymbol{k}}(\boldsymbol{x}) := {}_b\mathrm{wal}_{k_1,\ldots,k_s}(x_1,\ldots,x_s).$$

Again, if it is clear which base we mean we simply write $\mathrm{wal}_{\boldsymbol{k}}(\boldsymbol{x})$.

It is clear from the definitions that Walsh functions are piecewise constant. It can be shown that for any integers $s \geq 1$ and $b \geq 2$ the system $\{{}_b\mathrm{wal}_{k_1,\ldots,k_s} : k_1,\ldots,k_s \geq 0\}$ is a complete orthonormal system in $L_2([0,1)^s)$, see, for example, [2, 9]. More information on Walsh functions can be found, for example, in [2, 5, 7, 15].

We note that if Walsh functions, digital shifts or digital sequences are used in conjunction with each other, they are always in the same base b. Therefore we will often omit the b.

For the next lemma we introduce some notation: for a vector $\boldsymbol{u} \in \mathbb{N}_0^d$ and $\emptyset \neq X \subseteq \{1,\ldots,d\}$ we write $(\boldsymbol{u}_X, 0)$ for the vector $\boldsymbol{u}$ with all components whose index is not in X replaced by 0. For an $\infty \times \infty$ matrix C we write $C(m)$ for the $\infty \times m$ matrix which consists of the first m columns of the matrix C and where for $X = \{x_1,\ldots,x_e\} \subseteq \{1,\ldots,d\}$, $e = |X|$,

$$\mathcal{D}^*_{m,X} = \{\boldsymbol{l} \in \mathbb{N}^{|X|} : C_{x_1}(m)^\top \vec{l}_1 + \cdots + C_{x_e}(m)^\top \vec{l}_e = \vec{0}\}$$

with the vector $\boldsymbol{l} = (l_1, \ldots, l_e)$ and where each coordinate has base b representation $l_j = l_{j,0} + l_{j,1} b + \cdots$ and where $\vec{l_j} = (l_{j,0}, l_{j,1}, \ldots)^\top$ for $j = 1, \ldots, d$. Further

$$\widehat{K}(\boldsymbol{u}, \boldsymbol{w}) = \int_{[0,1]^d} \int_{[0,1]^d} \mathcal{K}_{d,\delta}(\boldsymbol{x}, \boldsymbol{y}) \overline{\mathrm{wal}_{\boldsymbol{u}}(\boldsymbol{x})} \mathrm{wal}_{\boldsymbol{w}}(\boldsymbol{y}) \,\mathrm{d}\boldsymbol{x} \,\mathrm{d}\boldsymbol{y}.$$

Lemma 1 *With the definition above we have $Z_{-1,-1} = 1$, $Z_{m,-1} = Z_{-1,m} = 0$ for $m \geq 0$ and for $m, k \geq 0$ we have*

$$|Z_{m,k}| \leq \sum_{\emptyset \neq X \subseteq \{1,\ldots,d\}} \sum_{\substack{\boldsymbol{u}, \boldsymbol{w} \in \mathbb{N}^{|X|} \\ \boldsymbol{u} \in \mathcal{D}^*_{m,X}, \boldsymbol{w} \in \mathcal{D}^*_{k,X}}} \left| \widehat{K}((\boldsymbol{u}_X, 0), (\boldsymbol{w}_X, 0)) \right|.$$

Proof. We can write $Z_{m,k}$ as

$$I_d^{(\boldsymbol{y})} I_d^{(\boldsymbol{z})} (\mathcal{K}_{d,\delta}(\boldsymbol{z}, \boldsymbol{y})) - I_d^{(\boldsymbol{y})} A^{(\boldsymbol{z})}_{\sigma,k,\boldsymbol{\nu}} (\mathcal{K}_{d,\delta}(\boldsymbol{z}, \boldsymbol{y})) - A^{(\boldsymbol{y})}_{\sigma,m,\boldsymbol{\nu}} I_d^{(\boldsymbol{z})} (\mathcal{K}_{d,\delta}(\boldsymbol{z}, \boldsymbol{y})) + A^{(\boldsymbol{y})}_{\sigma,m,\boldsymbol{\nu}} A^{(\boldsymbol{z})}_{\sigma,k,\boldsymbol{\nu}} (\mathcal{K}_{d,\delta}(\boldsymbol{z}, \boldsymbol{y})).$$

The reproducing kernel $\mathcal{K}_{d,\delta}$ can be written as a Walsh series, i.e., we have

$$\mathcal{K}_{d,\delta}(\boldsymbol{z}, \boldsymbol{y}) = \sum_{\boldsymbol{u}, \boldsymbol{w} \in \mathbb{N}_0^d} \widehat{K}(\boldsymbol{u}, \boldsymbol{w}) \mathrm{wal}_{\boldsymbol{u}}(\boldsymbol{z}) \overline{\mathrm{wal}_{\boldsymbol{w}}(\boldsymbol{y})},$$

where $\widehat{K}(\boldsymbol{u}, \boldsymbol{w})$ is as above.

We consider several cases now. First let $m = k = -1$. Then we have

$$Z_{-1,-1} = \int_{[0,1]^d} \int_{[0,1]^d} \mathcal{K}_{d,\delta}(\boldsymbol{z}, \boldsymbol{y}) \,\mathrm{d}\boldsymbol{y} \,\mathrm{d}\boldsymbol{z} = 1.$$

Further for $m \geq 0$ we have

$$Z_{m,-1} = (I_d^{(\boldsymbol{y})} - A^{(\boldsymbol{y})}_{\sigma,m,\boldsymbol{\nu}}) I_d^{(\boldsymbol{z})} (\mathcal{K}_{d,\delta}(\boldsymbol{z}, \boldsymbol{y})) = (I_d^{(\boldsymbol{y})} - A^{(\boldsymbol{y})}_{\sigma,m,\boldsymbol{\nu}})(1) = 0$$

and the same applies for $m = -1$ and $k \geq 0$, i.e., $Z_{-1,k} = 0$. Now note that $\widehat{K}(\boldsymbol{0}, \boldsymbol{0}) = 1$ and $\widehat{K}(\boldsymbol{u}, \boldsymbol{v}) = 0$ if there exists an index $1 \leq i \leq d$ such that $u_i = 0$ and $v_i \neq 0$ or contrariwise. Using Lemma 2 below, we can therefore write, for $m, k \geq 0$,

$$\begin{aligned}
|Z_{m,k}| &= \left| A^{(\boldsymbol{y})}_{\sigma,m,\boldsymbol{\nu}} A^{(\boldsymbol{z})}_{\sigma,k,\boldsymbol{\nu}} \left(\sum_{\emptyset \neq X \subseteq \{1,\ldots,d\}} \sum_{\boldsymbol{u}, \boldsymbol{w} \in \mathbb{N}^{|X|}} \widehat{K}((\boldsymbol{u}_X, 0), (\boldsymbol{w}_X, 0)) \mathrm{wal}_{(\boldsymbol{u}_X, 0)}(\boldsymbol{z}) \overline{\mathrm{wal}_{(\boldsymbol{w}_X, 0)}(\boldsymbol{y})} \right) \right| \\
&\leq \sum_{\emptyset \neq X \subseteq \{1,\ldots,d\}} \sum_{\boldsymbol{u}, \boldsymbol{w} \in \mathbb{N}^{|X|}} \left| \widehat{K}((\boldsymbol{u}_X, 0), (\boldsymbol{w}_X, 0)) A^{(\boldsymbol{y})}_{\sigma,m,\boldsymbol{\nu}} A^{(\boldsymbol{z})}_{\sigma,k,\boldsymbol{\nu}} \left(\mathrm{wal}_{(\boldsymbol{u}_X, 0)}(\boldsymbol{z}) \overline{\mathrm{wal}_{(\boldsymbol{w}_X, 0)}(\boldsymbol{y})} \right) \right| \\
&= \sum_{\emptyset \neq X \subseteq \{1,\ldots,d\}} \sum_{\substack{\boldsymbol{u}, \boldsymbol{w} \in \mathbb{N}^{|X|} \\ \boldsymbol{u} \in \mathcal{D}^*_{m,X}, \boldsymbol{w} \in \mathcal{D}^*_{k,X}}} \left| \widehat{K}((\boldsymbol{u}_X, 0), (\boldsymbol{w}_X, 0)) \right|.
\end{aligned}$$

Hence the result follows. $\square$

We need the following lemma.

Lemma 2 *Let $\{\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots\}$ be a digital sequence over $\mathbb{Z}_b$ generated by the $\infty \times \infty$ matrices $C_1, \ldots, C_d$ over $\mathbb{Z}_b$. Then for any $m \in \mathbb{N}$ and for any vector $\boldsymbol{k} = (k_1, \ldots, k_d)$ of non-negative integers we have*

$$\sum_{h=0}^{b^m-1} \mathrm{wal}_{\boldsymbol{k}}(\boldsymbol{x}_h) = \begin{cases} b^m & \text{if } \boldsymbol{k} \in \mathcal{D}_m^* \cup \{\boldsymbol{0}\}, \\ 0 & \text{otherwise.} \end{cases}$$

Here $\mathcal{D}_m^* := \mathcal{D}_{m,\{1,\ldots,d\}}^*$.

Proof. For short we write $\omega_b := \mathrm{e}^{2\pi \mathtt{i}/b}$. We have

$$\sum_{h=0}^{b^m-1} \mathrm{wal}_{\boldsymbol{k}}(\boldsymbol{x}_h) = \sum_{h=0}^{b^m-1} \omega_b^{\langle C_1(m)\vec{h}, \vec{k_1}\rangle + \cdots + \langle C_d(m)\vec{h}, \vec{k_d}\rangle} = \sum_{h=0}^{b^m-1} \omega_b^{\langle \vec{h}, C_1(m)^\top \vec{k_1} + \cdots + C_d(m)^\top \vec{k_d}\rangle}. \tag{4.5}$$

Let $C_1(m)^\top \vec{k_1} + \cdots + C_d(m)^\top \vec{k_d} =: (z_1, \ldots, z_m)^\top \in \mathbb{Z}_b^m$. Then we have

$$\sum_{h=0}^{b^m-1} \mathrm{wal}_{\boldsymbol{k}}(\boldsymbol{x}_h) = \sum_{h_0,\ldots,h_{m-1}=0}^{b-1} \omega_b^{h_0 z_1 + \cdots + h_{m-1} z_m} = \prod_{i=1}^{m} \sum_{h=0}^{b-1} \omega_b^{h z_i}. \tag{4.6}$$

For $z \in \mathbb{Z}_b$ (note that b is a prime) we have

$$\sum_{h=0}^{b-1} \omega_b^{hz} = \begin{cases} b & \text{if } z = 0, \\ 0 & \text{if } z \neq 0, \end{cases}$$

and the result follows. $\square$

The absolute convergence of (4.4) immediately follows from the following lemma.

Lemma 3 *Let $m, k \geq 0$, $\sigma \geq 1$ and $\delta \geq 2$. Then there is a constant $C_{d,b,\delta} > 0$ such that*

$$|Z_{m,k}| \leq C_{d,b,\delta} (\min(\delta,\sigma) m - t + \delta)^{d\delta} (\min(\delta,\sigma) k - t + \delta)^{d\delta} b^{-\min(\delta,\sigma)(k+m)+2t}.$$

Proof. From Lemma 1 together with [5, Eq. (6.3)] and a slight modification of [5, Lemma 5.2] we obtain that there is some constant $C_{d,b,\delta} > 0$ such that

$$\begin{aligned} |Z_{m,k}| &\leq \sum_{\emptyset \neq X \subseteq \{1,\ldots,d\}} \sum_{\substack{\boldsymbol{u},\boldsymbol{w} \in \mathbb{N}^{|X|} \\ \boldsymbol{u} \in \mathcal{D}_{m,X}^*, \boldsymbol{w} \in \mathcal{D}_{k,X}^*}} \left| \widehat{K}((\boldsymbol{u}_X, 0), (\boldsymbol{w}_X, 0)) \right| \\ &\leq \sum_{\emptyset \neq X \subseteq \{1,\ldots,d\}} C'_{|X|,b,\delta} (\min(\delta,\sigma) m - t + \delta)^{|X|\delta} (\min(\delta,\sigma) k - t + \delta)^{|X|\delta} b^{-\min(\delta,\sigma)(k+m)+2t} \\ &\leq C_{d,b,\delta} b^{-\min(\delta,\sigma)(k+m)+2t} (\min(\delta,\sigma) m - t + \delta)^{d\delta} (\min(\delta,\sigma) k - t + \delta)^{d\delta}, \end{aligned}$$

where $C'_{|X|,b,\delta} > 0$ is a constant appearing in [5]. Thus the result follows. $\square$

Now we can write

$$1=\int_{[0,1]^s}\int_{[0,1]^s}\mathcal{K}_{s,\delta}(\boldsymbol{x},\boldsymbol{y})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{y}=\sum_{\boldsymbol{m}\in\mathbb{N}_0^l}\sum_{\boldsymbol{k}\in\mathbb{N}_0^l}\bigotimes_{j=1}^{l}\Delta_{\sigma,m_j,\boldsymbol{\nu}_j,j}\Delta_{\sigma,k_j,\boldsymbol{\nu}_j,j}\left(\mathcal{K}_{s,\delta}(\boldsymbol{z},\boldsymbol{y})\right) \tag{4.7}$$

and further for any $\boldsymbol{y}\in[0,1]^s$ we have

$$1=\int_{[0,1]^s}\mathcal{K}_{s,\delta}(\boldsymbol{z},\boldsymbol{y})\,\mathrm{d}\boldsymbol{z}=\sum_{\boldsymbol{m}\in\mathbb{N}_0^l}\bigotimes_{j=1}^{l}\Delta_{\sigma,m_j,\boldsymbol{\nu}_j,j}\left(\mathcal{K}_{s,\delta}(\boldsymbol{z},\boldsymbol{y})\right).$$

Thus we also have

$$1=\sum_{\boldsymbol{m}\in\mathbb{N}_0^l}\sum_{\substack{\boldsymbol{k}\in\mathbb{N}_0^l\\ \|\boldsymbol{k}\|_1\le q}}\bigotimes_{j=1}^{l}\Delta_{\sigma,m_j,\boldsymbol{\nu}_j,j}\Delta_{\sigma,k_j,\boldsymbol{\nu}_j,j}\left(\mathcal{K}_{s,\delta}(\boldsymbol{z},\boldsymbol{y})\right) \tag{4.8}$$

and

$$1=\sum_{\substack{\boldsymbol{m}\in\mathbb{N}_0^l\\ \|\boldsymbol{m}\|_1\le q}}\sum_{\boldsymbol{k}\in\mathbb{N}_0^l}\bigotimes_{j=1}^{l}\Delta_{\sigma,m_j,\boldsymbol{\nu}_j,j}\Delta_{\sigma,k_j,\boldsymbol{\nu}_j,j}\left(\mathcal{K}_{s,\delta}(\boldsymbol{z},\boldsymbol{y})\right). \tag{4.9}$$

Thus using (4.3), (4.7), (4.8) and (4.9) we obtain

$$\mathrm{wce}^2(\mathcal{H}_{s,\delta},\mathcal{A}_{\sigma,q,\boldsymbol{\nu},s})=\sum_{\substack{\boldsymbol{m}\in\mathbb{N}_0^l\\ \|\boldsymbol{m}\|_1>q}}\sum_{\substack{\boldsymbol{k}\in\mathbb{N}_0^l\\ \|\boldsymbol{k}\|_1>q}}\prod_{j=1}^{l}\Delta_{\sigma,m_j,\boldsymbol{\nu}_j,j}\Delta_{\sigma,k_j,\boldsymbol{\nu}_j,j}\left(\mathcal{K}_{d,\delta}(\boldsymbol{z}_{d,j},\boldsymbol{y}_{d,j})\right). \tag{4.10}$$

Using the formula for the worst-case error above we obtain

$$\mathrm{wce}^2(\mathcal{H}_{s,\delta},\mathcal{A}_{\sigma,q,\boldsymbol{\nu},s})\le\sum_{\substack{\boldsymbol{m}\in\mathbb{N}_0^l\\ \|\boldsymbol{m}\|_1>q}}\sum_{\substack{\boldsymbol{k}\in\mathbb{N}_0^l\\ \|\boldsymbol{k}\|_1>q}}\prod_{j=1}^{l}\left|\Delta_{\sigma,m_j,\boldsymbol{\nu}_j,j}\Delta_{\sigma,k_j,\boldsymbol{\nu}_j,j}\left(\mathcal{K}_{d,\delta}(\boldsymbol{z}_{d,j},\boldsymbol{y}_{d,j})\right)\right|.$$

Lemma 3 now implies that there exists a constant $c_{b,d,\sigma}>0$ such that

$$\begin{aligned}&\left|\Delta_{\sigma,m_j,\boldsymbol{\nu}_j,j}\Delta_{\sigma,k_j,\boldsymbol{\nu}_j,j}\left(\mathcal{K}_{d,\delta}(\boldsymbol{z}_{d,j},\boldsymbol{y}_{d,j})\right)\right|\\ &\quad\le\; c_{b,d,\sigma}^2(\min(\delta,\sigma)m_j-t+\delta)^{d\delta}(\min(\delta,\sigma)k_j-t+\delta)^{d\delta}b^{-\min(\delta,\sigma)(k_j+m_j)+2t}\end{aligned}$$

for $m_j,k_j\ge 0$. Thus we obtain

$$\mathrm{wce}(\mathcal{H}_{s,\delta},\mathcal{A}_{\sigma,q,\boldsymbol{\nu},s})\le b^{tl}c_{b,d,\sigma}^{l}\sum_{\substack{\boldsymbol{m}\in\mathbb{N}_0^l\\ \|\boldsymbol{m}\|_1>q}}b^{-\min(\delta,\sigma)\|\boldsymbol{m}\|_1}\prod_{j=1}^{l}(\delta m_j+\delta)^{d\delta}.$$

Using the arithmetic-geometric mean inequality we obtain that (note that $s = dl$),

$$\prod_{j=1}^{l}(m_j+1)^{d\delta} \le l^{-s\delta}\left(m_1+\cdots+m_l+l\right)^{s\delta} \le \left(\|\boldsymbol{m}\|_1+1\right)^{s\delta}.$$

Therefore we have

$$\begin{aligned}
\mathrm{wce}(\mathcal{H}_{s,\delta},\mathcal{A}_{\sigma,q,\boldsymbol{\nu},s}) &\le \delta^{s\delta} b^{tl} c_{b,d,\sigma}^{l} \sum_{m=q+1}^{\infty} b^{-\min(\delta,\sigma)m}\left(m+1\right)^{s\delta}\binom{m+s-1}{s-1} \\
&\le \delta^{s\delta} b^{tl+\min(\delta,\sigma)} c_{b,d,\sigma}^{l} \sum_{m=q+2}^{\infty} b^{-\min(\delta,\sigma)m} m^{s(\delta+1)-1}.
\end{aligned}$$

It can be shown that there is a constant $c'_{s,\delta} > 0$ such that

$$\sum_{m=q+2}^{\infty} b^{-\min(\delta,\sigma)m} m^{s(\delta+1)-1} \le c'_{s,\delta} q^{s(\delta+1)-1} b^{-\min(\delta,\sigma)q}$$

and hence we obtain the following theorem.

Theorem 1 *There is a constant $C = C(s,d,\sigma,\delta,t,b) > 0$ (which does not depend on q) such that the worst-case error for multivariate integration in the space $\mathcal{H}_{s,\delta}$, $\delta \ge 2$, using the algorithm $\mathcal{A}_{\sigma,q,\boldsymbol{\nu},s}$ is bounded by*

$$\mathrm{wce}(\mathcal{H}_{s,\delta},\mathcal{A}_{\sigma,q,\boldsymbol{\nu},s}) \le C \cdot q^{s(\delta+1)-1}\frac{1}{b^{\min(\delta,\sigma)q}}.$$

Let N be the number of points used by the algorithm $\mathcal{A}_{\sigma,q,\boldsymbol{\nu},s}$. Then we have

$$b^q \le N \le b^q\binom{q+l}{l}.$$

Here the left inequality follows as the algorithm requires b^q points already in dimension one and the right inequality follows as $\bigotimes_{j=1}^{l}\Delta_{\sigma,m_j,\boldsymbol{\nu}_j,j}$ requires $b^{m_1+\cdots+m_l}$ points and the sum in (3.1) has $\binom{q+l}{l}$ summands. Hence we have $\log_b N > q$ and

$$b^{-q} \le N^{-1}\binom{q+l}{l} \le N^{-1}(1+\log_b N)^l.$$

Thus we have the following result.

Corollary 1 *There is a constant $C = C(s,d,\sigma,\delta,t,b) > 0$ (which does not depend on q) such that the worst-case error for multivariate integration in the space $\mathcal{H}_{s,\delta}$, $\delta \ge 2$, using the algorithm $\mathcal{A}_{\sigma,q,\boldsymbol{\nu},s}$ is bounded by*

$$\mathrm{wce}(\mathcal{H}_{s,\delta},\mathcal{A}_{\sigma,q,\boldsymbol{\nu},s}) \le C \cdot (1+\log_b N)^{s(\delta+1)-1+l\min(\delta,\sigma)}\frac{1}{N^{\min(\delta,\sigma)}},$$

where N is the number of points used by the algorithm $\mathcal{A}_{\sigma,q,\boldsymbol{\nu},s}$.

4.2 Random case error analysis

Let now $\boldsymbol{\nu} = (\boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_l) \in [0,1)^{dl}$, $s = dl$, uniformly i.i.d. and $f : [0,1]^s \to \mathbb{R}$ be a Lebesgue integrable function. Then for all $m_j \geq 0$ and $\sigma \geq 1$ we have

$$\int_{[0,1]^s} \bigotimes_{j=1}^{l} A_{\sigma,m_j,\boldsymbol{\nu}_j,j}(f)\,\mathrm{d}\boldsymbol{\nu} = \int_{[0,1]^s} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}.$$

Thus it follows that

$$\int_{[0,1]^s} \bigotimes_{j=1}^{l} \Delta_{\sigma,m_j,\boldsymbol{\nu}_j,j}(f)\,\mathrm{d}\boldsymbol{\nu} = 0$$

if there is at least one $m_j > 0$ and further

$$\int_{[0,1]^s} \bigotimes_{j=1}^{l} \Delta_{\sigma,0,\boldsymbol{\nu}_j,j}(f)\,\mathrm{d}\boldsymbol{\nu} = \int_{[0,1]^s} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}.$$

Thus it follows that the expectation with respect to the random digital shift $\boldsymbol{\nu} \in [0,1)^s$ of our approximation is the exact integral, i.e.,

$$\mathbb{E}_{\boldsymbol{\nu}}\left(\mathcal{A}_{\sigma,q,\boldsymbol{\nu},s}(f)\right) = \int_{[0,1]^s} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x},$$

which means that our approximation is an unbiased estimator of the integral.

The advantage of the randomization is that one can now obtain an estimate of the integration error via Chebychev's inequality, see [13, Remark 2]. For a positive integer w and $\boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_w \in [0,1)^s$ uniformly i.i.d., let

$$\overline{\mathcal{A}}_{\sigma,q,\boldsymbol{\nu}_1,\ldots,\boldsymbol{\nu}_w,s}(f) := \frac{1}{w}\sum_{k=1}^{w} \mathcal{A}_{\sigma,q,\boldsymbol{\nu}_k,s}(f).$$

Then $\overline{\mathcal{A}}_{\sigma,q,\boldsymbol{\nu}_1,\ldots,\boldsymbol{\nu}_w,s}(f)$ is an unbiased estimator of the integral with variance

$$\mathcal{V} = w^{-1}\mathbb{E}_{\boldsymbol{\nu}}\left(\left(\mathcal{A}_{\sigma,q,\boldsymbol{\nu},s}(f) - I_s(f)\right)^2\right).$$

Chebyshev's inequality then yields

$$\mathbb{P}\left(\left|\overline{\mathcal{A}}_{\sigma,q,\boldsymbol{\nu}_1,\ldots,\boldsymbol{\nu}_w,s}(f) - I_s(f)\right| < k\sqrt{\mathcal{V}}\right) \geq 1 - k^{-2} \qquad \forall k > 0.$$

Note that the bound from the previous case can be used to obtain a bound on the root-mean-square worst-case error, i.e.,

$$\widehat{\mathrm{wce}}(\mathcal{H}_{s,\delta}, \mathcal{A}_{\sigma,q,s}) = \left(\int_{[0,1)^s} \mathrm{wce}^2(\mathcal{H}_{s,\delta}, \mathcal{A}_{\sigma,q,\boldsymbol{\nu},s})\,\mathrm{d}\boldsymbol{\nu}\right)^{1/2},$$

as the bound from the previous section holds for all shifts $\boldsymbol{\nu} \in [0,1)^s$.

On the other hand a better bound can be obtained by dealing directly with the mean-square worst-case error $\widehat{\mathrm{wce}}^2(\mathcal{H}_{s,\delta}, \mathcal{A}_{\sigma,q,s})$. Using (4.10) and the definition of the mean-square worst-case error we obtain that

$$\widehat{\mathrm{wce}}^2(\mathcal{H}_{s,\delta}, \mathcal{A}_{\sigma,q,s}) = \sum_{\substack{\boldsymbol{m}\in\mathbb{N}_0^l \\ \|\boldsymbol{m}\|_1 > q}} \sum_{\substack{\boldsymbol{k}\in\mathbb{N}_0^l \\ \|\boldsymbol{k}\|_1 > q}} \prod_{j=1}^{l} \int_{[0,1]^d} \Delta_{\sigma,m_j,\boldsymbol{\nu}_j,j} \Delta_{\sigma,k_j,\boldsymbol{\nu}_j,j} \left(\mathcal{K}_{d,\delta}(\boldsymbol{z}_{d,j}, \boldsymbol{y}_{d,j})\right) \, \mathrm{d}\boldsymbol{\nu}_j.$$

For $\boldsymbol{y}, \boldsymbol{z} \in [0,1)^d$ we define $\widehat{Z}_{m,k} = \int_{[0,1]^d} (I_d^{(\boldsymbol{y})} - A_{\sigma,m,\boldsymbol{\nu}}^{(\boldsymbol{y})})(I_d^{(\boldsymbol{z})} - A_{\sigma,k,\boldsymbol{\nu}}^{(\boldsymbol{z})}) \left(\mathcal{K}_{d,\delta}(\boldsymbol{z}, \boldsymbol{y})\right) \mathrm{d}\boldsymbol{\nu}$. We obtain that $\widehat{Z}_{-1,-1} = 1$, $\widehat{Z}_{m,-1} = \widehat{Z}_{-1,m} = 0$ for $m \ge 0$ and for $m, k \ge 0$ we have

$$\widehat{Z}_{m,k} = \sum_{\emptyset \neq X \subseteq \{1,\ldots,d\}} \sum_{\substack{\boldsymbol{u}\in\mathbb{N}^{|X|} \\ \boldsymbol{u}\in\mathcal{D}^*_{m,X} \cap \mathcal{D}^*_{k,X}}} \widehat{K}((\boldsymbol{u}_X, 0), (\boldsymbol{u}_X, 0)).$$

Using results from [4] for $\delta \ge 2$ and from [6] for $\delta = 1$ we obtain that there is a constant $C_2 > 0$ such that

$$|\widehat{Z}_{m,k}| \le C_2 (\min(\delta,\sigma)\max(m,k) + \delta)^{2d\delta} b^{-2\min(\delta,\sigma)\max(m,k) + 2t}.$$

Hence there is a constant $c > 0$ such that

$$\begin{aligned} \left| \int_{[0,1]^d} \Delta_{\sigma,m_j,\boldsymbol{\nu}_j,j} \Delta_{\sigma,k_j,\boldsymbol{\nu}_j,j} \left(\mathcal{K}_{d,\delta}(\boldsymbol{z}_{d,j}, \boldsymbol{y}_{d,j})\right) \, \mathrm{d}\boldsymbol{\nu}_j \right| &\le c^2 (\min(\delta,\sigma)\max(m_j,k_j) + \delta)^{2d\delta} \\ &\quad \times b^{-2\min(\delta,\sigma)\max(m_j,k_j) + 2t} \end{aligned}$$

and therefore we have

$$\begin{aligned} \widehat{\mathrm{wce}}^2(\mathcal{H}_{s,\delta}, \mathcal{A}_{\sigma,q,s}) &\le c^{2l} \sum_{\substack{\boldsymbol{m}\in\mathbb{N}_0^l \\ \|\boldsymbol{m}\|_1 > q}} \sum_{\substack{\boldsymbol{k}\in\mathbb{N}_0^l \\ \|\boldsymbol{k}\|_1 > q}} \prod_{j=1}^{l} (\min(\delta,\sigma)\max(m_j,k_j) + \delta)^{2d\delta} b^{-2\min(\delta,\sigma)\max(m_j,k_j) + 2t} \\ &\le c^{2l} b^{2tl} \delta^{2\delta s} \sum_{\substack{\boldsymbol{m}\in\mathbb{N}_0^l \\ \|\boldsymbol{m}\|_1 > q}} \sum_{\substack{\boldsymbol{k}\in\mathbb{N}_0^l \\ \|\boldsymbol{k}\|_1 > q}} \frac{1}{b^{\min(\delta,\sigma)\|\boldsymbol{m}\|_1}} \frac{1}{b^{\min(\delta,\sigma)\|\boldsymbol{k}\|_1}} \prod_{j=1}^{l} (\max(m_j,k_j) + 1)^{2\delta d}. \end{aligned}$$

Again using the arithmetic-geometric mean inequality we obtain that

$$\prod_{j=1}^{l} (\max(m_j,k_j) + 1)^{2\delta d} \le (\|\boldsymbol{m}\|_1 + 1)^{2s\delta} (\|\boldsymbol{k}\|_1 + 1)^{2s\delta}.$$

Hence

$$\widehat{\mathrm{wce}}(\mathcal{H}_{s,\delta}, \mathcal{A}_{\sigma,q,s}) \le c^l b^{tl} \delta^{\delta s} \sum_{\substack{\boldsymbol{m}\in\mathbb{N}_0^l \\ \|\boldsymbol{m}\|_1 > q}} \frac{1}{b^{\min(\delta,\sigma)\|\boldsymbol{m}\|_1}} (\|\boldsymbol{m}\|_1 + 1)^{2s\delta}$$

$$
\begin{aligned}
&= c^l b^{tl} \delta^{\delta s} \sum_{m=q+1}^{\infty} \frac{1}{b^{\min(\delta,\sigma)m}} (m+1)^{2s\delta} \binom{m+s-1}{s-1} \\
&\leq c^l b^{tl+\delta} \delta^{\delta s} \sum_{m=q+2}^{\infty} \frac{1}{b^{\min(\delta,\sigma)m}} m^{s(2\delta+1)-1}.
\end{aligned}
$$

Again it can be shown that there exists a constant $c'_{s,\delta} > 0$ such that

$$
\sum_{m=q+2}^{\infty} b^{-\min(\delta,\sigma)m} m^{s(2\delta+1)-1} \leq c'_{s,\delta} q^{s(2\delta+1)-1} \frac{1}{b^{\min(\delta,\sigma)q}}
$$

and hence we obtain the following theorem.

Theorem 2 *There is a constant $\widetilde{C} = \widetilde{C}(s,d,\sigma,\delta,t,b) > 0$ (which does not depend on q) such that the root-mean-square worst-case error for multivariate integration in the space $\mathcal{H}_{s,\delta}$, $\delta \geq 1$, using the algorithm $\mathcal{A}_{\sigma,q,\boldsymbol{\nu},s}$, with $\boldsymbol{\nu} \in [0,1)^s$ uniformly i.i.d., is bounded by*

$$
\widehat{\mathrm{wce}}(\mathcal{H}_{s,\delta}, \mathcal{A}_{\sigma,q,\boldsymbol{\nu},s}) \leq \widetilde{C} \cdot q^{s(2\delta+1)-1} \frac{1}{b^{\min(\delta,\sigma)q}}.
$$

As in the previous section we obtain the following result.

Corollary 2 *There is a constant $\widetilde{C} = \widetilde{C}(s,d,\sigma,\delta,t,b) > 0$ (which does not depend on q) such that the root-mean-square worst-case error for multivariate integration in the space $\mathcal{H}_{s,\delta}$, $\delta \geq 1$, using the algorithm $\mathcal{A}_{\sigma,q,\boldsymbol{\nu},s}$, with $\boldsymbol{\nu} \in [0,1)^s$ uniformly i.i.d., is bounded by*

$$
\widehat{\mathrm{wce}}(\mathcal{H}_{s,\delta}, \mathcal{A}_{\sigma,q,\boldsymbol{\nu},s}) \leq \widetilde{C} \cdot (1 + \log_b N)^{s(2\delta+1)-1+l\delta} \frac{1}{N^{\min(\delta,\sigma)}},
$$

where N is the number of points used by the algorithm $\mathcal{A}_{\sigma,q,\boldsymbol{\nu},s}$.

5 Numerical results

In this section we provide some numerical experiments where we try our algorithms on some sample functions. We consider the following functions:

$$
\begin{aligned}
&\text{ExpSum} &:&\quad f(x_1,\ldots,x_s) := \exp(x_1 + \cdots + x_s) \\
&\text{SingSum} &:&\quad f(x_1,\ldots,x_s) := s! \log(s - (x_1 + \cdots + x_s)) \\
&\text{WedgeSum} &:&\quad f(x_1,\ldots,x_s) := (s+1)!(x_1 + \cdots + x_s - s/2) 1_{[s/2,s]}(x_1 + \cdots + x_s) \\
&\text{ProdCos} &:&\quad f(x_1,\ldots,x_s) := \cos(x_1) \cdots \cos(x_s) \\
&\text{Mix1} &:&\quad \text{ExpSum} + \text{ProdCos}
\end{aligned}
$$

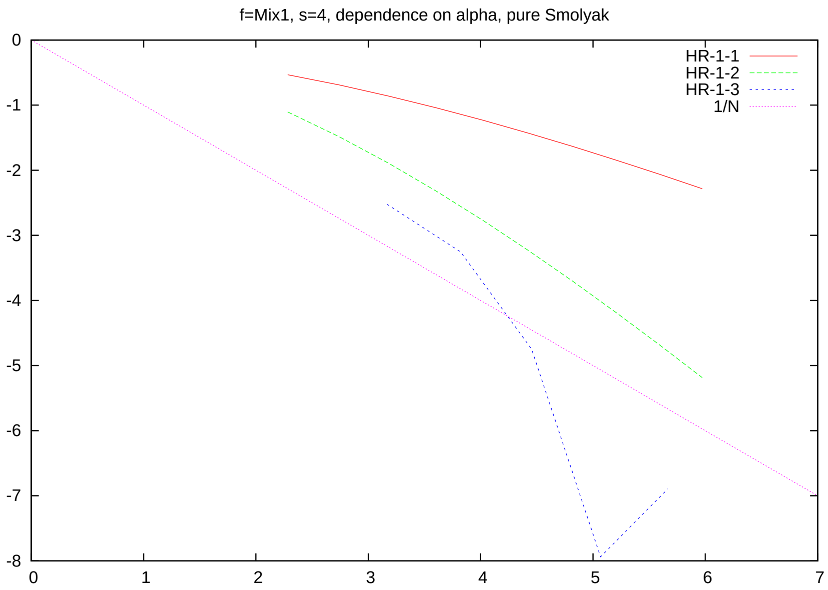

Figure 1: Log-error for rules with different degrees.

All our figures show the logarithmic relative error of integration drawn against the logarithm of the integration nodes used. Here logarithm means decadic logarithm. In the legends to the graphs, $HR-k-a$ means that we use a hybrid Smolyak quasi-Monte Carlo rule where the corresponding nested integration rule is a k-dimensional Faure rule with degree $\alpha = a$ (i.e., a (t, α, k)-sequence constructed as in [4] based on a $k\alpha$-dimensional Faure sequence).

Figure 1 shows the error for different values of α. One can see that higher values of α lead to faster convergence, as can be expected due to the smoothness of the integrand. It appears from Figures 2 and 3 that a pure quasi-Monte Carlo algorithm does better than the (hybrid) Smolyak algorithm based on a quasi-Monte Carlo rule. (It is of course possible that a Smolyak algorithm based on some different rule might perform better.) Figures 4 and 5 show the errors for a function with a singularity at one edge of the 4-dimensional unit cube. It seems that in that case greater values of α do not necessarily give better convergence rates. The convergence rates lie close to the $\frac{1}{N}$-line and pure quasi-Monte Carlo yields better results than pure Smolyak. A very similar behaviour is observed for a non-differentiable function in Figures 6 and 7. Figures 8 and 9 show how the dimension of the problem affects the convergence rate when using a pure Smolyak algorithm based on a $(t, \alpha, 1)$-sequence with $\alpha = 1$ for Figure 8 and $\alpha = 2$ for Figure 9. (Note also that in dimension 8 we need more than 10^5 integration nodes to build a Smolyak rule for which $q \geq s$.) As is to be expected, the error increases with increasing dimension and using an underlying digital $(t, 2, 1)$-sequence improves the convergence (compared to hybrid rules based on a digital $(t, 1, 1)$-sequence).

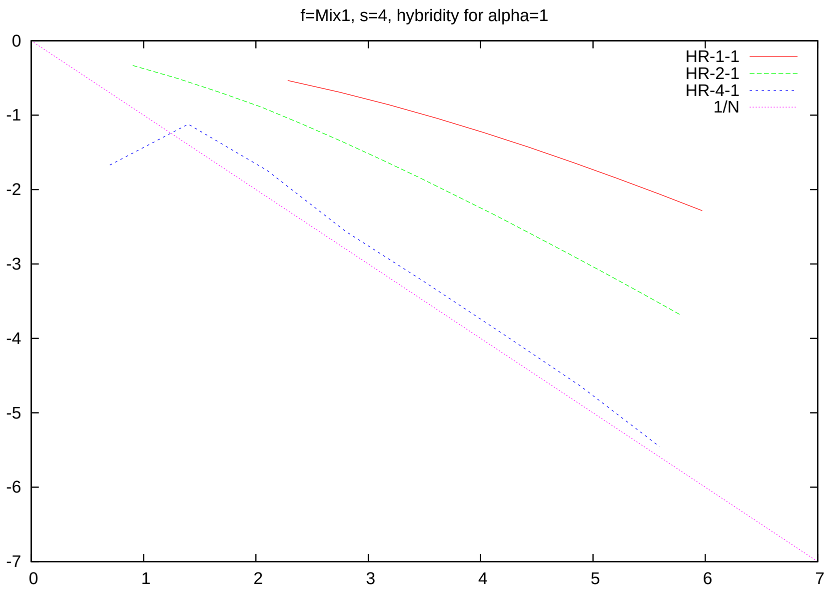

Figure 2: Log-error for rules with different hybridities.

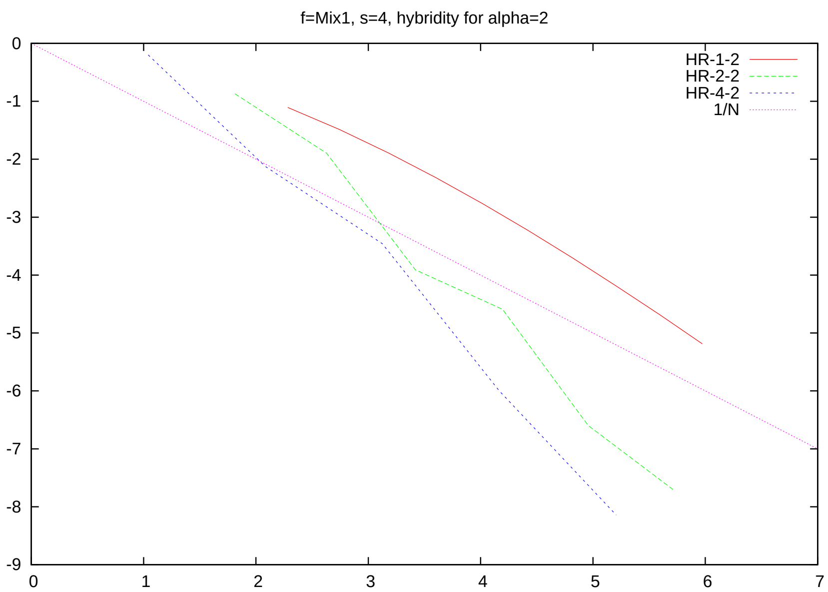

Figure 3: Log-error for rules with different hybridities.

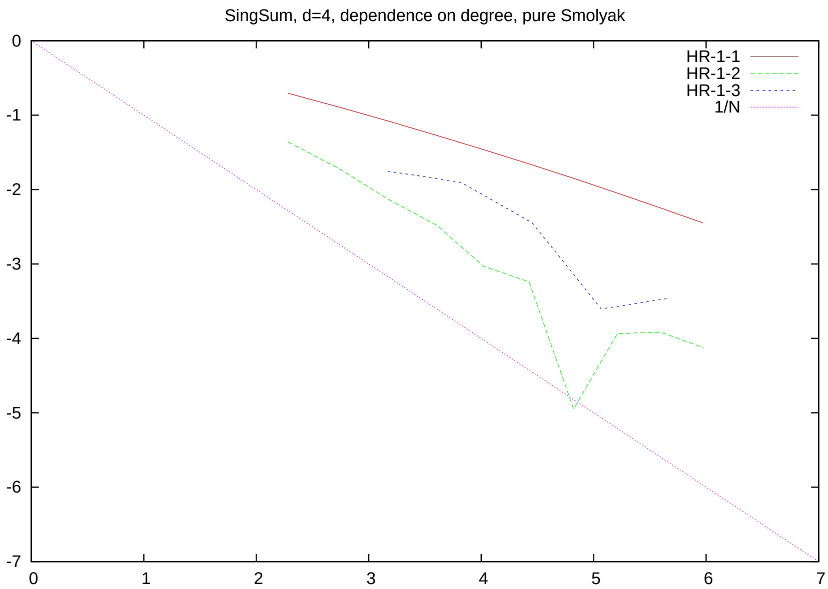

Figure 4: Log-error for rules with different α, singular function.

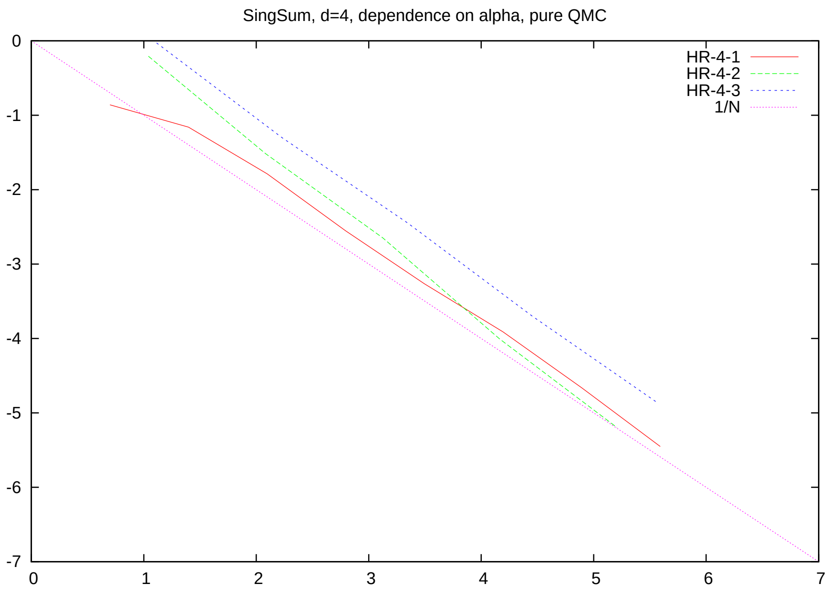

Figure 5: Log-error for rules with different α, singular function.

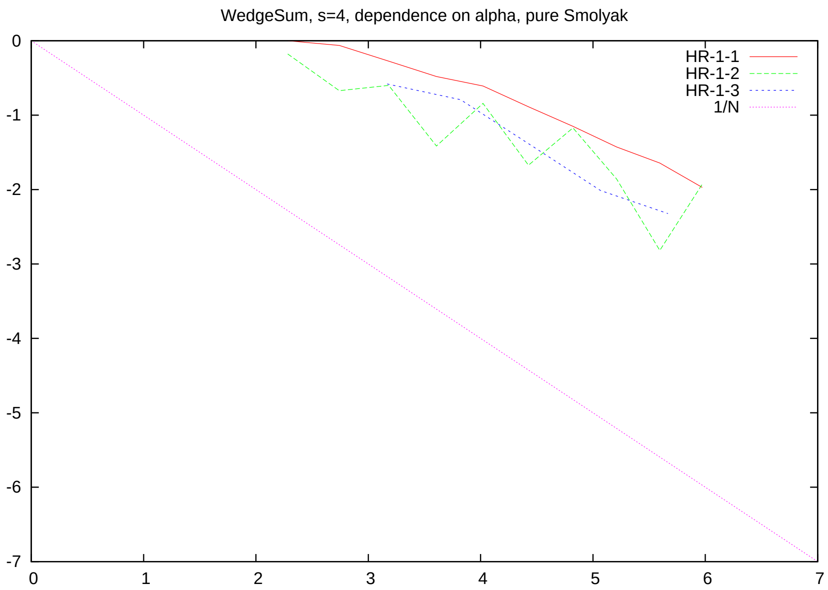

Figure 6: Log-error for rules with different α, non-differentiable function.

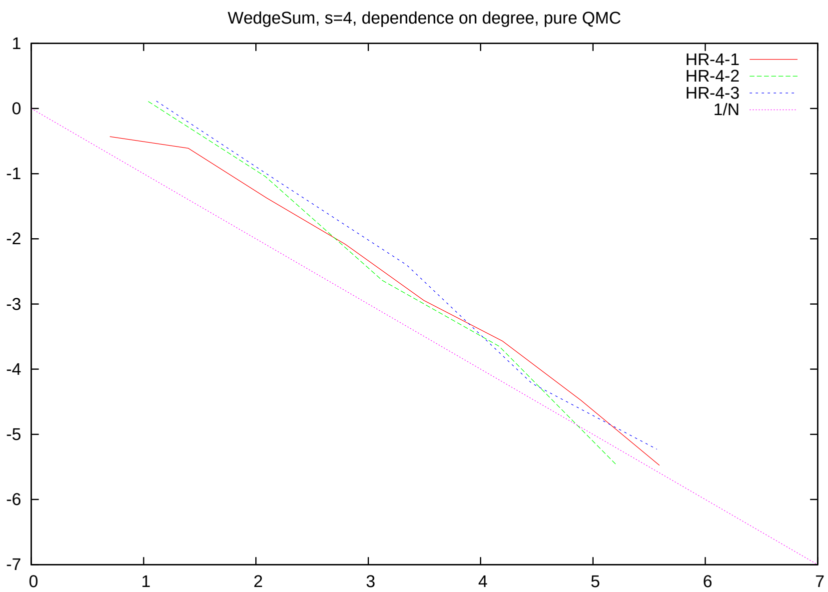

Figure 7: Log-error for rules with different α, non-differentiable function.

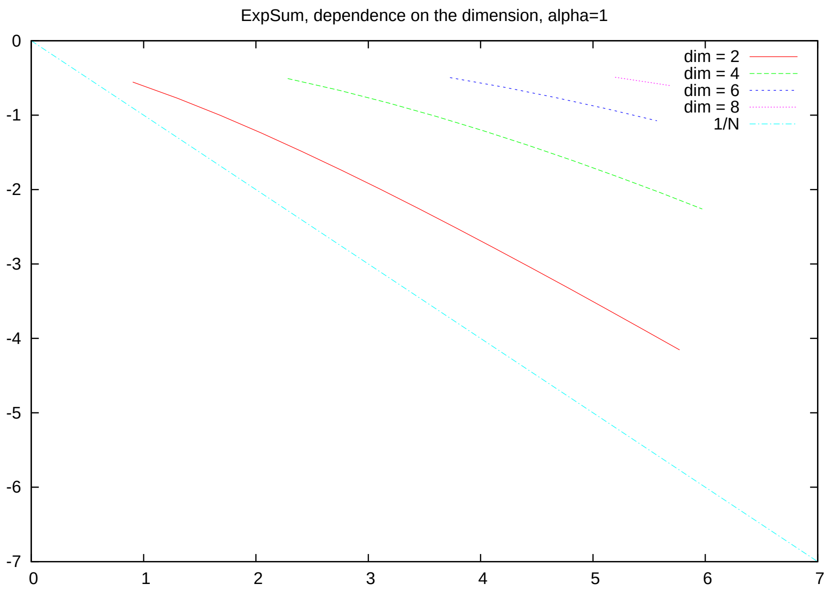

Figure 8: Log-error for different dimensions $\alpha = 1$.

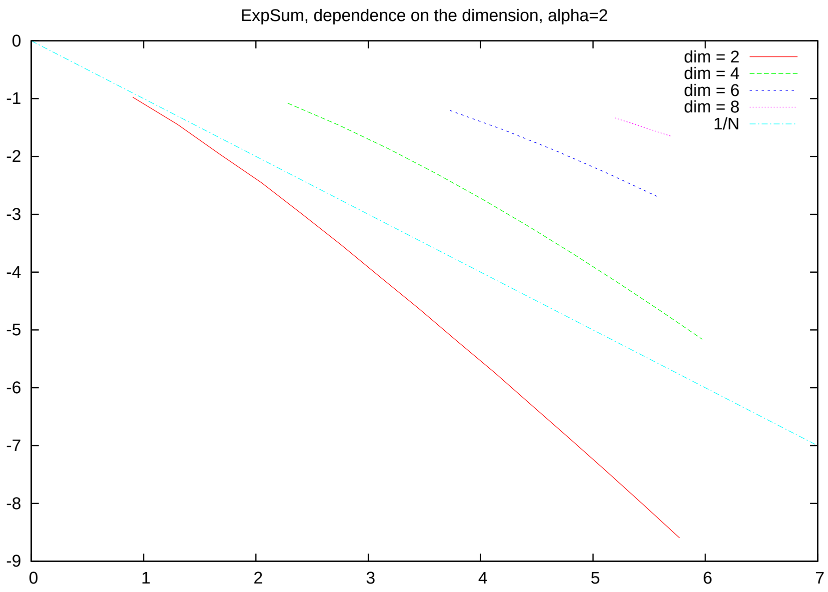

Figure 9: Log-error for different dimensions $\alpha = 2$.

References

[1] N. Aronszajn: Theory of reproducing kernels. Trans. Amer. Math. Soc., 68 (1950), 337–404.

[2] H.E. Chrestenson: A class of generalized Walsh functions. Pacific J. Math., 5 (1955), 17–31.

[3] R. Cranley and T.N.L. Patterson: Randomization of number theoretic methods for multiple integration. SIAM J. Numer. Anal., 13 (1976), 904–914.

[4] J. Dick: Explicit constructions of quasi-Monte Carlo rules of arbitrary high order. (2006), submitted. (available at http://www2.maths.unsw.edu.au/Contacts/profile.php?logname=josi)

[5] J. Dick: Walsh spaces containing smooth functions and quasi-Monte Carlo rules of arbitrary high order. (2006), submitted. (available at http://www2.maths.unsw.edu.au/Contacts/profile.php?logname=josi)

[6] J. Dick and F. Pillichshammer: Multivariate integration in weighted Hilbert spaces based on Walsh functions and weighted Sobolev spaces. J. Complexity, 21 (2005), 149–195.

[7] N.J. Fine: On the Walsh functions. Trans. Amer. Math. Soc., 65 (1949), 372–414.

[8] T. Gerstner and M. Griebel: Dimension-adaptive tensor-product quadrature. Computing, 71 (2003), 65–87.

[9] K. Niederdrenk: *Die endliche Fourier- und Walshtransformation mit einer Einführung in die Bildverarbeitung.* Vieweg, Braunschweig, 1982.

[10] H. Niederreiter: *Random Number Generation and Quasi-Monte Carlo Methods.* CBMS–NSF Series in Applied Mathematics 63, SIAM, Philadelphia, 1992.

[11] A.B. Owen: Monte Carlo variance of scrambled net quadrature. SIAM J. Numer. Anal., 34 (1997), 1884–1910.

[12] Plaskota, L., Wasilkowski, G.W.: Smolyak's algorithm for integration and L_1-approximation of multivariate functions with bounded mixed derivatives of second order. Numer. Algorithms, 36 (2004), 229–246.

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

[14] S.A. Smolyak: Quadrature and interpolation formulas for tensor products of certain classes of functions. Dokl. Akad. Nauk SSSR, 4 (1963), 240–243.

[15] J.L. Walsh:, A closed set of normal orthogonal functions. Amer. J. Math., 55 (1923), 5–24.

[16] G.W. Wasilkowski, H. Woźniakowski: Explicit cost bounds of algorithms for multivariate tensor product problems. J. Complexity, 11 (1995), 1–56.

Author's address:

Josef Dick, University of New South Wales Asia, 1 Kay Siang Road, Singapore 248922. *Email:* j.dick@unswasia.edu.sg
Gunther Leobacher, Institut für Finanzmathematik, Universität Linz, Altenbergerstraße 69, A-4040 Linz, Austria. *Email:* gunther.leobacher@jku.at
Friedrich Pillichshammer, Institut für Finanzmathematik, Universität Linz, Altenbergerstraße 69, A-4040 Linz, Austria. *Email:* friedrich.pillichshammer@jku.at