Numerical performance of optimized Frolov lattices in tensor product reproducing kernel Sobolev spaces

Christopher Kacwin Jens Oettershagen Mario Ullrich
Tino Ullrich

February 20, 2018

In this paper, we deal with several aspects of the universal Frolov cubature method, that is known to achieve optimal asymptotic convergence rates in a broad range of function spaces. Even though every admissible lattice has this favorable asymptotic behavior, there are significant differences concerning the precise numerical behavior of the worst-case error. To this end, we propose new generating polynomials that promise a significant reduction of the integration error compared to the classical polynomials. Moreover, we develop a new algorithm to enumerate the Frolov points from non-orthogonal lattices for numerical cubature in the d-dimensional unit cube $[0,1]^d$. Finally, we study Sobolev spaces with anisotropic mixed smoothness and compact support in $[0,1]^d$ and derive explicit formulas for their reproducing kernels. This allows for the simulation of exact worst-case errors which numerically validate our theoretical results.

1. Introduction

Many scientific approaches that are related to the treatment of real world phenomena rely on the computation of integrals on high-dimensional domains which often cannot be treated analytically. Examples include physics [4], computational finance [16], econometrics [18] and machine learning [2, 7, 30]. In this paper, we aim for efficient and stable numerical methods to approximately compute the integral

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

and give reliable error guarantees for a class F_d of d-variate functions. In fact, we are particularly interested in the worst-case error

$$e(n, F_d) := \sup_{\|f\|_{F_d} \le 1} |I_d(f) - Q_n^d(f)|, \tag{1.1}$$

for special cubature formulas of type

$$Q_n^d(f) \;=\; \frac{1}{n} \sum_{\boldsymbol{k} \in \mathbb{Z}^d} f(\boldsymbol{A}_n \boldsymbol{k}), \tag{1.2}$$

where $\boldsymbol{A}_n := n^{-1/d} \boldsymbol{A}$ is a suitable $d \times d$-matrix with $\det(\boldsymbol{A}) = 1$. This type of cubature rule has a long history going back to the 1970s, see Frolov [13]. In (1.2) the function f is assumed to be supported on a bounded domain Ω such that only finitely many summands contribute to the sum. Frolov noticed that the property

$$\mathrm{Nm}(\boldsymbol{A}) := \inf_{\boldsymbol{k} \in \mathbb{Z}^d \setminus \{0\}} \Big| \prod_{i=1}^{d} (\boldsymbol{A}\boldsymbol{k})_i \Big| > 0 \tag{1.3}$$

guarantees an optimal asymptotic worst-case behavior of (1.1) with respect to functions with L_p-bounded mixed derivative of order $r \in \mathbb{N}$ supported in $[0,1]^d$. In this context, optimality means that the worst-case error (1.1) can not be improved in the order sense by any other cubature formula using the same number of points. Note, that in case $|\det \boldsymbol{A}| = 1$ it can be shown that

$$n^{-1} \, |\{\boldsymbol{k} \colon \boldsymbol{A}_n \boldsymbol{k} \in \Omega\}| \to 1 \tag{1.4}$$

for every set Ω with (Lebesgue) volume 1 [32].

Frolov showed that the set of matrices satisfying (1.3) is not empty. Moreover, he gave a rather sophisticated number theoretic construction with a lot of potential for numerical analysis, as we will see in this paper. Starting with the irreducible (over $\mathbb{Q}$) polynomial

$$P_d(x) = \prod_{j=1}^{d} (x - 2j + 1) - 1 = \prod_{i=1}^{d} (x - \xi_i) \tag{1.5}$$

he defined the Vandermonde matrix

$$\boldsymbol{V} = \begin{pmatrix} 1 & \xi_1 & \cdots & \xi_1^{d-1} \\ 1 & \xi_2 & \cdots & \xi_2^{d-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \xi_d & \cdots & \xi_d^{d-1} \end{pmatrix} \in \mathrm{GL}_d(\mathbb{R}) \,. \tag{1.6}$$

One reason for the increasing interest in Frolov's cubature rule is certainly the fact that once a good matrix (1.3) is fixed the integration nodes are simply given as the rescaled image of the integer lattice points $\mathbb{Z}^d$ under the matrix $\boldsymbol{V}$. The method is therefore

comparably *simple*. Another striking aspect is a property which is sometimes called *universality*. The method (1.2) is not designed for a specific class of functions F_d as it is often the case for the commonly used quasi-Monte Carlo methods based on digital nets. In other words, we do not need to incorporate any a priori knowledge about the integrand (e.g. mixed or isotropic regularity etc.).

In this paper we are interested in an efficient implementation and the numerical performance of different Frolov type cubature methods for functions on $[0,1]^d$. First of all, this requires the efficient enumeration of Frolov lattice nodes in axis parallel boxes. It turned out that this is a highly non-trivial task which has been already considered by several authors [23], [24], [34] including three of the present ones. With a naive approach one may need to touch much more integer lattice points $\boldsymbol{k} \in \mathbb{Z}^d$ (overhead) to check whether $\boldsymbol{Ak} \in [0,1]^d$. This increases the runtime of an enumeration algorithm drastically in high dimensions. Here, the chosen irreducible polynomial for (1.6) has a significant effect. In [24] the authors observed that for $d = 2^m$ Chebyshev polynomials lead to an orthogonal lattice and an equivalent (orthogonal) lattice representation matrix with entries smaller than two in modulus. By exploiting rotational symmetry properties the mentioned overhead can be reduced and the enumeration procedure is less costly.

This observation already indicated that the choice of the polynomials in (1.6) is crucial. Unfortunately, Chebyshev polynomials and corresponding Vandermonde matrices (1.6) only provide (1.3) if $d = 2^m$. This has been shown for instance in Temlyakov [36]. The question remains how to fill the gaps. The classical Frolov polynomials are inappropriate in two respects. First, its roots spread in the range $[-d, d]$ such that (1.6) gets highly ill-conditioned. And secondly, although the lattice satisfies (1.3), the points are not really "spaces filling" meaning that the points accumulate around a lower dimensional manifold. This has a severe numerical impact for the worst-case error. In fact, the asymptotic rate of convergence is optimal but the preasymptotic behavior is useless for any practical issues.

One of the main contributions of the paper is the list of new improved generating polynomials given in Section 3 below. We give polynomials which are optimized according to the mentioned issues in dimensions $d = 1, ..., 10$, especially with a narrow distribution of its roots. As already mentioned above Chebyshev polynomials itself are not irreducible if d is not a power of two. However, they may provide admissible factors. This is the main idea of the construction and works if $d \in \{2, 3, 4, 5, 6, 8, 9, 10\}$. As for the case $d = 7$ a brute force search led to a polynomial with roots in $(-2.25, 1.75)$.

Due to the mentioned *universality* of Frolov's cubature rule, it is enough to fix the matrix and the corresponding lattice once and for all. In fact, the point construction does note depend on the respective framework. Therefore, it makes sense to generate the lattice points in a preprocessing step and make them available for practitioners. Our enumeration algorithm is similar to the one in [24] and extends to non-orthogonal lattices by exploiting a QR-factorization, see Section 4. Based on the above list of polynomials we generated a database of Frolov lattice nodes for dimensions up to $d = 10$ and $N \approx 10^6$ points. The points are available for download and direct use on the website

http://wissrech.ins.uni-bonn.de/research/software/frolov/

Having generated the cubature points we are now able to test the performance of various Frolov methods for functions with bounded mixed (weak) derivative, i.e,

$$\langle f^{(\boldsymbol{r})}, f^{(\boldsymbol{r})} \rangle_{L_2} = \|f^{(\boldsymbol{r})}\|_2^2 < \infty \,. \tag{1.7}$$

where $\boldsymbol{r} = (r_1, ..., r_d) \in \mathbb{N}^d$ is a smoothness vector with integer components satisfying

$$r = r_1 = ... = r_\nu < r_{\nu+1} \le r_{\nu+2} \le ... \le r_d \,. \tag{1.8}$$

A natural assumption, see (1.2), is the restriction to functions f supported inside the unit cube $\Omega = [0,1]^d$ satisfying (1.7). In this case the semi-norm (1.7) becomes a norm and the corresponding space a Hilbert space which will be denoted with $\mathring{H}^{\mathbf{r}}_{\mathrm{mix}}$.

The nowadays well-known worst-case error

$$e(n, \mathring{H}^{\mathbf{r}}_{\mathrm{mix}}) \asymp n^{-r} (\log n)^{(\nu-1)r}, \tag{1.9}$$

has been established in many classical papers [13], [11, 12], [36], see also the more recent papers [38] and [39] . Note, that we encounter another aspect of the *universality* property for this particular framework of *anisotropic mixed smoothness*. When using for instance a sparse grid approach (see e.g. Appendix A) for the numerical integration one has to know which direction is "rough" in the above sense to adapt the sparse grid accordingly. In fact, one samples more points in rough directions and less points in smoother direction. Frolov's method does not need this a priori information and behaves according to the optimal rate of convergence given in (1.9).

We will again provide a streamlined and self-contained proof in Section 6 pointing explicitly on the dependence of the constants on the dimension d, since the rate of convergence given by (1.9) completely hides this dependence. In fact, in case of one minimal smoothness component in (1.8) even the logarithm disappears completely and we have a pure polynomial rate as in the univariate setting. In Theorem 6.1 below we give a worst-case error bound which shows the influence of the dimension d. In addition, the result illustrates how the lattice invariants, like the polynomial discriminant D_P and the ℓ_∞-diameter of the smallest fundamental cell enter the error estimates.

Since $\mathring{H}^{\mathbf{r}}_{\mathrm{mix}}$ is embedded into the space of continuous functions a reproducing kernel exists [1]. We use the approach of Wahba [40] as a starting point to derive its reproducing kernel. Together with a standard correction procedure, cf. [3, Lem. 3, Thm. 11], we derive an explicit formula given in Theorem 5.4 and (5.15) below. The reproducing kernel is then being used to simulate the exact worst-case errors which represent the norm of the error functional, i.e. its Riesz representer, which can be computed exactly. This approach allows to gain insights into the true behavior of the constants that are involved in the bounds for the integration error and usually only are estimated. Let us emphasize once again that we simulate the worst-case error with respect to a whole function class rather than testing the algorithm on a single prototype test function.

Finally, in Section 7 we show the results of several numerical experiments. In the first part of the experiment section we compare different well-known methods for numerical integration in the reproducing kernel Hilbert space framework which we established in Sections 5 and 6. In particular, we compare Frolov lattices based on different generating polynomials, the classical Frolov polynomials and the improved polynomials from Section 3. As one would expect, the numerical behavior of the respective worst-case errors differ significantly for small n. Where the improved polynomials lead to a rather satisfactory error decay, the classical method is numerically completely useless if the dimension increases. Interestingly, in case $d = 2$ Frolov lattices according to the golden ratio polynomial compete with the Fibonacci lattice rule. We also compare Frolov lattices and sparse grids with respect to the numerical performance. Note, that the sparse grid cubature method represents a further method which is able to benefit from higher (mixed) smoothness. However, it is well known [9] that sparse grids show a worse behavior in the logarithm compared to Frolov lattices. Our experiments validate this theoretical fact. Among the considered methods (sparse grids, quasi-Monte Carlo) Frolov lattices behave best in our setting. In addition, Frolov lattices do not have to be adapted to the present anisotropy when considering anisotropic mixed smoothness. When considering one minimal smoothness component (1.8) we observe the same pure polynomial rate in different dimensions, only the constant differs. Note, that this effect would also be present for sparse grids adapted to the smoothness vector, which one has to know in advance.

Notation. As usual $\mathbb{N}$ denotes the natural numbers, $\mathbb{Z}$ denotes the integers, and $\mathbb{R}$ the real numbers . The letter d is always reserved for the underlying dimension in $\mathbb{R}^d, \mathbb{Z}^d$ etc. We denote with $\boldsymbol{x} \cdot \boldsymbol{y}$ the usual Euclidean inner product in $\mathbb{R}^d$. For $0 < p \leq \infty$ we denote with $|\cdot|_p$ and $\|\cdot\|_p$ the (d-dimensional) discrete ℓ_p-norm and the continuous L_p-norm on $\mathbb{R}$, respectively, where B_p^d denotes the respective unit ball in $\mathbb{R}^d$. The function $(\cdot)_+$ is given by $\max\{\cdot, 0\}$. With $\mathcal{F}$ we denote the Fourier transform given by $\mathcal{F}f(\boldsymbol{\xi}) := \int_{\mathbb{R}} f(x) \exp(-2\pi i \boldsymbol{x} \cdot \boldsymbol{\xi}) \, \mathrm{d}\boldsymbol{x}$ for a function $f \in L_1(\mathbb{R}^d)$ and $\boldsymbol{\xi} \in \mathbb{R}^d$. For two sequences of real numbers a_n and b_n we will write $a_n \lesssim b_n$ if there exists a constant $c > 0$ such that $a_n \leq c\, b_n$ for all n. We will write $a_n \asymp b_n$ if $a_n \lesssim b_n$ and $b_n \lesssim a_n$. With $\mathrm{GL}_d := \mathrm{GL}_d(\mathbb{R})$ we denote the group of invertible matrices over $\mathbb{R}$, whereas $\mathrm{SO}_d := \mathrm{SO}_d(\mathbb{R})$ denotes the group of orthogonal matrices over $\mathbb{R}$ with unit determinant. With $\mathrm{SL}_d(\mathbb{Z})$ we denote the group of invertible matrices over $\mathbb{Z}$ with unit determinant. The notation $D := \mathrm{diag}(x_1, ..., x_d)$ with $\boldsymbol{x} = (x_1, ..., x_d) \in \mathbb{R}^d$ refers to the diagonal matrix $D \in \mathbb{R}^{d \times d}$ with $\boldsymbol{x}$ at the diagonal. With $\gcd(a, b)$ we denote the greatest common divisor of two positive integers a, b. And finally, by $\mathbb{Z}[x]$ we denote the ring of polynomials with integer coefficients. Although we consider different generating matrices for admissible lattices in the forthcoming, we do not specify the matrix in the denotation Q_n^d. This is, because we will fix, for every dimension d under consideration, a matrix that is optimal in a sense that will be explained later. To be precise, for a given dimension d, the matrix A will be a multiple of the Vandermonde matrix as defined in Theorem 2.2 with the specific polynomials (and roots) as given in Table 1.

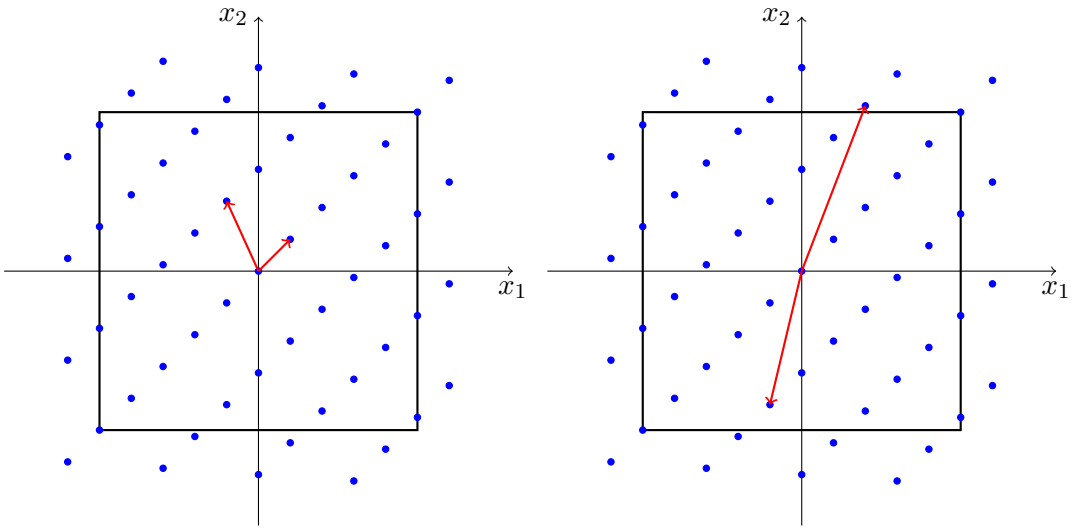


Figure 1: Equivalent lattice representations within the unit cube $\Omega = [-1/2, 1/2]^2$. The highlighted lattice elements are the column vectors of the corresponding lattice representation matrix.

2. Admissible lattices and their representation

For a matrix $\boldsymbol{T} \in \mathrm{GL}_d(\mathbb{R})$, we call $\{\boldsymbol{T}\boldsymbol{k} : \boldsymbol{k} \in \mathbb{Z}^d\} = \boldsymbol{T}(\mathbb{Z}^d)$ a (full-rank) lattice with lattice representation matrix $\boldsymbol{T}$.

For a matrix $\boldsymbol{U} \in \mathrm{SL}_d(\mathbb{Z})$, the matrices $\boldsymbol{T}$ and $\boldsymbol{T}\boldsymbol{U}$ generate the same lattice, and it can easily be shown that all possible lattice representations of $\boldsymbol{T}(\mathbb{Z}^d)$ are given this way. Therefore, it makes sense to define the determinant of a lattice $\boldsymbol{T}(\mathbb{Z}^d)$ as $|\det \boldsymbol{T}|$. We want to mention that for a given lattice, it is often preferred to have a lattice representation matrix $\boldsymbol{T} = (\boldsymbol{t}_1|\cdots|\boldsymbol{t}_d) \in \mathrm{GL}_d(\mathbb{R})$ with column vectors $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_d \in \mathbb{R}^d$ that are small with respect to some norm, cf. Figure 1.

Crucial for the performance of the Frolov cubature formula (1.2) will be the notion of *admissibility* which is settled in the following definition.

Definition 2.1 (Admissible lattice)**.** A lattice $\boldsymbol{T}(\mathbb{Z}^d)$ is called admissible if

$$\mathrm{Nm}(\boldsymbol{T}) \coloneqq \inf_{\boldsymbol{k}\in\mathbb{Z}^d\setminus\{0\}} \left| \prod_{i=1}^{d} (\boldsymbol{T}\boldsymbol{k})_i \right| > 0 \tag{2.1}$$

holds true.

Figure 2 illustrates this property. In fact, lattice points different from 0 lie outside of a hyperbolic cross with 'radius' $\mathrm{Nm}(\boldsymbol{T})$. Our construction of choice for admissible lattices is given by the following procedure.

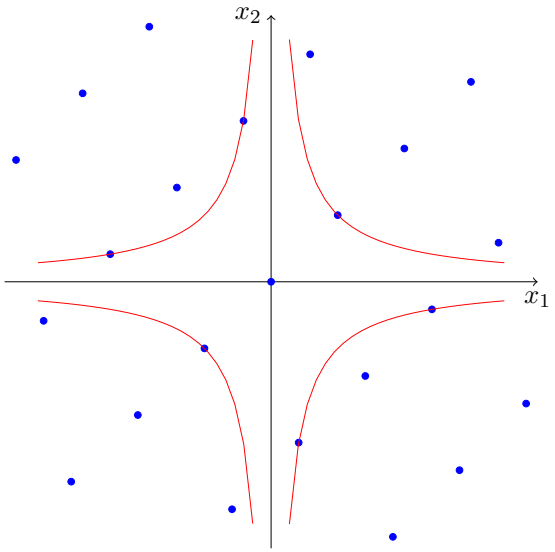


Figure 2: Admissible lattice and hyperbolic cross.

Proposition 2.2. *Let $P(x)$ be a polynomial of degree d satisfying*

- *P has integer coefficients,*
- *P has leading coefficient 1,*
- *P is irreducible over $\mathbb{Q}$,*
- *P has d different real roots $\xi_1, \dots, \xi_d$.*

The Vandermonde matrix

$$\boldsymbol{V} = \begin{pmatrix} 1 & \xi_1 & \cdots & \xi_1^{d-1} \\ 1 & \xi_2 & \cdots & \xi_2^{d-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \xi_d & \cdots & \xi_d^{d-1} \end{pmatrix} \in GL_d(\mathbb{R}) \tag{2.2}$$

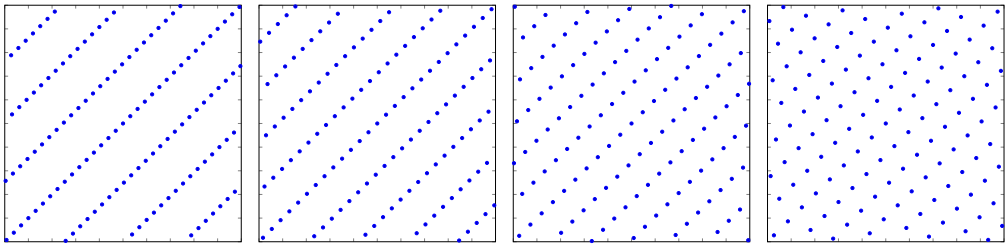
generates an admissible lattice $\boldsymbol{V}(\mathbb{Z}^d)$ with $\mathrm{Nm}(\boldsymbol{V}) = 1$. Its determinant equals the polynomial discriminant D_P of P:

$$|\det \boldsymbol{V}| = \prod_{k<l} |\xi_k - \xi_l| = D_P \,. \tag{2.3}$$

Moreover, it holds

$$\mathrm{Nm}(\boldsymbol{V}^{-\top}) = |\det \boldsymbol{V}|^{-2} = D_P^{-2} \,. \tag{2.4}$$

The necessary prequisites of P can be reformulated with concepts of algebraic number theory: P is the minimal polynomial of an algebraic integer of order d. For the proof



(a) $P(x) = x^2 - 17$, $D_P = 2\sqrt{17}$ (b) $P(x) = x^2 - 8$, $D_P = 2\sqrt{8}$ (c) $P(x) = x^2 - 3$, $D_P = 2\sqrt{3}$ (d) $P(x) = x^2 - x - 1$, $D_P = \sqrt{5}$

Figure 3: Lattices corresponding to different polynomials for $d = 2$. A small discriminant correlates with a good distribution of lattice points.

of this statement we refer to [23], or [19] and [26] for a thorough introduction into the theory of algebraic integers. The quantity (2.4) has a direct impact on the convergence behavior of the Frolov cubature formula and we therefore are interested in polynomials which maximize this quantity for a fixed d, i.e. have a small (or the smallest) polynomial discriminant D_P, cf. Figure 3.

Using Proposition 2.2, we obtain the lattice $\boldsymbol{V}(\mathbb{Z}^d)$ represented by a Vandermonde matrix $\boldsymbol{V}$. There are two problems with such matrices from the numerical point of view: First, they have large column vectors and therefore a large condition number, and second, its entries are of the form $\boldsymbol{V}_{ij} = \xi_i^{j-1}$ for which the calculation gets unstable for increasing j. However, we can bypass this problem using special polynomials, which will be discussed in the next section.

Lemma 2.3. *Let P be a polynomial which satisfies the prequisites in Proposition 2.2, and has roots $\xi_1, \dots, \xi_d$ which lie in $(-2, 2)$. Furthermore, let $\omega_1, \dots, \omega_d \in (-1, 1)$ be defined via the equation*

$$2\cos(\pi\omega_k) = \xi_k\,, \quad k = 1, \dots, d\,.$$

The lattice $\boldsymbol{V}(\mathbb{Z}^d)$ generated by the associated Vandermonde matrix

$$\boldsymbol{V} = \begin{pmatrix} 1 & \xi_1 & \cdots & \xi_1^{d-1} \\ 1 & \xi_2 & \cdots & \xi_2^{d-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \xi_d & \cdots & \xi_d^{d-1} \end{pmatrix} \in GL_d(\mathbb{R})$$

is also generated by the matrix $\boldsymbol{T}$ with

$$\boldsymbol{T}_{kl} = \begin{cases} 1 & l = 1\,, \\ 2\cos\left(\pi(l-1)\omega_k\right) & l = 2, \dots, d\,. \end{cases}$$

The resulting matrix $\boldsymbol{T}$ has entries in $(-2, 2)$ that can be calculated in a numerically stable way, which is optimal for our purposes. The proof is a straightforward application of Euler's identity and can be found in [23, 24].

Finally, we specify our choice of the matrices $\boldsymbol{A}$ and $\boldsymbol{A}_n$ used in Frolov's cubature formula (1.2) as

$$\boldsymbol{A} = |\det \boldsymbol{T}|^{-1/d}\boldsymbol{T}, \quad \boldsymbol{A}_n = n^{-1/d}\boldsymbol{A} \tag{2.5}$$

with $\boldsymbol{T}$ as in Lemma 2.3.

3. Improved generating polynomials

In this section, we consider polynomials which can be used to create admissible lattices. We will call such polynomials *admissible*, i.e. a d-th order polynomial P is admissible if it satisfies the prequisites of Proposition 2.2. At the end of this section, we provide a list of admissible polynomials with small discriminant for $d = 2, \ldots, 10$.

The study of Chebyshev Polynomials of the first and second kind provides us with a wide range of admissible polynomials. The most important features are their real and pairwise different roots, as well as the narrow distribution thereof. It is also quite fortunate to us that the decomposition into irreducible factors is well-understood and can be stated explicitly, see [31].

Definition 3.1. The Chebyshev Polynomials of the first kind $T_d(x)$ are defined recursively via

$$\begin{aligned} T_0(x) &= 1, \\ T_1(x) &= x, \\ T_d(x) &= 2xT_{d-1}(x) - T_{d-2}(x), \quad d \geq 2\,. \end{aligned}$$

The Chebyshev Polynomials of the second kind $U_d(x)$ are defined recursively via

$$\begin{aligned} U_0(x) &= 1, \\ U_1(x) &= 2x, \\ U_d(x) &= 2xU_{d-1}(x) - U_{d-2}(x), \quad d \geq 2\,. \end{aligned}$$

Lemma 3.2. *The Chebyshev polynomial $T_d(x)$ has the roots*

$$\cos\left(\frac{\pi(2k-1)}{2d}\right), \quad k = 1, \ldots, d\,.$$

The Chebyshev polynomial $U_d(x)$ has the roots

$$\cos\left(\frac{\pi k}{d+1}\right), \quad k = 1, \ldots, d\,.$$

The polynomials $T_d(x)$ and $U_d(x)$ are not admissible since they do not have leading coefficient 1. But they can be scaled appropriately to achieve this.

Lemma 3.3. *The scaled Chebyshev polynomials* $\widetilde{T}_d(x) = 2T_d(x/2)$ *and* $\widetilde{U}_d(x) = U_d(x/2)$ *have leading coefficient* 1 *and belong to* $\mathbb{Z}[x]$. *The scaled Chebyshev polynomial* $\widetilde{T}_d(x)$ *has the roots*

$$t_{d,k} = 2\cos\left(\frac{\pi(2k-1)}{2d}\right), \quad k = 1, \ldots, d\,.$$

The scaled Chebyshev polynomial $\widetilde{U}_d(x)$ *has the roots*

$$u_{d,k} = 2\cos\left(\frac{\pi k}{d+1}\right), \quad k = 1, \ldots, d\,.$$

¿From this lemma it follows directly that irreducible factors of $\widetilde{T}_d(x)$ and $\widetilde{U}_d(x)$ are admissible and have roots that lie in $(-2, 2)$. As already stated above, [31] lists the complete decomposition of Chebyshev Polynomials into irreducible factors, which we reformulate for the scaled versions in the next lemma.

Lemma 3.4. *For a fixed* $d > 1$, *we have*

$$\widetilde{T}_d(x) = \prod_h D_{d,h}(x)\,,$$

where $h \leq d$ *runs through all odd positive divisors of* d *and*

$$D_{d,h}(x) = \prod_{\substack{k=1 \\ \gcd(2k-1,d)=h}}^{d} (x - t_{d,k})$$

are all irreducible. It also holds

$$\widetilde{U}_d(x) = \prod_h E_{d,h}(x)\,,$$

where $h \leq d$ *runs through all positive divisors of* $2d+2$ *and*

$$E_{d,h}(x) = \prod_{\substack{k=1 \\ \gcd(k,2d+2)=h}}^{d} (x - u_{d,k})$$

are all irreducible.

It has been shown in [36] that $\widetilde{T}_d(x)$ is irreducible for $d = 2^m, m \in \mathbb{N}$ and that the corresponding lattice is orthogonal [24]. However, in this paper we are more interested in the irreducible factors of $\widetilde{U}_d(x)$, mainly for two reasons. First, it can be easily seen that the irreducible factors of $\widetilde{T}_d(x)$ have paired roots, i.e.

$$D_{d,h}(t_{d,k}) = 0 \Rightarrow D_{d,h}(t_{d,d-k+1}) = 0\,.$$

This means that either $D_{d,h}(x) = x$ or $D_{d,h}(x)$ is a polynomial of even degree, limiting the usefulness to our purposes. Second, it appears to be the case that the discriminant of $\widetilde{U}_d(x)$ is smaller than the discriminant of $\widetilde{T}_d(x)$, which makes the factors of $\widetilde{U}_d(x)$ more attractive to us. The following lemma is a consequence of Lemma 3.4.

dimension d	notation	polynomial & roots	discriminant D_P
2	$E_{4,2}(x)$	x^2+x-1 $2\cos\left(\pi\frac{2i}{2d+1}\right),\ i=1,\dots,d$	2.24
3	$E_{6,2}(x)$	x^3+x^2-2x-1 $2\cos\left(\pi\frac{2i}{2d+1}\right),\ i=1,\dots,d$	7
4	$E_{14,2}(x)$	$x^4-x^3-4x^2+4x+1$ $2\cos\left(\pi\frac{2}{15}\right), 2\cos\left(\pi\frac{4}{15}\right), 2\cos\left(\pi\frac{8}{15}\right), 2\cos\left(\pi\frac{14}{15}\right)$	33.54
5	$E_{10,2}(x)$	$x^5+x^4-4x^3-3x^2+3x+1$ $2\cos\left(\pi\frac{2i}{2d+1}\right),\ i=1,\dots,d$	121
6	$E_{12,2}(x)$	$x^6+x^5-5x^4-4x^3+6x^2+3x-1$ $2\cos\left(\pi\frac{2i}{2d+1}\right),\ i=1,\dots,d$	609.34
7	$P_7(x)$	$x^7+x^6-6x^5-4x^4+10x^3+4x^2-4x-1$ no explicit formula available	4487.14
8	$E_{16,2}(x)$	$x^8+x^7-7x^6-6x^5+15x^4+10x^3-10x^2-4x+1$ $2\cos\left(\pi\frac{2i}{2d+1}\right),\ i=1,\dots,d$	20256.8
9	$E_{18,2}(x)$	$x^9+x^8-8x^7-7x^6+21x^5+15x^4-20x^3-10x^2+5x+1$ $2\cos\left(\pi\frac{2i}{2d+1}\right),\ i=1,\dots,d$	130321
10	$E_{24,2}(x)$	$x^{10}-10x^8+35x^6+x^5-50x^4-5x^3+25x^2+5x-1$ $2\cos\left(\pi\frac{2}{25}\right), 2\cos\left(\pi\frac{4}{25}\right), 2\cos\left(\pi\frac{6}{25}\right), 2\cos\left(\pi\frac{8}{25}\right), 2\cos\left(\pi\frac{12}{25}\right),$ $2\cos\left(\pi\frac{14}{25}\right), 2\cos\left(\pi\frac{16}{25}\right), 2\cos\left(\pi\frac{18}{25}\right), 2\cos\left(\pi\frac{22}{25}\right), 2\cos\left(\pi\frac{24}{25}\right)$	873464

Table 1: Admissible polynomials with small discriminants for $d=2,\dots,10$.

Lemma 3.5. *Let $d > 1$. If $p = 2d + 1$ is a prime, the dth-order polynomial*

$$E_{2d,2}(x) = \prod_{k=1}^{d} (x - u_{2d,2k})$$

is admissible.

Proof. Consider the factorization of $\widetilde{U}_{2d}(x)$. We have $2(2d) + 2 = 4d + 2 = 2p$, therefore we have for $1 \leq k \leq 2d$

$$\gcd(k, 2(2d) + 2) = \begin{cases} 1 & k \text{ odd} \\ 2 & k \text{ even}\,. \end{cases}$$

This implies that $\widetilde{U}_{2d}(x) = E_{2d,1}(x)E_{2d,2}(x)$, which both are of order d. □

This simple rule covers the cases $d \in \{2, 3, 5, 6, 8, 9\}$. For the cases $d = 4$ and $d = 10$ we also did find good factors.

Lemma 3.6. *The polynomial $E_{14,2}(x)$ is of order 4 and admissible and the polynomial $E_{24,2}(x)$ is of order 10 and admissible.*

Proof. Both polynomials are admissible by definition, it remains to compute their order. We first consider $E_{14,2}(x)$. Here, $2 \cdot 14 + 2 = 30$, and for $1 \leq k \leq 14$ one has $\gcd(k, 30) = 2$ if and only if $k \in \{2, 4, 8, 14\}$. Therefore, $E_{14,2}(x)$ is a polynomial of order 4. Now consider $E_{24,2}(x)$. We have $2 \cdot 24 + 2 = 50$, and for $1 \leq k \leq 24$ one has $\gcd(k, 50) = 2$ if and only if $k \in \{2, 4, 6, 8, 12, 14, 16, 18, 22, 24\}$. Therefore, $E_{24,2}(x)$ is a polynomial of order 10. □

Unfortunately, the case $d = 7$ is not covered by the factorization of all $\widetilde{U}_d(x)$ and $\widetilde{T}_d(x)$. However, using a numerical brute force approach, we found the following polynomial.

Lemma 3.7. *The polynomial*

$$P_7(x) = x^7 + x^6 - 6x^5 - 4x^4 + 10x^3 + 4x^2 - 4x - 1$$

is of order 7 and admissible.

Proof. We have to prove that $P_7(x)$ is irreducible over $\mathbb{Q}$. It has leading coefficient 1 and coefficients in $\mathbb{Z}$, therefore it is irreducible over $\mathbb{Q}$ if and only if it is irreducible over $\mathbb{Z}$. Here, we consider irreducibility over $\mathbb{F}_2$, which is a sufficient condition for irreducibility over $\mathbb{Z}$. In $\mathbb{F}_2$, one has

$$P_7(x) \equiv x^7 + x^6 + 1\,.$$

Assume that this polynomial is reducible. Because it has no roots in $\mathbb{F}_2$, it would have to contain a factor of degree less then 4 which also has no root in $\mathbb{F}_2$. The possible

candidates are therefore x^2+x+1, x^3+x+1 and x^3+x^2+1. Doing a polynomial division with these three polynomials, one finds that

$$\begin{aligned} x^7+x^6+1 &\equiv (x^2+x+1)(x^5+x^3+x^2+1) && +x \\ &\equiv (x^3+x+1)(x^4+x^3+x^2) && +x^2+1 \\ &\equiv (x^3+x^2+1)(x^4+x+1) && +x^2+x \end{aligned}$$

and we have a contradiction. Therefore, $P_7(x)$ is irreducible over $\mathbb{F}_2$, and subsequently also over $\mathbb{Q}$. □

Even though Lemma 2.3 is not applicable for this polynomial because its roots lie in $(-2.25, 1.75)$, they still lie close to each other, which results in a good polynomial discriminant. Regarding the lattice representation issue in the $d = 7$ case, one has to compute the Vandermonde matrix $\boldsymbol{V}$ explicitly (using an arbitrary precision data type to avoid stability issues) and find a good lattice representation matrix $\boldsymbol{T}$ by means of a lattice reduction algorithm, see for instance [25].

This completes our list of polynomials used for the dimensions $2 \leq d \leq 10$. We attach Table 1 collecting all polynomials and useful information.

4. Efficient enumeration of Frolov lattices in d-cubes

In this section we present an enumeration algorithm to determine the set of integration points for the Frolov cubature formula. The approach is similar to the one in [24] for orthogonal lattices, Here, we generalize the method for arbitrary lattices.

4.1. Enumeration of non-orthogonal Frolov lattices

We fix the integration domain $\Omega = [-1/2, 1/2]^d$ and a lattice $\boldsymbol{T}(\mathbb{Z}^d)$ with lattice representation matrix $\boldsymbol{T}$. We are interested in the discrete set

$$\mathcal{N} = \Omega \cap \boldsymbol{T}(\mathbb{Z}^d) = \{\boldsymbol{T}\boldsymbol{k} \in \Omega : \ \boldsymbol{k} \in \mathbb{Z}^d\}\,.$$

Our strategy is to consider a slightly larger set $\mathcal{B} \supset \mathcal{N}$ which allows for explicit enumeration in a straightforward way. We choose

$$\mathcal{B} = B_{\sqrt{d}/2}(0) \cap \boldsymbol{T}(\mathbb{Z}^d) = \left\{\boldsymbol{T}\boldsymbol{k} : \ \|\boldsymbol{T}\boldsymbol{k}\|_2^2 \leq \frac{d}{4}, \boldsymbol{k} \in \mathbb{Z}^d\right\}\,.$$

Using the matrix decomposition

$$\boldsymbol{T} = \boldsymbol{Q}\boldsymbol{R}\,,$$

Algorithm 1: Assemblation of the set $\mathcal{N} = \Omega \cap \boldsymbol{T}(\mathbb{Z}^d)$.

Input:
Integration domain $\Omega = [-1/2, 1/2]^d$,
Lattice representation matrix $\boldsymbol{T} = \boldsymbol{Q}\boldsymbol{R}$

---

set $\mathcal{N} = \emptyset$
set $\boldsymbol{m} = (0, \ldots, 0)^\top$
run `assemble` $(\mathcal{N}, d, \boldsymbol{m})$

Function `assemble` $(\mathcal{N}, j, \boldsymbol{m})$
- **if** $j \geq 2$ **then**
  - Determine the set $K_j = \{k_j \in \mathbb{Z} : \; g_j(0, \ldots, 0, k_j, m_{j+1}, \ldots, m_d) \leq \frac{d}{4} - \sum_{i=j+1}^{d} g_i(\boldsymbol{m})\}$
  - **forall** $k_j \in K_j$ **do**
    - **set** $m_j = k_j$
    - `assemble` $(\mathcal{N}, j-1, \boldsymbol{m})$
  - **set** $m_j = 0$
- **if** $j = 1$ **then**
  - Determine the set $K_1 = \{k_1 \in \mathbb{Z} : \; g_j(k_1, m_2, \ldots, m_d) \leq \frac{d}{4} - \sum_{i=2}^{d} g_i(\boldsymbol{m})\}$
  - **forall** $k_1 \in K_1$ **do**
    - **set** $m_1 = k_1$
    - **if** $\boldsymbol{T}\boldsymbol{m} \in \Omega$ **then**
      - **set** $\mathcal{N} = \mathcal{N} \cup \{\boldsymbol{T}\boldsymbol{m}\}$
  - **set** $m_1 = 0$

Output: Set of lattice points $\mathcal{N}$

where $\boldsymbol{Q}$ is an orthogonal matrix and $\boldsymbol{R}$ is an upper triangular matrix, we can rewrite this set as

$$\mathcal{B} = \left\{ \boldsymbol{T}\boldsymbol{k} : \; \|\boldsymbol{R}\boldsymbol{k}\|_2^2 \leq \frac{d}{4}, \boldsymbol{k} \in \mathbb{Z}^d \right\} .$$

The function $\|\boldsymbol{R} \cdot\|_2^2$ can be split up into additive parts

$$\begin{aligned} \|\boldsymbol{R}\boldsymbol{k}\|_2^2 &= \sum_{i=1}^{d} g_i(\boldsymbol{k}) \\ g_i(\boldsymbol{k}) &= (\boldsymbol{R}\boldsymbol{k})_i^2, \quad i = 1, \ldots, d \,, \end{aligned}$$

and from the upper triangular structure of $\boldsymbol{R}$ it follows that $g_j(\boldsymbol{k})$ only depends on the components $k_j, \ldots, k_d$. For an integer vector $\boldsymbol{k}$ we therefore have

$$\|\boldsymbol{R}\boldsymbol{k}\|_2^2 \leq \frac{d}{4} \Longleftrightarrow g_j(\boldsymbol{k}) \leq \frac{d}{4} - \sum_{i=j+1}^{d} g_i(\boldsymbol{k}) \,, \quad j = 1, \ldots, d \,. \tag{4.1}$$

Fixing the coordinates $k_{j+1}, \ldots, k_d$ results in explicitly solvable inequalities for k_j, since the right hand side is constant and the left hand side is a quadratic function in k_j. Therefore, the set $\mathcal{N}$ can be assembled with Algorithm 1.

This algorithm iterates over all elements of $\mathcal{B}$, which determines the complexity that is of order

$$\mathrm{vol}_d\left(B_{\sqrt{d}/2}(0)\right)/|\det \boldsymbol{T}| \asymp 2^d \cdot |\mathcal{N}|\,.$$

This is certainly true if the sets K_j appearing in the algorithm are all nonempty, and this should be the case for a lattice with a small determinant and a good choice of its representation matrix. The exponential dependence on d is of minor importance here; Once the Frolov integration points are computed and stored, they can be reused for numerical integration.

4.2. Numerical results

In Table 2 the running times for the enumeration of the Frolov lattice points in $[0,1]^d$ with Algorithm 1 are provided for dimensionalities $d \in \{2, 3, \ldots, 9\}$. Firstly, we observe that the number of points N converges to the scaling factor n, as n becomes large, cf. (1.4).

Moreover, one can observe the *linear runtime* of the algorithm in terms of the number of points N: If the number of points N is quadrupled, then also the required time to assemble these $4N$ points is approximatively quadrupled. However, comparing the runtimes for small d and large d, it is apparent that a dimension-dependent constant is involved. This is analogous to the orthogonal setting for $d = 2^k, k \in \mathbb{N}$, as it was treated in [24].

The resulting point sets for dimension $d \in \{2, 3, \ldots, 10\}$ are available for download at `http://wissrech.ins.uni-bonn.de/research/software/frolov/`.

5. Compactly supported functions with bounded mixed derivative in L_2

5.1. Characterization of the space

We denote with $S(\mathbb{R}^d)$ the usual Schwartz space. Let $\mathbf{r} = (r_1, ..., r_d) \in \mathbb{N}^d$ be a smoothness vector with integer components. Then we define the semi-norm

$$\|\varphi\|^2_{H^{\mathbf{r}}_{\mathrm{mix}}} := \sum_{e \subset [d]} \Big\| \Big(\prod_{i \in e} \frac{\partial^{r_i}}{\partial x_i^{r_i}} \Big) \varphi \Big\|_2^2$$

Dim. d	Scaling n	Points N	Time (s)
2	1024	1023	4.4e-05
2	4096	4093	0.000158
2	16384	16387	0.00053
2	65536	65533	0.002117
2	262144	262147	0.00823
2	1048576	1048575	0.096369
3	1024	1021	0.000105
3	4096	4093	0.000341
3	16384	16387	0.001213
3	65536	65537	0.004547
3	262144	262149	0.017474
3	1048576	1048581	0.114605
4	1024	1023	0.00024
4	4096	4103	0.000805
4	16384	16395	0.002844
4	65536	65551	0.010464
4	262144	262155	0.038923
4	1048576	1048579	0.248508
5	1024	1021	0.00061
5	4096	4093	0.002072
5	16384	16359	0.007013
5	65536	65533	0.025019
5	262144	262141	0.129366
5	1048576	1048591	0.473579

Dim. d	Scaling n	Points N	Time (s)
6	1024	1005	0.00146
6	4096	4087	0.004961
6	16384	16401	0.016533
6	65536	65513	0.059226
6	262144	262161	0.241978
6	1048576	1048585	0.943112
7	1024	1009	0.003208
7	4096	4099	0.011418
7	16384	16383	0.039014
7	65536	65531	0.13972
7	262144	262117	0.513067
7	1048576	1048573	2.0007
8	1024	1029	0.007961
8	4096	4051	0.025833
8	16384	16441	0.094269
8	65536	65539	0.329561
8	262144	262207	1.20636
8	1048576	1048767	4.59066
9	1024	997	0.017742
9	4096	4035	0.066017
9	16384	16517	0.223132
9	65536	65557	0.76848
9	262144	262107	2.77068
9	1048576	1048631	10.4136

Table 2: Running times for the assemblation of Frolov cubature points in $[0,1]^d$.

where $\|\cdot\|_2$ denotes the $L_2(\mathbb{R}^d)$-norm. Clearly this norm is induced by an inner product. By Plancherel's theorem together with well-known properties of the Fourier transform, see (6.2) below, we may rewrite

$$\begin{aligned}\|\varphi\|_{H^{\mathbf{r}}_{\mathrm{mix}}} &= \Big\|\mathcal{F}^{-1}\Big[\Big(\prod_{i=1}^{d}(1+|2\pi\xi_i|^{2r_i})\Big)^{1/2}\mathcal{F}\varphi(\boldsymbol{\xi})\Big]\Big\|_2 \\ &= \Big\|\Big(\prod_{i=1}^{d}(1+|2\pi\xi_i|^{2r_i})\Big)^{1/2}\mathcal{F}\varphi(\boldsymbol{\xi})\Big\|_2 = \|v_{\boldsymbol{r}}(\boldsymbol{\xi})\mathcal{F}\varphi\|_2\,,\end{aligned} \tag{5.1}$$

where we define

$$v_{\mathbf{r}}(\boldsymbol{x}) := \Big(\prod_{i=1}^{d}(1+|2\pi x_i|^{2r_i})\Big)^{1/2}. \tag{5.2}$$

Let now Ω be a bounded domain in $\mathbb{R}^d$. We denote with $C_0^\infty(\Omega)$ the space of all infinitely many times differentiable (real-valued) functions $\varphi : \mathbb{R}^d \to \mathbb{R}$ with $\operatorname{supp}\varphi \subset \Omega$. Finally,

we define the space

$$\mathring{H}^{\mathbf{r}}_{\mathrm{mix}}(\overline{\Omega}) := \overline{C_0^\infty(\Omega)}^{\|\cdot\|_{H^{\mathbf{r}}_{\mathrm{mix}}}} \tag{5.3}$$

by completion with respect to the norm $\|\cdot\|_{H^{\mathbf{r}}_{\mathrm{mix}}}$. As a consequence we get that $\mathring{H}^{\mathbf{r}}_{\mathrm{mix}}(\overline{\Omega})$ is a Hilbert space which consists of $r_i - 1$ times continuously differentiable functions (mixed in each component) on $\mathbb{R}^d$ which vanish on $\mathbb{R}^d \setminus \Omega$.

We will now consider a more specific situation. Let $\Omega = (0,1)^d$. Then it holds

$$\mathring{H}^{\mathbf{r}}_{\mathrm{mix}} := \mathring{H}^{\mathbf{r}}_{\mathrm{mix}}([0,1]^d) = \mathring{H}^{r_1}([0,1]) \otimes \cdots \otimes \mathring{H}^{r_d}([0,1]) \tag{5.4}$$

in the sense of tensor products of Hilbert spaces, where $\mathring{H}^r = \mathring{H}^{r_i}([0,1])$ is the univariate version of the above defined spaces. Functions in this class satisfy a left and a right boundary condition, namely $f^{(j)}(0) = f^{(j)}(1) = 0$ for $j = 0, ..., r-1$.

The first assertion in the following lemma is a direct consequence of Taylor's theorem and the homogeneous boundary condition of the function and all its derivatives. The second one follows from (i) together with Hölder's inequality.

Lemma 5.1. *Let $\boldsymbol{r} \in \mathbb{N}^d$.* (i) *Every function $\varphi \in C_0^\infty((0,1)^d)$ admits the following representation*

$$\varphi(x_1, ..., x_d) = \int_0^1 \cdots \int_0^1 \varphi^{(\boldsymbol{r})}(t_1, ..., t_d) \prod_{i=1}^d \frac{(x_i - t_i)_+^{r_i - 1}}{(r_i - 1)!}\, dt_1 ... dt_d\,.$$

(ii) *Let $e \subset [d]$. Then*

$$\Big\| \Big(\prod_{i \in e} \frac{\partial^{r_i}}{\partial x_i^{r_i}} \Big) \varphi \Big\|_2^2 \le \|\varphi^{(\boldsymbol{r})}\|_2^2 \prod_{i \in e} \frac{1}{[(r_i - 1)!]^2 (2r_i - 1) 2r_i}$$

and therefore

$$\|\varphi\|^2_{H^{\boldsymbol{r}}_{mix}} \le \|\varphi^{(\boldsymbol{r})}\|_2^2 \sum_{e \subset [d]} \prod_{i \in e} \frac{1}{[(r_i - 1)!]^2 (2r_i - 1) 2r_i}\,.$$

Remark 5.2. *(a)* Note, that the assertions in Lemma 5.1 hold true for any function $\varphi \in S(\mathbb{R}^d)$ with $\operatorname{supp} \varphi \subset \mathbb{R}^d_+$. We only need zero boundary values at 0.

(b) The previous lemma shows that the semi-norm $\|\cdot\|_{\mathring{H}^{\boldsymbol{r}}_{\mathrm{mix}}}$ induced by the bilinear form

$$\langle \varphi, \psi \rangle_{\mathring{H}^{\boldsymbol{r}}_{\mathrm{mix}}} := \int_{[0,1]^d} \varphi^{(\boldsymbol{r})}(\boldsymbol{x}) \psi^{(\boldsymbol{r})}(\boldsymbol{x})\, d\boldsymbol{x} \tag{5.5}$$

is actually a norm on $C_0^\infty((0,1)^d)$ since the bilinear form is positive definite as a consequence of (ii). Hence, we could have also used this semi-norm for the completion in (5.3). As it turns out Lemma 5.1 and (5.5) are actually the key to derive the reproducing kernel for the space $\mathring{H}^{\boldsymbol{r}}_{\mathrm{mix}}$.

(c) We have an explicit upper bound for the norm equivalence constant in (ii). Suppose that we have a constant smoothness vector $\boldsymbol{r} = (r, ..., r)$ with $r \in \mathbb{N}$. Then it holds

$$\begin{aligned}\sum_{e\subset[d]}\prod_{i\in e}\frac{1}{[(r_i-1)!]^2(2r_i-1)2r_i} &= \sum_{i=0}^{d}\binom{d}{i}\Big(\frac{1}{[(r-1)!]^2(2r-1)2r}\Big)^i \\ &= \Big(1+\frac{1}{[(r-1)!]^2(2r-1)2r}\Big)^d.\end{aligned} \tag{5.6}$$

Hence, if $r = 1$ the constant is bounded by $(3/2)^d$, in case $r = 2$ we have $(13/12)^d$ and in case $r = 3$ already $(61/60)^d$.

5.2. The reproducing kernel of $\mathring{H}^{\mathbf{r}}_{\mathrm{mix}}$

In the sequel we will identify the space $\mathring{H}^{\mathbf{r}}_{\mathrm{mix}}$ as a reproducing kernel Hilbert space. We are looking for a kernel function $\mathring{K}^{\boldsymbol{r}}_d(\boldsymbol{x}, \boldsymbol{y})$ such that for every $f \in \mathring{H}^{\mathbf{r}}_{\mathrm{mix}}$

$$\langle f(\cdot), \mathring{K}^{\boldsymbol{r}}_d(\boldsymbol{x}, \cdot)\rangle_{\mathring{H}^{\boldsymbol{r}}_{\mathrm{mix}}} = f(\boldsymbol{x}) \quad , \quad \boldsymbol{x} \in [0,1]^d.$$

To this end, we may derive the reproducing kernels of the univariate spaces $\mathring{H}^{r_i}$. The reproducing kernel of the tensor product space (5.4) is then given by the point-wise product of the univariate kernels

$$\mathring{K}^{\boldsymbol{r}}_d(\boldsymbol{x}, \boldsymbol{y}) = \prod_{\ell=1}^{d}\mathring{K}^{r}_1(x_\ell, y_\ell). \tag{5.7}$$

Therefore, the problem of computing $\mathring{K}^{\boldsymbol{r}}_d(\boldsymbol{x}, \boldsymbol{y})$ is reduced to the construction of $\mathring{K}^{r}_1 : [0,1] \times [0,1] \to \mathbb{R}$.

Let us first recall a general fact for Hilbert spaces and orthogonal sums. To this end, let $U := \mathrm{span}\{u_0, \ldots, u_{r-1}\} \subset \mathcal{H}$ be an r-dimensional subspace of a Hilbert space $\mathcal{H}$. Using Gram-Schmidt orthogonalization, the orthogonal projection $P_U : \mathcal{H} \to U$ is given by

$$P_U(f)(x) = \sum_{j=0}^{r-1}\left(\sum_{k=0}^{r-1} G^{-1}_{j,k}\cdot\langle f, u_k\rangle_{\mathcal{H}}\right)u_j, \tag{5.8}$$

where the Gramian matrix $\mathbf{G} = (\langle u_j, u_k\rangle_{\mathcal{H}})_{j,k=0}^{r-1} \in \mathbb{R}^{r\times r}$. Moreover, the projection onto the orthogonal complement $U^\perp = \mathcal{H} \ominus U$ is $P_{U^\perp} f = (\mathrm{Id} - P_U)f$.

The next Lemma provides the necessary utilities to compute the reproducing kernel of closed subspaces that are defined via homogeneous boundary conditions.

Lemma 5.3. *Let $\mathcal{H}_K$ be a RKHS with kernel $K : [0,1] \times [0,1] \to \mathbb{R}$. Assuming that $K(x, \cdot)$ is r times weakly differentiable, let $u_j := K^{(0,j)}(\cdot, 1) := \frac{\partial^j}{\partial y^j}K(\cdot, y)_{|y=1}$ for $j = 0, \ldots, r-1$ and $U = \mathrm{span}\{u_0, \ldots, u_{r-1}\}$. Then it holds that*

(i) For $j = 0, \dots, r-1$, the Riesz representer of the functional $f \mapsto f^{(j)}(1)$ in $\mathcal{H}_K$ is given by u_j, i.e.

$$\langle f, u_j \rangle_{\mathcal{H}_K} = f^{(j)}(1) \quad \text{for all } f \in \mathcal{H}_K.$$

(ii) The reproducing kernel $K_{U^\perp}$ of $U^\perp \subset \mathcal{H}_K$, i.e. the orthogonal complement of U in $\mathcal{H}_K$, is given by

$$K_{U^\perp}(x,y) = P_{U^\perp} K(\cdot, y)(x) = K(x,y) - \sum_{j=0}^{r-1} \sum_{k=0}^{r-1} G_{j,k}^{-1} u_j(x) u_k(y). \tag{5.9}$$

(iii) It holds that

$$U^\perp = \{ f \in \mathcal{H}_K : f^{(j)}(1) = 0, j = 0, \dots, r-1 \}.$$

Proof. (i) is [3, Lem. 10] for the linear functional $f \mapsto f^{(j)}(1)$ and (ii) follows by applying [3, Thm. 11] to (5.8). Finally, regarding (iii) we note that it holds for all $f \in U^\perp$ that

$$\langle f, u_j \rangle_{\mathcal{H}_K} = \langle f, K^{(0,j)}(\cdot, 1) \rangle_{\mathcal{H}_K} = f^{(j)}(1) = 0.$$

□

We want to apply this machinery to $\mathring{H}^r$ with $r \in \mathbb{N}$. The observation in Lemma 5.1 together with (5.5) gives rise to use the approach of Wahba [40, 1.2] as a starting point. Let us define the kernel function

$$K_1^r(x,y) := \int_0^1 \frac{(x-t)_+^{r-1}}{(r-1)!} \cdot \frac{(y-t)_+^{r-1}}{(r-1)!} \, dt \quad , \quad x, y \in [0,1]\,. \tag{5.10}$$

Then it is immediately clear from Lemma 5.1,(i) (and a straight-forward density argument) that

$$f(x) = \langle f(\cdot), K_1^r(x, \cdot) \rangle_{\mathring{H}^r} \quad , \quad x \in [0,1]\,.$$

Indeed, recall that the inner product $\langle \cdot, \cdot \rangle_{\mathring{H}^r}$ stems from (5.5) and that

$$(K_1^r)^{(0,r)}(x,y) = \frac{(x-y)_+^{r-1}}{(r-1)!}\,.$$

It is possible to give an explicit formula for (5.10) by using that

$$K_1^r(x,y) := \int_0^{\min\{x,y\}} \frac{(x-t)_+^{r-1}}{(r-1)!} \cdot \frac{(y-t)_+^{r-1}}{(r-1)!} \, dt\,. \tag{5.11}$$

Interpreting this as a Taylor remainder term we find

$$K_1^r(x,y) = \frac{(-1)^r}{(2r-1)!} \Big[\sum_{k=r}^{2r-1} \binom{2r-1}{k} (-\min\{x,y\})^k \max\{x,y\}^{2r-1-k} \Big]\,. \tag{5.12}$$

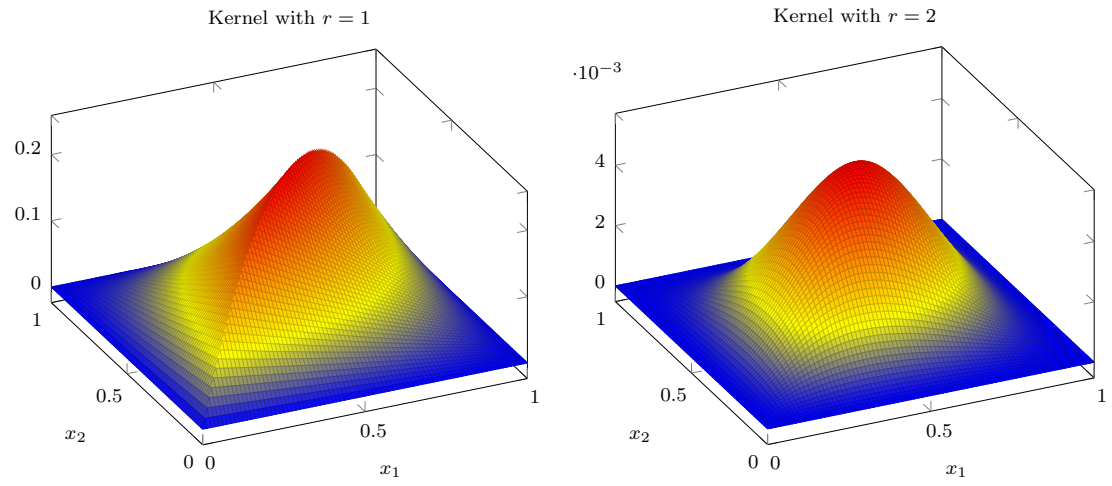


Figure 4: Plots of the kernel $\mathring{K}_1^r : [0,1] \times [0,1] \to \mathbb{R}$ with smoothness $r = 1$ (left) and smoothness $r = 2$ (right).

However, $\mathring{H}^r$ is a only a closed subspace of $\mathcal{H}_{K_1^r}$ since the functions $f \in \mathcal{H}_{K_1^r}$ may lack the right boundary condition which is $f^{(j)}(1) = 0$ if $j = 0, ..., r-1$, whereas the left boundary condition $f^{(j)}(0) = 0$ if $j = 0, ..., r-1$ is for free due to the construction. Let us now apply the construction from Lemma 5.3 to K_1^r to construct a reproducing kernel $\mathring{K}_1^r$ for the closed subspace $\mathring{H}^r$.

First we compute the functions $u_j(\cdot) = (K_1^r)^{(0,j)}(\cdot, 1)$ for $j = 0, ..., r-1$ explicitly. Using again the formula (5.10) we find

$$\begin{aligned} u_j(x) &= \Big(\frac{d}{dy}\Big)^j \int_0^{\min\{x,y\}} \frac{(x-t)_+^{r-1}}{(r-1)!} \cdot \frac{(y-t)_+^{r-1}}{(r-1)!}\, dt \Big|_{y=1} \\ &= \int_0^x \frac{(x-t)_+^{r-1}}{(r-1)!} \cdot \frac{(1-t)^{r-1-j}}{(r-1-j)!}\, dt\,, \end{aligned} \tag{5.13}$$

where we used the well-known formula for the differentiation of integrals. Similar as above in (5.11) we interpret this as a Taylor's remainder term for a specific polynomial. It is not hard to verify that this polynomial is given by

$$u_j(x) = \frac{(-1)^r}{(2r-1-j)!}\Big[\sum_{k=r}^{2r-1-j} \binom{2r-1-j}{k} (-x)^k\Big] \quad , \quad j = 0, ..., r-1\,. \tag{5.14}$$

Looking at the functions u_j, $j = 0, ..., r-1$, we see immediately that $\{x^r, ..., x^{2r-1}\}$ is a basis of their span. Hence we may use the system $\tilde{u}_j(x) = x^{j+r}/(j+r)!$ in (5.9). This gives the following representation for the kernel $\mathring{K}_1^r(x,y)$, namely

$$\mathring{K}_1^r(x,y) = K_1^r(x,y) - \sum_{j=0}^{r-1}\sum_{k=0}^{r-1} \frac{G_{j,k}^{-1}}{(j+r)!(k+r)!} x^{j+r} y^{k+r}\,, \tag{5.15}$$

where $K_1^r(x,y)$ is given by (5.12) and

$$\mathbf{G} = \Big(\frac{1}{j!k!(j+k+1)}\Big)_{\substack{j=0,\dots,r-1\\ k=0,\dots,r-1}}.$$

Let us give two examples. Putting $r=1$ in (5.15) we have

$$\mathring{K}_1^1(x,y) = \min\{x,y\} - xy, \quad , \quad x,y \in [0,1]\,.$$

Furthermore, in case $r=2$ we obtain

$$\mathring{K}_1^2(x,y) = K_1^2(x,y) - x^2y^2 + x^2y^3/2 + x^3y^2/2 - x^3y^3/3\,,$$

where

$$K_1^2(x,y) = \frac{1}{2}\min\{x,y\}^2 \max\{x,y\} - \frac{1}{6}\min\{x,y\}^3,$$

For $r=1,2,3$ we obtain the associated Gramian matrices

$$(\mathbf{G}^1)^{-1} = \big(1\big) \quad , \quad (\mathbf{G}^2)^{-1} = \begin{pmatrix} 4 & -6 \\ -6 & 12 \end{pmatrix} \quad , \quad (\mathbf{G}^3)^{-1} = \begin{pmatrix} 9 & -36 & 60 \\ -36 & 192 & -360 \\ 60 & -360 & 720 \end{pmatrix}.$$

In the case $d=1$ the kernels for $r=1$ and $r=2$ are depicted in Figure 4. The smoothness can be observed along the diagonal $x=y$, where the kernel for $r=1$ exhibits a kink.

Regarding the multivariate kernel, we have arrived at the following result.

Theorem 5.4. *Given a smoothness vector $\mathbf{r} = (r_1, r_2, \dots, r_d) \in \mathbb{N}^d$, the reproducing kernel of the tensor product space $\mathring{H}^{\mathbf{r}}_{\mathrm{mix}} = \mathring{H}^{r_1} \otimes \cdots \otimes \mathring{H}^{r_d}$ is given by*

$$\mathring{K}_d^{\mathbf{r}}(\boldsymbol{x},\boldsymbol{y}) = \prod_{\ell=1}^d \mathring{K}_1^{r_\ell}(x_\ell, y_\ell) \tag{5.16}$$

$$= \prod_{\ell=1}^d \Big(K_1^{r_\ell}(x_\ell,y_\ell) - \sum_{j=0}^{r_\ell-1}\sum_{k=0}^{r_\ell-1} (\mathbf{G}^{r_\ell})^{-1}_{j,k} (K_1^{r_\ell})^{(0,j)}(x_\ell,1)\,(K_1^{r_\ell})^{(0,k)}(y_\ell,1)\Big), \tag{5.17}$$

where $u_j(x_\ell) = (K_1^{r_\ell})^{(0,j)}(x_\ell,1)$ are given in (5.14) and $(\mathbf{G}^{r_\ell})^{-1}$ are given in (5.5).

The explicit expression for the reproducing kernel of $\mathring{H}^{\mathbf{r}}_{\mathrm{mix}}$ allows to compute the norms of arbitrary bounded linear functionals $L \in (\mathring{H}^{\mathbf{r}}_{\mathrm{mix}})^\star$, since it holds

$$\|L\|_{(\mathring{H}^{\mathbf{r}}_{\mathrm{mix}})^\star} = \sup_{\|f\|_{\mathring{H}^{\mathbf{r}}_{\mathrm{mix}}} \le 1} |L(f)| = \sqrt{L^{(\boldsymbol{x})} L^{(\boldsymbol{y})} \mathring{K}_d^{\mathbf{r}}(\boldsymbol{x},\boldsymbol{y})}. \tag{5.18}$$

The right-hand side involves the application of the functional L to both components of the kernel. We will use this in Section 7 for the simulation of worst-case integration errors which can be rewritten as norms of certain functionals (7.1) involving the integration functional $L(f) = I_d(f) = \int_{[0,1]^d} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$. In the sequel we will compute the norm and its Riesz representer. We have

$$\|I_d\|^2_{(\mathring{H}^{\mathbf{r}}_{\mathrm{mix}})^\star} = \sup_{\|f\|_{\mathring{H}^{\mathbf{r}}_{\mathrm{mix}}}\le 1} |I_d(f)|^2 = \prod_{\ell=1}^{d} \left(\int_0^1 \int_0^1 \mathring{K}_1^{r_\ell}(x_\ell, y_\ell)\,\mathrm{d}x_\ell\,\mathrm{d}y_\ell \right) \tag{5.19}$$

where

$$\begin{aligned}
\int_0^1 \int_0^1 \mathring{K}_1^{r}(x,y)\,\mathrm{d}x\,\mathrm{d}y &= \int_0^1 \int_0^1 K_1^{r}(x,y) - \sum_{j=0}^{r-1}\sum_{k=0}^{r-1} \frac{G^{-1}_{j,k}}{(j+r)!(k+r)!} x^{j+r} y^{k+r}\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_0^1 \int_0^1 K_1^{r}(x,y)\,\mathrm{d}x\,\mathrm{d}y - \sum_{j=0}^{r-1}\sum_{k=0}^{r-1} \frac{G^{-1}_{j,k}}{(j+r+1)!(k+r+1)!} \\
&= \frac{1}{(r!)^2(2r+1)} - \sum_{j=0}^{r-1}\sum_{k=0}^{r-1} \frac{G^{-1}_{j,k}}{(j+r+1)!(k+r+1)!}\,.
\end{aligned}$$

The last identity follows from the representation (5.11) and

$$\begin{aligned}
\int_0^1 \int_0^1 K_1^{r}(x,y)\,\mathrm{d}x\,\mathrm{d}y &= \int_0^1 \left(\int_0^1 \int_0^1 \frac{(x-t)_+^{r-1}}{(r-1)!} \frac{(y-t)_+^{r-1}}{(r-1)!}\,\mathrm{d}x\,\mathrm{d}y \right) \mathrm{d}t \\
&= \int_0^1 \left(\int_t^1 \frac{(x-t)^{r-1}}{(r-1)!} \int_t^1 \frac{(y-t)^{r-1}}{(r-1)!}\,\mathrm{d}x\,\mathrm{d}y \right) \mathrm{d}t \\
&= \int_0^1 \left(\frac{(1-t)^{r}}{r!} \frac{(1-t)^{r}}{r!} \right) \mathrm{d}t = \int_0^1 \frac{(1-t)^{2r}}{(r!)^2}\,\mathrm{d}t \\
&= \frac{1}{(2r)!(2r+1)}.
\end{aligned}$$

For the Riesz representer of $f \mapsto \int_{[0,1]^d} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \langle f, R_{I_d}\rangle_{\mathring{H}^{\mathbf{r}}_{\mathrm{mix}}}$ it holds

$$R_{I_d}(\boldsymbol{y}) = \int_{[0,1]^d} \mathring{K}^{\mathbf{r}}_d(\boldsymbol{x},\boldsymbol{y})\,\mathrm{d}\boldsymbol{x} = \prod_{\ell=1}^{d} \left(\int_0^1 \mathring{K}_1^{r_\ell}(x_\ell, y_\ell)\,\mathrm{d}x_\ell \right) \quad , \quad \boldsymbol{y} = (y_1, ..., y_d)\,.$$

Clearly, we have

$$\int_0^1 \mathring{K}_1^{r_\ell}(x,y)\,\mathrm{d}x = \int_0^1 K_1^{r_\ell}(x,y) - \sum_{j=0}^{r-1}\sum_{k=0}^{r-1} \frac{G^{-1}_{j,k}}{(j+r)!(k+r)!} x^{j+r} y^{k+r}\,\mathrm{d}x\,.$$

A similar computation as above together with the identity

$$\int_0^1 \frac{(y-t)_+^{r-1}}{(r-1)!} \frac{(1-t)^{r}}{r!}\,\mathrm{d}t = \frac{(-1)^r}{(2r)!} \sum_{k=r}^{2r} \binom{2r}{k} (-y)^k$$

(see the computation after (5.13)) leads to the following explicit formula

$$\int_0^1 \mathring{K}_1^{r_\ell}(x,y)\,\mathrm{d}x = \frac{(-1)^r}{(2r)!}\sum_{k=r}^{2r}\binom{2r}{k}(-y)^k - \sum_{j=0}^{r-1}\sum_{k=0}^{r-1}\frac{G^{-1}_{j,k}}{(j+r+1)!(k+r)!}y^{k+r}\,. \quad (5.20)$$

6. Worst-case error estimates with respect to $\mathring{H}^{\mathbf{r}}_{\mathbf{mix}}$

In this section, we are interested in the behavior of the worst-case error

$$e(n,d,\boldsymbol{r}) := \sup_{\|f\|_{\mathring{H}^{\boldsymbol{r}}_{\mathrm{mix}}}\le 1}\Big|Q_n^d(f) - I_d(f)\Big|\,. \quad (6.1)$$

of Frolov's cubature rule Q_n^d with respect to the unit ball in the norm $\|\cdot\|_{\mathring{H}^{\boldsymbol{r}}_{\mathrm{mix}}}$, see (5.5). Recall that

$$Q_n^d(f) \;=\; \frac{1}{n}\sum_{\boldsymbol{k}\in\mathbb{Z}^d} f(\boldsymbol{A}_n\boldsymbol{k}),$$

where $\boldsymbol{A}_n = n^{-1/d}\boldsymbol{A}$ and $\boldsymbol{A} = \big(\det(\boldsymbol{V})\big)^{-1/d}\boldsymbol{V}$ with $\boldsymbol{V}$ from Theorem 2.2. Let further $\boldsymbol{B}_n = (\boldsymbol{A}_n)^{-\top}$

The main tool for analyzing (6.1) is Poisson's summation formula. Let $\varphi \in S(\mathbb{R}^d)$ be a multivariate Schwartz-function. With $\mathcal{F}\varphi$ we denote the Fourier transform

$$\mathcal{F}\varphi(\xi) = \int_{-\infty}^{\infty}\varphi(x)\exp(-2\pi i\boldsymbol{x}\cdot\xi)\,d\boldsymbol{x} \quad , \quad \xi\in\mathbb{R}^d\,. \quad (6.2)$$

Then it holds

$$\sum_{\boldsymbol{m}\in\mathbb{Z}^d}\varphi(x+\boldsymbol{m}) = \sum_{\boldsymbol{k}\in\mathbb{Z}^d}\mathcal{F}\varphi(\boldsymbol{k})\exp(2\pi i\boldsymbol{k}\cdot\boldsymbol{x})$$

with absolute convergence on both sides. The following consequence is of particular importance. Let $\boldsymbol{A}:\mathbb{R}^d\to\mathbb{R}^d$ be a regular matrix with $\det\boldsymbol{A}\neq 0$. Let further $\boldsymbol{B} = \boldsymbol{A}^{-\top}$. Then we have

$$\det\boldsymbol{A}\sum_{\boldsymbol{m}\in\mathbb{Z}^d}\varphi(\boldsymbol{A}(\boldsymbol{x}+\boldsymbol{m})) = \sum_{\boldsymbol{k}\in\mathbb{Z}^d}\mathcal{F}\varphi(\boldsymbol{B}\boldsymbol{k})\exp(2\pi i\boldsymbol{k}\cdot\boldsymbol{x}) \quad (6.3)$$

Let us finally mention the following special case by putting $\boldsymbol{x}=0$

$$\det\boldsymbol{A}\sum_{\boldsymbol{m}\in\mathbb{Z}^d}\varphi(\boldsymbol{A}\boldsymbol{m}) = \sum_{\boldsymbol{k}\in\mathbb{Z}^d}\mathcal{F}\varphi(\boldsymbol{B}\boldsymbol{k}) \quad (6.4)$$

A more general variant (with respect to the regularity of the participating functions) can be found in [39, Thm. 3.1, Cor. 3.2]

In this section we show the by now well-known upper bounds on the worst-case error of Frolov's cubature formula for the Sobolev spaces $\mathring{H}^{\mathbf{r}}_{\mathrm{mix}}$. We give relatively short proofs

here with special emphasis on the constants. In particular, we will see how the invariants of the used lattice will affect the error estimates.

We will see that only two invariants will play a role in the upper bounds, which we want to discuss shortly. For this note that the lattices under consideration are generated by a multiple of a Vandermonde matrix $\boldsymbol{V}$, which is defined via a generating polynomial P as in Theorem 2.2. The first invariant is the determinant, or in other words the discriminant of the generating polynomial

$$D_P \,=\, \det(\boldsymbol{V}).$$

For example, we know from Theorem 2.2 that $\mathrm{Nm}(\boldsymbol{V}^{-\top}) = 1/D_P^2$.

The second invariant is

$$B_P \,:=\, \min_{\boldsymbol{U}} \|\boldsymbol{V}\boldsymbol{U}\|_\infty, \tag{6.5}$$

where the minimum is over all $\boldsymbol{U} \in \mathrm{SL}_d(\mathbb{Z})$. This constant is an upper bound for the diameter (in ℓ_∞) of the "smallest" fundamental cell of the lattice. To see this, note that every *fundamental cell*, i.e. a parallelepiped with corners on the lattice with no lattice point in the interior, is of the form $T([0,1]^d)$, where $T \in \mathbb{R}^{d\times d}$ is a generating matrix for the lattice. Moreover, it is well-known that every generating matrix of the lattice that is generated by $\boldsymbol{V}$ is of the form $\boldsymbol{V}\boldsymbol{U}$ for some unimodular, integer-valued matrix $\boldsymbol{U}$. We will see that both, D_P and B_P, should be small to obtain a small upper bound on the errors. This justifies the choice of the generating polynomials in the previous section. Here is the main result of this section.

Theorem 6.1. *Let $\boldsymbol{r} = (r_1, \dots, r_d) \in \mathbb{N}^d$ and $\eta := \#\{j \colon r_j = r\}$. Then we have for any $f \in \mathring{H}^{\boldsymbol{r}}_{\mathrm{mix}}$*

$$\begin{aligned}
&\left|Q_n^d(f) - I_d(f)\right| \\
&\leq\, C(d,\eta,\boldsymbol{r}) \cdot \max\left\{D_P, \frac{(2B_P)^d}{n}\right\}^{1/2} \left(\frac{D_P}{n}\right)^r \left(2 + \log\left(n/D_P\right)\right)^{(\eta-1)/2} \|f\|_{\mathring{H}^{\boldsymbol{r}}_{mix}},
\end{aligned} \tag{6.6}$$

where

$$\begin{aligned}
&C(d,\eta,\boldsymbol{r}) \\
&:=\, 2^{d+1} \left(1 - 2^{-2(r'-r)}\right)^{-(d-\eta)/2} \left(1 - 2^{(1-2r)}\right)^{-\eta/2} \left(\sum_{e\subset[d]} \prod_{i\in e} \frac{1}{[(r_i-1)!]^2(2r_i-1)2r_i}\right)^{1/2}
\end{aligned}$$

with $r' := \min_j\{r_j \colon r_j \neq r\}$.

Let us prove the following estimate first.

Proposition 6.2. *Let $\varphi \in C_0^\infty((0,1)^d)$. Then*

$$\Big|Q_n^d(\varphi) - I_d(\varphi)\Big| \le \sqrt{\frac{M(\boldsymbol{A}_n)}{\det \boldsymbol{B}_n}}\Big(\sum_{\boldsymbol{k}\in\mathbb{Z}^d\setminus\{\boldsymbol{0}\}} |v_{\boldsymbol{r}}(\boldsymbol{B}_n\boldsymbol{k})|^{-2}\Big)^{1/2} \|\varphi\|_{H^{\boldsymbol{r}}_{mix}}, \tag{6.7}$$

where

$$M(\boldsymbol{A}_n) := \min_{\boldsymbol{U}\in SL_d(\mathbb{Z})} \#\Big\{\boldsymbol{m}\in\mathbb{Z}^d \colon \boldsymbol{U}\boldsymbol{A}_n\big(\boldsymbol{m}+(0,1)^d\big)\cap[0,1]^d \neq \emptyset\Big\}, \tag{6.8}$$

is the minimal number of fundamental cells of the integration lattice necessary to cover the unit cube.

Proof. The above special case of Poisson's summation formula (6.4) gives

$$\begin{aligned}\Big|Q_n^d(\varphi) - I_d(\varphi)\Big| &= \Big|\sum_{\boldsymbol{k}\in\mathbb{Z}^d\setminus\{\boldsymbol{0}\}} \mathcal{F}\varphi(\boldsymbol{B}_n\boldsymbol{k})\Big| \\ &\le \Big(\sum_{\boldsymbol{k}\in\mathbb{Z}^d\setminus\{\boldsymbol{0}\}} v_{\mathbf{r}}(\boldsymbol{B}_n\boldsymbol{k})^{-2}\Big)^{1/2}\Big(\sum_{\boldsymbol{k}\in\mathbb{Z}^d} |v_{\mathbf{r}}(\boldsymbol{B}_n\boldsymbol{k})\mathcal{F}\varphi(\boldsymbol{B}_n\boldsymbol{k})|^2\Big)^{1/2}\end{aligned} \tag{6.9}$$

By the definition of $v_{\boldsymbol{r}}$ we may rewrite

$$|v_{\mathbf{r}}(\boldsymbol{B}_n\boldsymbol{k})\mathcal{F}\varphi(\boldsymbol{B}_n\boldsymbol{k})|^2 = \sum_{e\subset[d]} \Big|\mathcal{F}\Big[\Big(\prod_{i\in e}\frac{\partial^{r_i}}{\partial x_i^{r_i}}\Big)\varphi\Big](\boldsymbol{B}_n\boldsymbol{k})\Big|^2$$

Using this for the second factor in (6.9) we find

$$\begin{aligned}&\sum_{\boldsymbol{k}\in\mathbb{Z}^d} |v_{\mathbf{r}}(\boldsymbol{B}_n\boldsymbol{k})\mathcal{F}\varphi(\boldsymbol{B}_n\boldsymbol{k})|^2 \\ &\quad= \sum_{e\subset[d]}\int_{[0,1]^d} \Big|\sum_{\boldsymbol{k}\in\mathbb{Z}^d} \mathcal{F}\Big[\Big(\prod_{i\in e}\frac{\partial^{r_i}}{\partial x_i^{r_i}}\Big)\varphi\Big](\boldsymbol{B}_n\boldsymbol{k})\exp(2\pi i\boldsymbol{k}\cdot\boldsymbol{x})\Big|^2 d\boldsymbol{x}\,.\end{aligned}$$

Now we apply Poisson's summation formula in the form (6.3) to the integrand and find

$$\begin{aligned}&\sum_{\boldsymbol{k}\in\mathbb{Z}^d} |v_{\mathbf{r}}(\boldsymbol{B}_n\boldsymbol{k})\mathcal{F}\varphi(\boldsymbol{B}_n\boldsymbol{k})|^2 \\ &\quad= (\det\boldsymbol{A}_n)^2\sum_{e\subset[d]}\int_{[0,1]^d} \Big|\sum_{\boldsymbol{m}\in\mathbb{Z}^d} \Big[\Big(\prod_{i\in e}\frac{\partial^{r_i}}{\partial x_i^{r_i}}\Big)\varphi\Big](\boldsymbol{A}_n(\boldsymbol{x}+\boldsymbol{m}))\Big|^2 d\boldsymbol{x} \\ &\quad\le (\det\boldsymbol{A}_n)^2 M(\boldsymbol{A}_n)\sum_{e\subset[d]}\sum_{\boldsymbol{m}\in\mathbb{Z}^d}\int_{[0,1]^d} \Big|\Big[\Big(\prod_{i\in e}\frac{\partial^{r_i}}{\partial x_i^{r_i}}\Big)\varphi\Big](\boldsymbol{A}_n(\boldsymbol{x}+\boldsymbol{m}))\Big|^2 d\boldsymbol{x} \\ &\quad= (\det\boldsymbol{A}_n)M(\boldsymbol{A}_n)\sum_{e\subset[d]}\int_{\mathbb{R}^d} \Big|\Big[\Big(\prod_{i\in e}\frac{\partial^{r_i}}{\partial x_i^{r_i}}\Big)\varphi\Big](\boldsymbol{y})\Big|^2 d\boldsymbol{y} \\ &= \frac{M(\boldsymbol{A}_n)}{\det\boldsymbol{B}_n}\|\varphi\|^2_{H^{\boldsymbol{r}}_{\mathrm{mix}}},\end{aligned} \tag{6.10}$$

where we used Hölder's inequality and the fact that φ and all its partial derivatives have compact support in $(0,1)^d$ together with (6.8). □

Remark 6.3. Let us comment on the number $M(\boldsymbol{A}_n)$. Clearly, all the fundamental cells are contained in $[-L(n,P), 1+L(n,P)]^d$ with $L(n,P) := (D_P n)^{-1/d} B_P$ and B_P from (6.5). Here, we used that $\boldsymbol{A}_n = (D_P n)^{-1/d} \boldsymbol{V}$. Therefore, $M(\boldsymbol{A}_n)$ is bounded by the number of lattice points $\boldsymbol{A}_n(\mathbb{Z}^d)$ in this set. This number can be controlled by (6.11) below, which will be also of some importance later. For a proof see e.g. [37, Lem. 5]. In fact, for every axis-parallel box $\Omega \subset \mathbb{R}^d$ and every $T \in \mathbb{R}^{d\times d}$ we have

$$\# \left(\boldsymbol{T}(\mathbb{Z}^d) \cap \Omega \right) \,\leq\, \frac{\mathrm{vol}_d(\Omega)}{\mathrm{Nm}(\boldsymbol{T})} + 1. \tag{6.11}$$

With all the definitions from above and $\mathrm{Nm}(\boldsymbol{V}) = 1$, we obtain that

$$M(\boldsymbol{A}_n) \,\leq\, n\, D_P \left(1 + \frac{2B_P}{(D_P n)^{1/d}}\right)^d + 1 \,\leq\, n\, 2^d \max\left\{D_P,\, \frac{(2B_P)^d}{n}\right\}. \tag{6.12}$$

We see that the bound of the second factor of the above error bound depends asymptotically only on $\sqrt{D_P}$ (and the norm of f). However, for preasymptotic bounds also the term $B_P^{d/2}/\sqrt{n}$ plays an important role.

Proof. To finish the proof of Theorem 6.1 it remains to estimate the middle factor in (6.7). In fact, the statement (6.6) then follows by a straight-forward density argument recalling (5.3).

If $\boldsymbol{r} = (r, \dots, r)$ with $r \in \mathbb{N}_0$ is a constant smoothness vector, the following proof can be found in several articles, see e.g. [36] or [38, p. 580]. Note, that it also works for fractional $r > 1/2$, which is essentially shown in [39]. Although the optimal order of convergence is known also in the non-constant case, we were not able to find a proof with explicit constants. Therefore, we give it here. We assume without restriction that $r_1 = \cdots = r_\eta < r_{\eta+1} \leq \cdots \leq r_d$ for some $\eta \in \{1, \dots, d\}$.

First, for $\boldsymbol{m} = (m_1, \dots, m_d) \in \mathbb{N}_0^d$, we define the sets

$$\rho(\boldsymbol{m}) \,:=\, \{\boldsymbol{x} \in \mathbb{R}^d \colon \lfloor 2^{m_j - 1} \rfloor \leq |x_j| < 2^{m_j} \text{ for } j = 1, \dots, d\}.$$

Note that $\prod_{j=1}^d |x_j| < 2^{\|\boldsymbol{m}\|_1}$ for all $x \in \rho(\boldsymbol{m})$. Since $B_n = n^{1/d} B = (D_p n)^{1/d}\, \boldsymbol{V}^{-\top}$ we have

$$\mathrm{Nm}(B_n) \,=\, \inf_{\boldsymbol{k} \in \mathbb{Z}^d \setminus \{0\}} \prod_{j=1}^d (B_n \boldsymbol{k})_j \,=\, \frac{n}{D_p}.$$

This shows that $|(B_n(\mathbb{Z}^d) \setminus 0) \cap \rho(\boldsymbol{m})| = 0$ for all $\boldsymbol{m} \in \mathbb{N}_0^d$ with $\|\boldsymbol{m}\|_1 < R_n$, where

$$R_n \,:=\, \left\lceil \log_2\!\big(n/D_p\big) \right\rceil.$$

Moreover, for $B_n\boldsymbol{k} \in \rho(\boldsymbol{m})$, we have

$$\nu_{*,\boldsymbol{r}}(B_n\boldsymbol{k}) \,\geq\, \nu_{\boldsymbol{r}}(B_n\boldsymbol{k}) \,\geq\, \prod_{j=1}^{d} \max\{1, 2\pi\lfloor 2^{m_j-1}\rfloor\}^{r_j} \,\geq\, 2^{r_1m_1+\cdots+r_dm_d}.$$

Since $\rho(\boldsymbol{m})$ is a union of 2^d axis-parallel boxes each with volume less than $2^{\|\boldsymbol{m}\|_1}$, (6.11) implies that $\big|B_n(\mathbb{Z}^d)\cap\rho(\boldsymbol{m})\big| \leq 2^d(D_p2^{\|\boldsymbol{m}\|_1}/n+1) \leq 2^{d+2}2^{\|\boldsymbol{m}\|_1-R_n}$ if $\|\boldsymbol{m}\|_1 \geq R_n$. Additionally, note that $|\{\boldsymbol{m}\in\mathbb{N}_0^\eta\colon \|\boldsymbol{m}\|_1=\ell\}| = \binom{\eta-1+\ell}{\eta-1}$. With $r := r_1$ and $r' := r_{\eta+1}$, we obtain

$$\begin{aligned}
\sum_{\boldsymbol{k}\in\mathbb{Z}^d\setminus 0} |\nu_{*,\boldsymbol{r}}(B_n\boldsymbol{k})|^{-2} \,&\leq \sum_{m:\|m\|_1\geq R_n} \Big|B_n(\mathbb{Z}^d)\cap\rho(m)\Big|\, 2^{-2r_1m_1-\ldots-2r_dm_d}\\
&\leq\, 2^{d+2} \sum_{m:\|m\|_1\geq R_n} 2^{\|m\|_1-R_n}\, 2^{-2r(m_1+\ldots+m_\eta)-2r'(m_{\eta+1}+\ldots+m_d)}\\
&=\, 2^{d+2} \sum_{\ell=R_n}^{\infty} \sum_{m:\|m\|_1=\ell} 2^{\|m\|_1-R_n}\, 2^{-2r(m_1+\ldots+m_\eta)-2r'(m_{\eta+1}+\ldots+m_d)}\\
&=\, 2^{d+2} \sum_{\ell=R_n}^{\infty} \sum_{\ell'=0}^{\ell} \sum_{\substack{m_1,\ldots,m_\eta:\\ \sum_{j=1}^{\eta} m_j=\ell-\ell'}} \sum_{\substack{m_{\eta+1},\ldots,m_d:\\ \sum_{j=\eta+1}^{d} m_j=\ell'}} 2^{\ell-R_n}\, 2^{-2r(\ell-\ell')-2r'\ell'}\\
&=\, 2^{d+2} \sum_{\ell=R_n}^{\infty} \sum_{\ell'=0}^{\ell} \binom{\eta-1+\ell-\ell'}{\eta-1} \binom{d-\eta-1+\ell'}{d-\eta-1} 2^{\ell-R_n-2r\ell}\, 2^{-2(r'-r)\ell'}\\
&\leq\, 2^{d+2} \sum_{\ell=R_n}^{\infty} \binom{\eta-1+\ell}{\eta-1} 2^{\ell-R_n-2r\ell} \sum_{\ell'=0}^{\ell} \binom{d-\eta-1+\ell'}{d-\eta-1} 2^{-2(r'-r)\ell'}
\end{aligned}$$

In the last estimate we used that $\binom{k+\ell}{k} \leq \binom{k+\ell+1}{k}$ for every $k,\ell\in\mathbb{N}$. To bound the two sums above we use the well-known binomial identity

$$\sum_{\ell=0}^{\infty} \binom{D+\ell}{D} x^\ell \,=\, \frac{1}{(1-x)^{D+1}}$$

as well as the bound

$$\binom{D+\ell+R}{D} \,\leq\, \binom{D+\ell}{D} (1+R)^D$$

for $D,\ell,R\in\mathbb{N}$ and $x\in\mathbb{C}$ with $|x|<1$. We obtain for the second sum that

$$\sum_{\ell'=0}^{\ell} \binom{d-\eta-1+\ell'}{d-\eta-1} 2^{-2(r'-r)\ell'} \,\leq\, \Big(1-2^{-2(r'-r)}\Big)^{-(d-\eta)}$$

and for the first sum that

$$\begin{aligned}\sum_{\ell=R_n}^{\infty} \binom{\eta-1+\ell}{\eta-1} 2^{\ell-R_n-2r\ell} &= \sum_{\ell=0}^{\infty} \binom{\eta-1+\ell+R_n}{\eta-1} 2^{\ell-2r(\ell+R_n)} \\ &\le 2^{-2rR_n}\,(1+R_n)^{\eta-1} \sum_{\ell=0}^{\infty} \binom{\eta-1+\ell}{\eta-1} 2^{(1-2r)\ell} \\ &= 2^{-2rR_n}\,(1+R_n)^{\eta-1} \left(1-2^{(1-2r)}\right)^{-\eta}\end{aligned}$$

for $r > 1/2$. If we use $\log_2\bigl(n/D_p\bigr) \le R_n \le 1+\log_2\bigl(n/D_p\bigr)$ we finally obtain Theorem 6.1. □

7. Numerical results: Exact worst-case errors in $\mathring{H}^{\mathbf{r}}_{\mathrm{mix}}$

In Section 6 it has been shown that the Frolov method achieves the optimal rate of convergence in Sobolev spaces with both, uniform and anisotropic mixed smoothness. However, as we have seen in Section 3, there are different ways to choose the polynomials, which significantly influence the numerical performance. Therefore, even though the asymptotic convergence rate of all (admissible) Frolov cubature rules have the optimal order $\mathcal{O}(N^{-r}(\log N)^{(d-1)/2})$ for uniform smoothness f, there might be huge constants involved. In order to investigate the influence of different Frolov polynomials on the preasymptotic behavior of the integration error, we use a well-known technique for reproducing kernel Hilbert spaces to compute the worst-case error explicitly. This supplements the theoretical bounds from Section 6. Moreover, we compare the worst-case errors of Frolov cubature, the sparse grid method and quasi–Monte Carlo methods in $\mathring{H}^{r}_{\mathrm{mix}}$.

7.1. Exact worst-case errors via reproducing kernels

The worst-case error of any linear cubature rule $Q_N(f) = \sum_{i=1}^{N} w_i f(\boldsymbol{x}_i)$ with prescribed weights and nodes can be computed *exactly* via the norm of the error functional $R_N(f) := I_d(f) - Q_N(f)$, cf. Eq. (5.18). Applying R_N to both components of the kernel $\mathring{K}^{\mathbf{r}}_d(\boldsymbol{x},\boldsymbol{y})$, the well-known formula for the (absolute) worst-case error is obtained, i.e

$$\begin{aligned}\sup_{\|f\|_{\mathring{H}^{\mathbf{r}}_{\mathrm{mix}}}\le 1} |R_N(f)|^2 = &\int_{[0,1]^d}\int_{[0,1]^d} \mathring{K}^{\mathbf{r}}_d(\boldsymbol{x},\boldsymbol{y})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{y} - 2\sum_{i=1}^{N} w_i \int_{[0,1]^d} \mathring{K}^{\mathbf{r}}_d(\boldsymbol{x}_i,\boldsymbol{y})\,\mathrm{d}\boldsymbol{y} \\ &+ \sum_{i=1}^{N}\sum_{j=1}^{N} w_i w_j \mathring{K}^{\mathbf{r}}_d(\boldsymbol{x}_i,\boldsymbol{x}_j).\end{aligned} \tag{7.1}$$

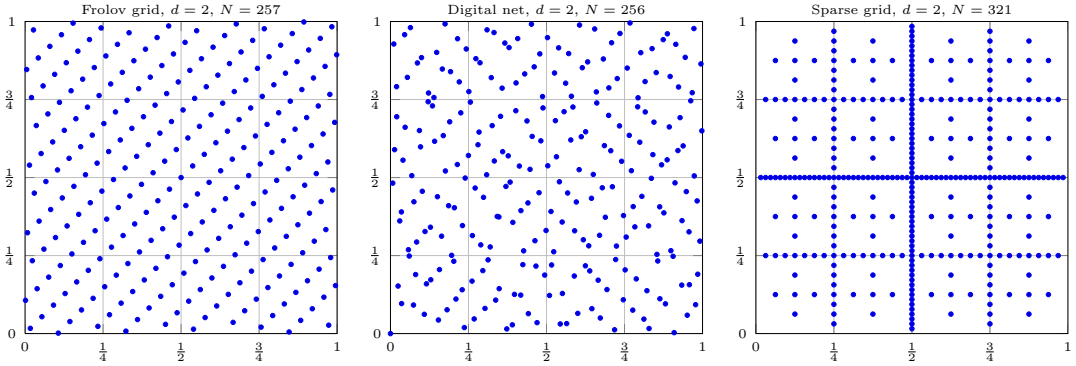


Figure 5: A Frolov lattice (left), an order-2 digital net (middle) and a zero boundary sparse grid (right).

Often, (7.1) is normalized with respect to norm of I_d in the dual-space $(\mathring{H}^{\boldsymbol{r}}_{\mathrm{mix}})^\star$, i.e. (7.1) is divided by $\|I_d\|_{(\mathring{H}^{\boldsymbol{r}}_{\mathrm{mix}})^\star} = (\int_{[0,1]^d}\int_{[0,1]^d} \mathring{K}^{\boldsymbol{r}}_d(\boldsymbol{x},\boldsymbol{y})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{y})^{1/2}$, cf. (5.19). The resulting quantity is called *normalized worst-case error*.

In order to evaluate (7.1) for an arbitrary given cubature rule we use the closed-form representation of the kernel $\mathring{K}^{\boldsymbol{r}}_d$ from Theorem 5.4 as well as the closed-form representation of the Riesz-representer (5.20).

Besides Frolov cubature rules, we will consider the sparse grid construction, which goes back to Smolyak [33], and also higher-order quasi–Monte Carlo integration [22]. Examples for the different point constructions are given in Figure 5. Their properties will be discussed below.

The Frolov points are generated using our newly developed Algorithm 1. The resulting points obtained by the improved polynomial construction can also be downloaded from `http://wissrech.ins.uni-bonn.de/research/software/frolov`.

7.2. Uniform mixed smoothness

As a first step we compare worst-case errors for cubature formulas that are known to work well in periodic Sobolev spaces, of which $\mathring{H}^r_{\mathrm{mix}}$ is a subset. These are different Frolov cubature rules, that are based on different choices of the generating polynomial. In the following, "Classical Frolov" will refer to the classical generating polynomial in (1.5), while "Improved Frolov" will refer to the lattices that are generated by the improved polynomials from Section 3. Moreover, we consider the sparse grid method that is based on the trapezoidal rule, see Appendix A. Due to the zero-boundary condition in $\mathring{H}^r$, all points with one component equal to zero are left out, cf. Figure 5.[1] It achieves a convergence rate of order $\mathcal{O}(N^{-r}(\log N)^{(d-1)(r+1/2)})$ in $\mathring{H}^r_{\mathrm{mix}}$, which is best possible

[1] This is similar to the open trapezoidal rule which, however, uses different weights, cf. [14].

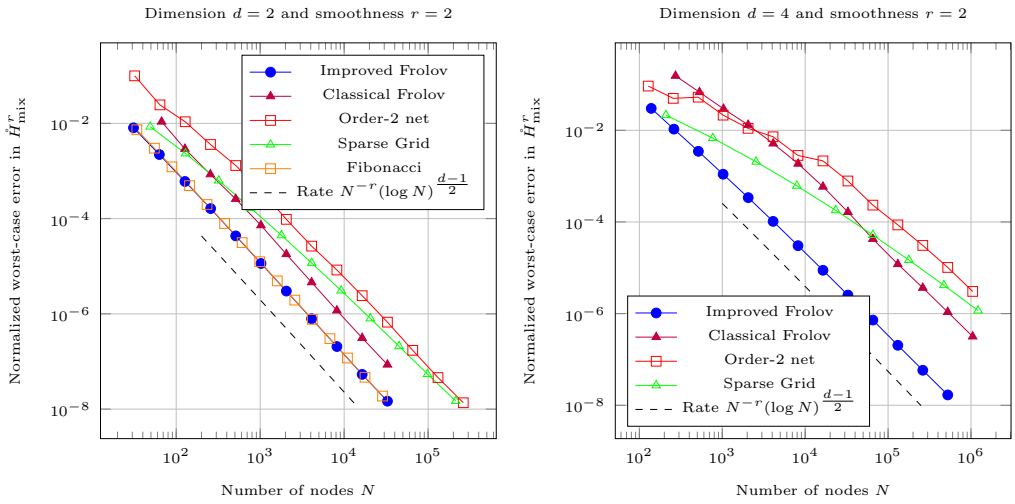


Figure 6: Worst-case errors for different cubature rules for uniform mixed smoothness $r = 2$ in dimensional $d = 2$ (left) and dimension $d = 4$ (right).

for a sparse grid method, cf. Theorem A.1 below. As an example for a higher order quasi–Monte Carlo method we use a digital net of order 2 that is obtained by interlacing the digits of a $(2d)$-dimensional Niederreiter-Xing net. This is obtained by using the implementation of Pirsic [29] of Xing-Niederreiter sequences [27] for rational places in dimension $2d-1$. These are known to yield smaller t-values than e.g. Sobol- or classical Niederreiter-sequences [10]. Then, a $2d$-dimensional digital net is obtained by employing the sequence-to-net propagation rule, cf. [22, 28] for more details. It is known that order-2 nets yield the optimal rate of convergence in periodic Sobolev spaces with bounded mixed derivatives of order $r < 2$, see [20] and also [17], since $\mathring{H}^r_{\mathrm{mix}} \subset H^r_{\mathrm{mix}}(\mathbb{T}^d)$.

Moreover, in the bivariate setting we also consider the Fibonacci lattice, which is not just known to be an order-optimal cubature rule for periodic Sobolev spaces with dominating mixed smoothness [8], but also represents the best possible point set for quasi – Monte Carlo intergation in this space, at least for small point numbers [21].

In the left-hand-side picture of Figure 6, the worst-case errors for smoothness $r = 2$ are computed in dimension $d = 2$. Clearly, the Frolov lattice based on the improved polynomial performs best in $\mathring{H}^r_{\mathrm{mix}}$. Of similar quality is the Fibonacci lattice and the classical Frolov lattice is slightly worse. The sparse grid also achieves the optimal main rate of N^{-r}, but it is known that the exponent of its logarithm is smoothness dependent. This is also apparent in Figure 6, where the sparse grid has an asymptotic behavior that is inferior to all the other considered methods. On the right-hand-side of Figure 6, the worst-case errors for smoothness $r = 2$ are computed in dimension $d = 4$. Here, the Fibonacci lattice is not considered. However, for all the other methods we note that the picture does not change much, compared to the case $d = 2$. As before, the improved Frolov method performs best and the classical Frolov obtains the same optimal

asymptotic convergence rate but a substantially worse constant. This effect is now much stronger than in the bivariate setting, i.e. the classical Frolov lattice has a worst-case error that is about two magnitudes larger than the one of the improved Frolov lattice. Moreover, the order-2 digital net seems to be competitive too, albeit with a substantially larger constant and longer pre-asymptotic regime. Again, the worse logarithmic exponent of the sparse grid method can be clearly observed.

In the Figures 8, 9 and 10 the influence of the dimensionality and the smoothness onto the performance of the Frolov cubature method is considered in more detail. As an example for a cubature method with a less than optimal complexity, the sparse grid method is also included. Especially the classical construction suffers from a strong growth of the constant as the dimensionality increases. Also, the pre-asymptotic regime seems to last longer. This effect can so far not be thoroughly explained by the existing theory. In dimension $d = 7$, the classical Frolov construction needs more than 10^6 points to achieve the error level of the zero-algorithm, i.e. normalized worst-case error 1. Note at this point, that all given errors are normalized worst-case errors, which can, for non optimally weighted cubature rules, be substantially larger than 1. It is apparent that the classical Frolov method is practically useless in dimension $d \geq 5$, due to its unfavorable pre-asymptotic behavior. Our new approach, however, shows a much better dependence onto the dimensionality and certainly allows the treatment of moderate-dimensional integrals from Sobolev spaces with dominating mixed smoothness of uniform type.

Moreover, we observe the universality of Frolov's method, i.e. without adaption to the respective parameters it achieves the best possible rate of convergence in every $\mathring{H}^r_{\mathrm{mix}}$, $r \in \{1, 2, 3\}$.

7.3. Anisotropic mixed smoothness

It has been shown in Theorem 6.1 that in Sobolev spaces with dominating mixed smoothness of different orders in each direction, only the lowest smoothness and associated dimension enters the error estimate. In order to make this phenomenon visible from a numerical perspective, we compute explicit worst-case errors in

$$\mathring{H}^{\boldsymbol{r}}_{\mathrm{mix}} = \mathring{H}^{r_1} \otimes \cdots \otimes \mathring{H}^{r_d},$$

where $r_1 = r$ and $r_2 = r_3 = \cdots = r_d = r+1$. Then, Theorem 6.1 predicts that the worst-case error asymptotically behaves like in the univariate setting, i.e. decays at a rate of $\mathcal{O}(N^{-r})$. The question that is investigated in Figure 7 is how long it takes to overcome the preasymptotic regime until this favorable convergence rate becomes visible.

On the left-hand-side of Figure 7, i.e. for $r = 1$, already with less than 3000 points the Frolov method follows the asymptotic regime of N^{-1} in all the considered cases $d \in \{2, 3, \ldots, 7\}$.

In contrast, on the right-hand-side of Figure 7, i.e. for $r = 2$, the dimension seems to have a much larger impact onto the length of the sub-optimal preasymptotic regime.

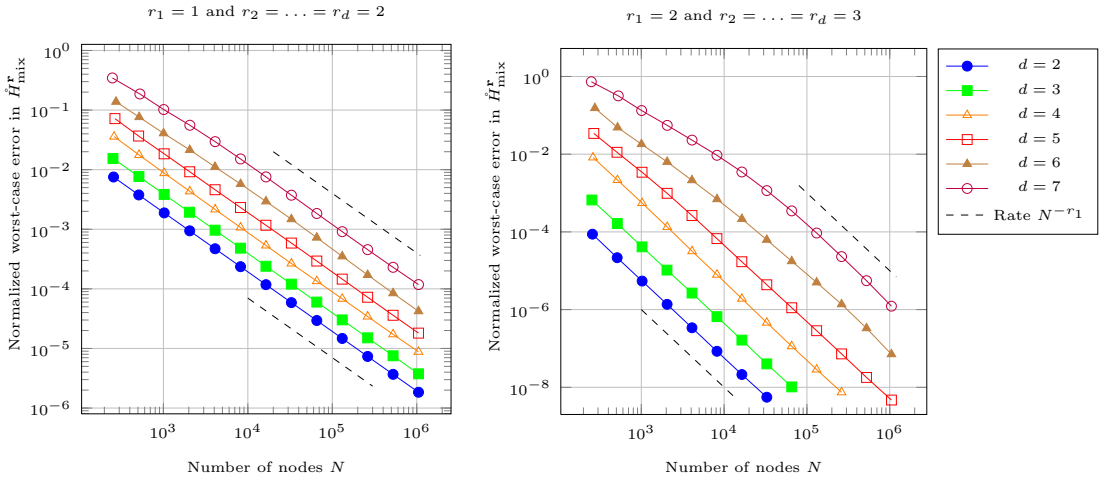


Figure 7: Worst-case errors for uniform mixed smoothness in various dimensions. Left-hand side: $r_1 = 1$ and $r_2 = \cdots = r_d = 2$. Left-hand side: $r_1 = 2$ and $r_2 = \cdots = r_d = 3$.

For example, in $d = 7$ the N^{-2}-rate becomes visible only when the number of points N is larger than $\approx 10^5$.

We remark that the sparse grid method is also able to deal with anisotropic mixed smoothness vectors $\boldsymbol{r} = (r_1, \ldots, r_d)$. Then, however, the construction needs to be adjusted to the smoothness vector which has to be known in advance, see [35, pp. 32,36,72], the recent survey [8, Sect. 10.1] and the references therein. The resulting sparse grid construction therefore is not a universal cubature formula.[2]

However, both plots in in Figure 7 were computed with the exact same set of Frolov points, which automatically benefit from the anisotropic smoothness that is present in a given integration problem, i.e. in this case $r = 1$ or $r = 2$. Therefore, it is not necessary to estimate the smoothness of the integrand and tune the method appropriately.

Acknowledgement

T.U. wishes to thank Winfried Bruns (Osnabrueck) for several fruitful discussions. T.U. and J.O. gratefully acknowledge support by the German Research Foundation (DFG) Ul-403/2-1, GR-1144/21-1 and the Emmy-Noether programme, Ul-403/1-1.

[2]Note that it is also possible to construct dimension-adaptive spare grids, which are are able to detect the smoothness vector in the process of approximation adaptively, cf. [15].

A. Appendix: Sparse grid cubature in $H^r_{\mathrm{mix}}(\mathbb{T}^d)$

Let

$$Q_N(f) := \sum_{j=0}^{N-1} \frac{1}{N} f\Big(\frac{j}{N}\Big) \tag{A.1}$$

denote the uniformly weighted N-point trapezoidal rule. It is known that it achieves the optimal rate of convergence N^{-r} in the periodic Sobolev space $H^r(\mathbb{T}), r \in \mathbb{N}$ (our proof below also works for the univariate case). In order to obtain a multivariate integration method we define the hierarchical quadrature rules

$$\Delta_k := \Delta_k(Q) := Q_{2^k} - Q_{2^{k-1}} \quad \text{for all } k = 1, 2, \dots \tag{A.2}$$

and $\Delta_0 = Q_1$. Their tensor product is denoted by $\Delta_{\boldsymbol{k}} := \bigotimes_{j=1}^d \Delta_{k_j}$, $\boldsymbol{k} \in \mathbb{N}_0^d$. The sparse grid cubature rule of level $L \in \mathbb{N}$ is then given by

$$Q_L^{sg} := \sum_{|\boldsymbol{k}|_1 \le L} \Delta_{\boldsymbol{k}}, \tag{A.3}$$

with multi-indices $\boldsymbol{k} = (k_1, \dots, k_d) \in \mathbb{N}_0^d$. The cubature rule Q_L^{sg} uses

$$N_L = \sum_{|\boldsymbol{k}|_1 \le L} \prod_{j=1}^d 2^{k_j - 1} = \mathcal{O}(2^L \cdot L^{d-1}) \tag{A.4}$$

function values combined with non-equal weights. The following theorem gives the well-known error bound in $H^r_{\mathrm{mix}}(\mathbb{T}^d)$. For the convenience of the reader we will also give a proof.

Theorem A.1. *Consider the sparse grid cubature rule Q_L^{sg} as it is defined in* (A.2) *and* (A.3) *based on the univariate trapezoidal rule* (A.1)*. The worst-case integration error of Q_L^{sg} in $H^r_{\mathrm{mix}}(\mathbb{T}^d)$ can be bounded by*

$$\sup_{\|f\|_{H^r_{\mathrm{mix}}} \le 1} \left| \int_{[0,1]^d} f(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} - Q_L^{sg}(f) \right| \asymp N^{-r} (\log N)^{(d-1)(r+1/2)}, \tag{A.5}$$

where $N = N_L$ denotes the number of points used by Q_L^{sg}.

Proof. The lower bound follows from [9, Thm. 5.2]. Note, that the lower bound also holds true for the smaller space $\mathring{H}^r_{\mathrm{mix}}(\mathbb{T}^d)$ since the constructed fooling functions also belong to this space. For the upper bound we use the detour to sampling recovery. In the recent paper [6, Thm. 4.7, 4.8, 5.13, 5.14] it has been observed that nested trigonometric interpolation operators

$$I_{2^k}[f](x) = \frac{1}{2^k} \sum_{u=0}^{2^k-1} f\Big(\frac{u}{2^k}\Big) \mathcal{D}^1_{2^k}\Big(x - \frac{u}{2^k}\Big) \quad , \quad k = 0, 1, 2, ..., \tag{A.6}$$

based upon the modified (nested) Dirichlet kernel $\mathcal{D}^1_{2^k}(x) := \mathcal{D}_{2^{k-1}}(x) - e^{2\pi i 2^{k-1}x}$ may be used to characterize $H^r_{\mathrm{mix}}(\mathbb{T}^d)$. In fact, the tensor products $\Delta_{\boldsymbol{k}}(I)$, $\boldsymbol{k} \in \mathbb{N}_0^d$, are defined analogously to (A.2) using this time (A.6) (note that $\mathcal{D}^1_1(x) \equiv 1$). Then we have

$$\|f\|^2_{H^r_{\mathrm{mix}}} \asymp \sum_{\boldsymbol{k}\in\mathbb{N}_0^d} 2^{2r|\boldsymbol{k}|_1} \|\Delta_{\boldsymbol{k}}(I)[f]\|_2^2 \,. \tag{A.7}$$

See also [5, Prop. 3.3] for the classical (non-nested) trigonometric interpolation. The associated sparse grid interpolation operator I^{sg}_L is defined in the same way as above in (A.3). Now we argue similar as in [5, Thm. 5.4]. Indeed, Hölder's inequality together with (A.7) gives

$$\begin{aligned} \|f - I^{sg}_L[f]\|_2 &\le \Big(\sum_{|\boldsymbol{k}|_1 > L} 2^{-2|\boldsymbol{k}|_1 r} \Big)^{1/2} \cdot \Big(\sum_{|\boldsymbol{k}|_1 > L} 2^{2r|\boldsymbol{k}|_1} \|\Delta_{\boldsymbol{k}}(I)[f]\|_2^2 \Big)^{1/2} \\ &\le 2^{-rL} L^{(d-1)/2} \|f\|_{H^r_{\mathrm{mix}}} \,. \end{aligned} \tag{A.8}$$

Noting further that

$$Q^{sg}_L(f) = \int_{[0,1]^d} I^{sg}_L[f](\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}$$

we have by Hölder's inequality and (A.8)

$$\Big| \int_{[0,1]^d} f(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} - Q^{sg}_L(f) \Big| \le \|f - I^{sg}_L[f]\|_2 \le 2^{-rL} L^{(d-1)/2} \|f\|_{H^r_{\mathrm{mix}}} \,,$$

see also [8, Rem. 8.9] . Finally, the bound (A.5) follows from (A.4) . □

Remark A.2. The above multivariate cubature rule on the sparse grid uses a number of nodes on the boundary of $[0,1]^d$ which are not needed when dealing with functions from $\mathring{H}^r_{\mathrm{mix}} \subset H^r_{\mathrm{mix}}(\mathbb{T}^d)$. However, as already mentioned in the proof of Theorem A.1, with respect to the asymptotic rate of convergence we can not do essentially better. However, to do a fair cost comparison for the different methods considered in Section 7 we only counted the interior nodes (see the diagrams above, e.g. Figure 6).

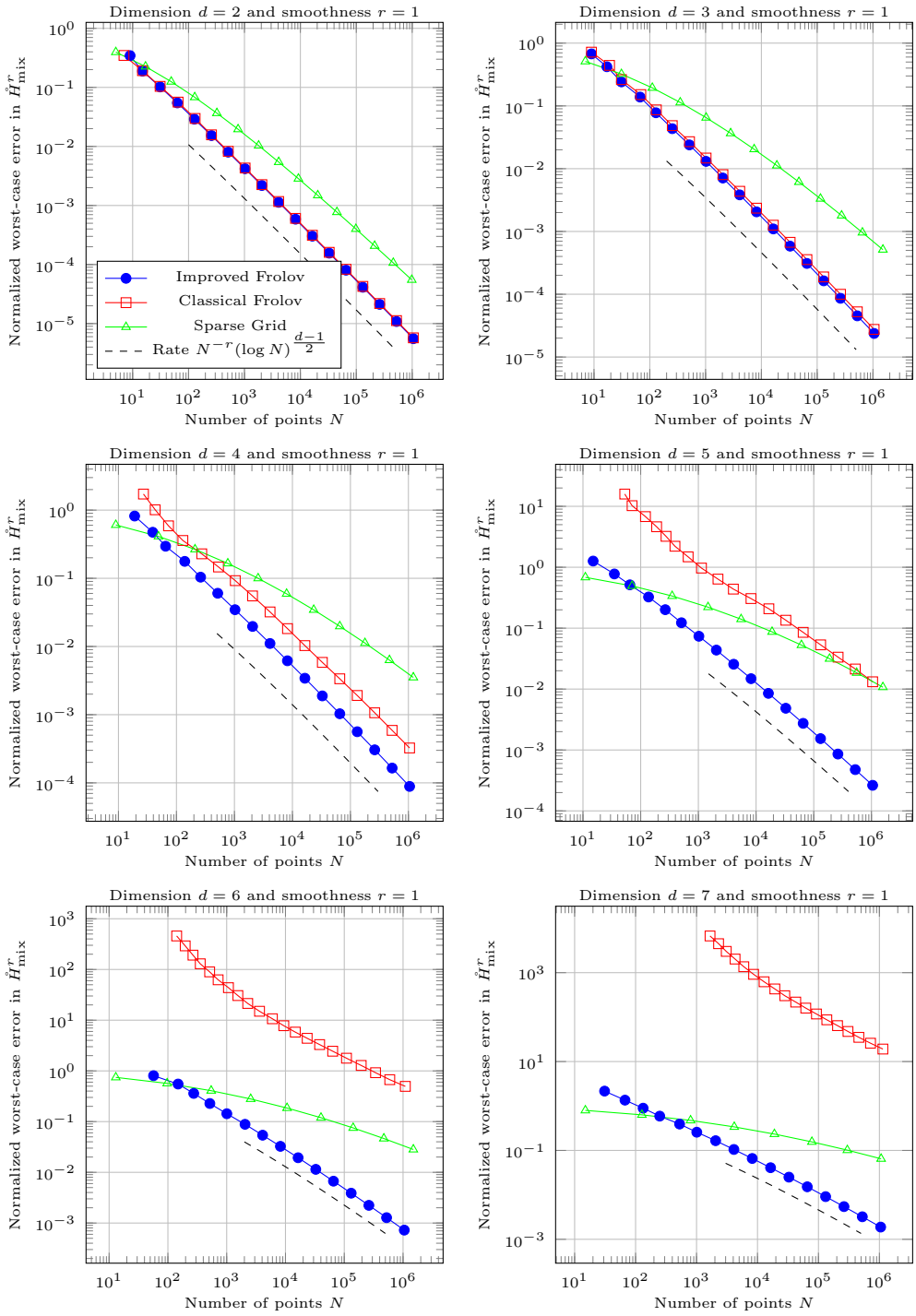


Figure 8: Normalized worst-case errors for uniform smoothness parameter $r = 1$ in dimension $d \in \{2, 3, 4, 5, 6, 7\}$ for different Frolov constructions and sparse grids.

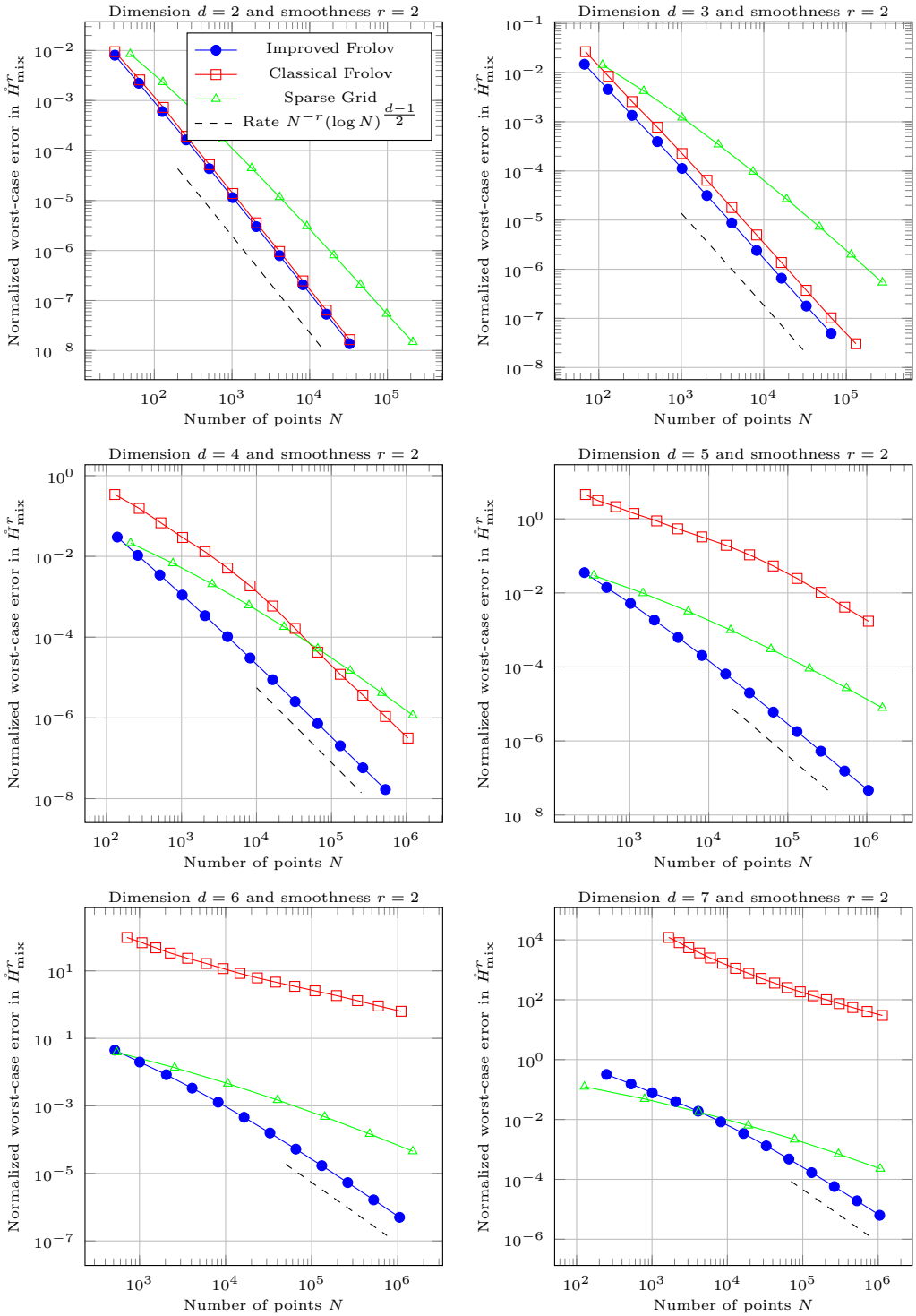


Figure 9: Normalized worst-case errors for uniform smoothness parameter $r = 2$ in dimension $d \in \{2, 3, 4, 5, 6, 7\}$ for different Frolov constructions and sparse grids.

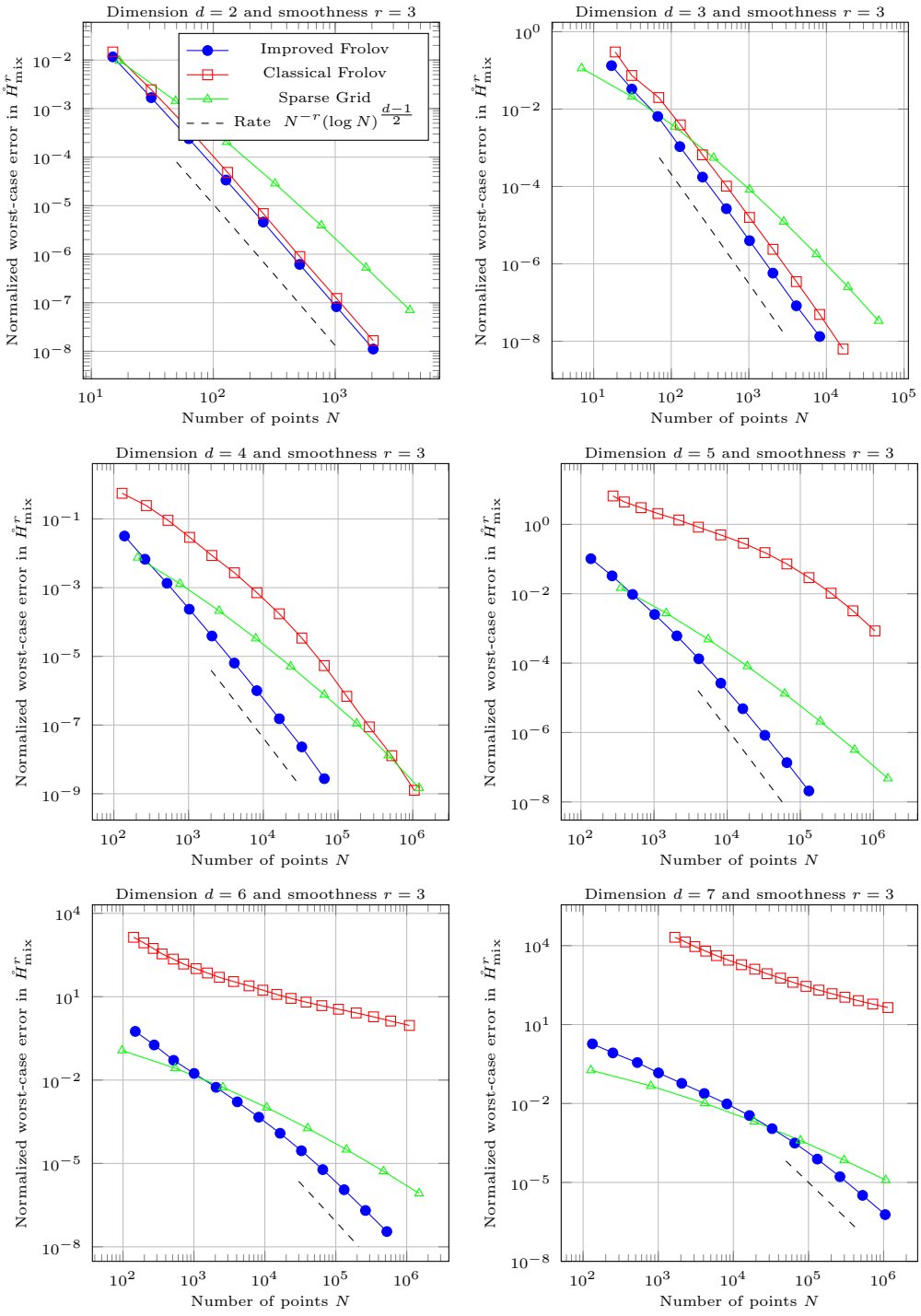


Figure 10: Normalized worst-case errors for uniform smoothness parameter $r = 3$ in dimension $d \in \{2, 3, 4, 5, 6, 7\}$ for different Frolov constructions and sparse grids.

References

[1] N. Aronszajn. Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 68(3):337–404, 1950.

[2] H. Avron, V. Sindhwani, J. Yang, and M. Mahoney. Quasi-monte carlo feature maps for shift-invariant kernels. *Journal of Machine Learning Research*, 17(120):1–38, 2016.

[3] A. Berlinet and C. Thomas-Agnan. *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Springer, 2004.

[4] K. Binder and D. Heermann. *Monte Carlo Simulation in Statistical Physics: An Introduction*. Springer-Verlag Berlin Heidelberg, 2010.

[5] G. Byrenheid, D. Dũng, W. Sickel, and T. Ullrich. Sampling on energy-norm based sparse grids for the optimal recovery of Sobolev type functions in H^γ. *J. Approx. Theory*, 207:207–231, 2016.

[6] G. Byrenheid and T. Ullrich. Optimal sampling recovery of mixed order Sobolev embeddings via discrete Littlewood-Paley type characterizations. *Anal. Math.*, 43(2):133–191, 2017.

[7] W.-C. Chang, C.-L. Li, Y. Yang, and B. Póczos. Data-driven random Fourier features using Stein effect. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, pages 1497–1503, 2017.

[8] D. Dũng, V.N. Temlyakov, and T. Ullrich. Hyperbolic Cross Approximation. *Advanced Courses in Mathematics. CRM Barcelona. Birkhäuser/Springer*, to appear.

[9] D. Dũng and T. Ullrich. Lower bounds for the integration error for multivariate functions with mixed smoothness and optimal Fibonacci cubature for functions on the square. *Math. Nachr.*, 288(7):743–762, 2015.

[10] J. Dick and H. Niederreiter. On the exact t-value of Niederreiter and Sobol' sequences. *Journal of Complexity*, 24(5–6):572 – 581, 2008.

[11] V. V. Dubinin. Cubature formulas for classes of functions with bounded mixed difference. *Mat. Sb.*, 183(7):23–34, 1992.

[12] V. V. Dubinin. Cubature formulas for Besov classes. *Izv. Ross. Akad. Nauk Ser. Mat.*, 61(2):27–52, 1997.

[13] K. K. Frolov. Upper bounds for the errors of quadrature formulae on classes of functions. *Dokl. Akad. Nauk SSSR*, 231(4):818–821, 1976.

[14] T. Gerstner and M. Griebel. Numerical integration using sparse grids. *Numer. Algorithms*, 18:209–232, 1998.

[15] T. Gerstner and M. Griebel. Dimension–adaptive tensor–product quadrature. *Computing*, 71(1):65–87, 2003.

[16] P. Glasserman. *Monte Carlo Methods in Financial Engineering*. Springer, 2003.

[17] T. Goda, K. Suzuki, and T. Yoshiki. An explicit construction of optimal order quasi–Monte Carlo rules for smooth integrands. *SIAM J. Numer. Anal*, 54(5):2664–2683, 2016.

[18] C. Gouriéroux and A. Monfort. *Simulation-Based Econometric Methods*. Oxford University Press, 1997.

[19] P. M. Gruber and C. G. Lekkerkerker. *Geometry of numbers*, volume 37 of *North-Holland Mathematical Library*. North-Holland Publishing Co., Amsterdam, second edition, 1987.

[20] A. Hinrichs, L. Markhasin, J. Oettershagen, and T. Ullrich. Optimal quasi-Monte Carlo rules on higher order digital nets for the numerical integration of multivariate periodic functions. *Numerische Mathematik*, 134(1):163–196, 2016.

[21] A. Hinrichs and J. Oettershagen. Optimal point sets for quasi-Monte Carlo integration of bivariate periodic functions with bounded mixed derivatives. In Ronald Cools and Dirk Nuyens, editors, *Monte Carlo and Quasi-Monte Carlo Methods: MCQMC, Leuven, Belgium, April 2014*, pages 385–405. Springer International Publishing, 2016.

[22] J. Dick and F. Pillichshammer. *Digital nets and sequences. Discrepancy theory and quasi-Monte Carlo integration*. Cambridge University Press, Cambridge, 2010.

[23] C. Kacwin. Realization of the Frolov cubature formula via orthogonal Chebyshev-Frolov lattices. Masterarbeit, Institut für Numerische Simulation, Universität Bonn, 2016.

[24] C. Kacwin, J. Oettershagen, and T. Ullrich. On the orthogonality of the Chebyshev-Frolov lattice and applications. *Monatsh. Math.*, 184(3):425–441, 2017.

[25] A. K. Lenstra, H. W. Lenstra, and L. Lovász. Factoring polynomials with rational coefficients. *Mathematische Annalen*, 261(4):515–534, 1982.

[26] D. A. Marcus. *Number Fields (Universitext)*. Springer, 1995.

[27] H. Niederreiter and C. Xing. A construction of low-discrepancy sequences using global function fields. *Acta Arithmetica*, 73(1):87–102, 1995.

[28] H. Niederreiter and C. Xing. Low-discrepancy sequences and global function fields with many rational places. *Finite Fields and Their Applications*, 2(3):241–273, 1996.

[29] G. Pirsic. A software implementation of Niederreiter-Xing sequences. In *Monte Carlo and quasi-Monte Carlo methods 2000*. Springer, Berlin, 2002.

[30] A. Rahimi and B. Recht. Random features for large-scale kernel machines. In J. C. Platt, D. Koller, Y. Singer, and S. T. Roweis, editors, *Advances in Neural Information Processing Systems 20*, pages 1177–1184. Curran Associates, Inc., 2008.

[31] M. O. Rayes, V. Trevisan, and P. S. Wang. Factorization of Chebyshev polynomials, 1998.

[32] M. M. Skriganov. Constructions of uniform distributions in terms of geometry of numbers. *Algebra i Analiz*, 6(3):200–230, 1994.

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

[34] K. Suzuki and T. Yoshiki. Enumeration of the Chebyshev-Frolov lattice points in axis-parallel boxes. arXiv:1612.05342.

[35] V.N. Temlyakov. Approximation of functions with bounded mixed derivative. *Proc. Steklov Inst. Math.*, (1(178)):vi+121, 1989. A translation of Trudy Mat. Inst. Steklov **1**78 (1986), Translated by H. H. McFaden.

[36] V.N. Temlyakov. *Approximation of periodic functions*. Computational Mathematics and Analysis Series. Nova Science Publishers, Inc., Commack, NY, 1993.

[37] M. Ullrich. A Monte Carlo method for integration of multivariate smooth functions. *ArXiv e-prints*, 2016. arXiv:1604.06008 [math.NA].

[38] M. Ullrich. On "Upper error bounds for quadrature formulas on function classes" by K. K. Frolov. *Springer Proc. Math. Stat, Series Monte Carlo and quasi-Monte Carlo methods*, 163:571–582, 2016.

[39] M. Ullrich and T. Ullrich. The role of Frolov's cubature formula for functions with bounded mixed derivative. *SIAM J. Numer. Anal.*, 54(2):969–993, 2016.

[40] G. Wahba. *Spline models for observational data*, volume 59 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1990.