Institut für Numerische Simulation
Rheinische Friedrich-Wilhelms-Universität Bonn

Wegelerstraße 6 • 53115 Bonn • Germany
phone +49 228 73-3427 • fax +49 228 73-7527
www.ins.uni-bonn.de

C. Rieger and H. Wendland

Sampling Inequalities for Sparse Grids

INS Preprint No. 1609

2015

Noname manuscript No.
(will be inserted by the editor)

Sampling Inequalities for Sparse Grids

Christian Rieger · Holger Wendland

Received: date / Accepted: date

Abstract Sampling inequalities play an important role in deriving error estimates for various reconstruction processes. They provide quantitative estimates on a Sobolev norm of a function, defined on a bounded domain, in terms of a discrete norm of the function's sampled values and a smoothness term which vanishes if the sampling points become dense. The density measure, which is typically used to express these estimates, is the mesh norm or Hausdorff distance of the discrete points to the bounded domain. Such a density measure intrinsically suffers from the curse of dimension. The curse of dimension can be circumvented, at least to a certain extend, by considering additional structures. Here, we will focus on bounded mixed regularity. In this situation sparse grid constructions have been proven to overcome the curse of dimension to a certain extend

In this paper, we will concentrate on a special construction for such sparse grids, namely Smolyak's method and provide sampling inequalities for mixed regularity functions on such sparse grids in terms of the number of points in the sparse grid. Finally, we will give some applications of these sampling inequalities.

Keywords High dimensional approximation · sparse grids · sampling inequalities · smoothing Splines

Christian Rieger
Institute for Numerical Simulation
University of Bonn
53115 Bonn
Germany
E-mail: rieger@ins.uni-bonn.de

Holger Wendland
University of Bayreuth
95440 Bayreuth
Germany
E-mail: holger.wendland@uni-bayreuth.de

1 Introduction

Sampling inequalities provide an efficient, theoretical tool to prove convergence results for a large class of numerical reconstruction processes, see for example [13, 22, 12, 1]. They are based on the observation that a sufficiently regular function that is small on a sufficiently dense discrete set has to be globally small. However, the derived estimates suffer from the so-called curse of dimension. To make this more precise, let us consider a typical variant of a sampling inequality for a function $u \in W_p^k(\Omega)$ on a bounded Lipschitz domain $\Omega \subseteq \mathbb{R}^d$. Here, $W_p^k(\Omega)$ denotes the usual Sobolev space of smoothness k, where the smoothness is measured in the $L_p(\Omega)$-norm. A typical sampling inequality provides an estimate of the form

$$\|u\|_{L_\infty(\Omega)} \leq C h_{X_N,\Omega}^{k-d/p} |u|_{W_p^k(\Omega)} + \|u|_{X_N}\|_{\ell_\infty(X_N)}, \tag{1}$$

where the quantity

$$h_{X_N,\Omega} := h_{X,\Omega} := \sup_{x\in\Omega} \min_{x_j \in X} \|x - x_j\|_{\ell_2} \tag{2}$$

is called the fill distance of the discrete set $X_N = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\} \subseteq \Omega$. To illustrate the curse of dimension, we will apply the sampling inequality (1) to a norm-minimal interpolation process, i.e. we assume that the function u from (1) is of the form $u = f - I_{X_N} f$, where $I_{X_N} : W_2^k(\Omega) \to W_2^k(\Omega)$ satisfies

$$I_{X_N} f|_{X_N} = f|_{X_N} \quad \text{and} \quad \|I_{X_N} f\|_{W_2^k(\Omega)} \leq c\|f\|_{W_2^k(\Omega)}$$

with $c > 0$ being a constant. For $k > d/2$, the existence of such stable interpolation processes is for example given if interpolation by radial basis functions or other kernel-based methods is employed, see for example [21].

For *quasi-uniform* point sets, the number of points N is related to the fill distance via $h_{X_N,\Omega} \sim N^{-\frac{1}{d}}$, see [21, Proposition 14.1]. Thus we can derive

$$\begin{aligned}\|f - I_{X_N} f\|_{L_\infty(\Omega)} &\leq C h_{X_N,\Omega}^{k-d/2} \|f\|_{W_2^k(\Omega)} \\ &\leq C N^{-\frac{k}{d}+\frac{1}{2}} \|f\|_{W_2^k(\Omega)}.\end{aligned}$$

Obviously, a large dimension slows the error bound dramatically down. Even worse, the smoothness k must satisfy $k > d/2$ to guarantee convergence at all. This is due to the well-known Sobolev embedding theorem $W_2^k(\Omega) \subseteq C(\Omega)$, which requires $k > d/2$ and is necessary for interpolation. While this assumption is not too restrictive in moderate dimensions, say $1 \leq d \leq 4$, it becomes unacceptable for higher dimensions.

The first step to remedy this problem is to replace the smoothness assumption $u \in W_2^k(\Omega)$ by something more restricting. In this context, typically tensor product spaces are considered, i.e. Sobolev spaces of mixed regularity. However, this approach is then limited to tensor product domains. Typically, one is mainly concerned with the d dimensional hypercube, i.e.,

$$I^d := I^{(1)} \times \cdots \times I^{(d)} \quad \text{with} \quad I^{(j)} = I = [-1, 1]. \tag{3}$$

As smoothness spaces, we will consider tensor product spaces, i.e.,

$$W_p^{k;\otimes^d}(I^d) := \bigotimes_{j=1}^{d} W_p^k(I) = \{f \in L_p(I^d) : D^{\boldsymbol{\alpha}} f \in L_p(I^d), \|\boldsymbol{\alpha}\|_\infty \leq k\}, \quad (4)$$

equipped with the norm

$$\|f\|^p_{W_p^{k;\otimes^d}(I^d)} := \sum_{0 \leq |\boldsymbol{\alpha}| \leq k} \|D^{\boldsymbol{\alpha}} f\|^p_{L_p(I^d)}.$$

Here, $k \in \mathbb{N}_0$ is an integer and $1 \leq p < \infty$ but we will also allow a fractional order of smoothness $\tau > 0$ or $p = \infty$, using, for example interpolation between Sobolev spaces.

Because of the tensor product structure, obviously the Sobolev embedding theorem now becomes

$$W_2^{k;\otimes^d}(I^d) \subseteq C(I^d), \quad \text{for} \quad k > \frac{1}{2}.$$

Hence, here the dependence on the spatial dimension is removed.

If applied to residuals of stable reconstruction processes, sampling inequalities imply error estimates that can cope with noisy data (see [14]). Sampling inequalities have proven useful in various applications, e.g. spline smoothing (see [22]), support vector regression algorithms (see [15,7]), and integral equations (see [11, 10]).

So far, all these sampling inequlities are built using the above mentioned fill distance. For higher dimensional spaces this is, because of the reasons outlined above, not practicable. Hence, in high dimension sampling inequalities will only make sense if they use the number of points instead of the fill distance. It is the goal of this paper, to derive such sampling inequalities for specific, so-called *sparse grids*, which play a prominent role in numerical methods for high dimensional problems, see for example [3,8]. For specific interpolation methods on sparse grids over the torus, error estimates have previously been derived in [19,4].

The paper is organised as follows. In the next section, we will introduce Smolyak's algorithm to construct high dimensional approximations from one dimensional ones. Intrinsic to this algorithm are sparse grids, which we also introduce here, particularly, those which we will use. Also in Section 2, we will discuss different methods of polynomial reproductions, in particular those, which use oversampling to have a constant Lebesgue function.

Section 3 is devoted to deriving sampling inequalities on sparse grids and in Section 4 we will give two possible applications: Interpolation and penalised least-squares approximation with tensor product reproducing kernels in mixed-derivative Sobolev spaces.

2 Smolyak's Method and Polynomial Reproductions

2.1 Smolyak's Algorithm

In this paper, we will use a general construction, given by Smolyak [18]. Smolyak provided a construction technique that used univariate operators to build a multivariate operator on a sparse grid with approximately the same convergence properties as the univariate operators.

To be more precise, suppose we are given a sequence of point sets $X^{(i)} := \{x_1^{(i)}, \dots, x_{m_i}^{(i)}\} \subseteq I^{(i)}$ for $i \in \mathbb{N}$. Suppose further that for each i, we can define an operator

$$\mathcal{U}(X^{(i)}) = \mathcal{U}^{(i)} : C(I^{(i)}) \to C(I^{(i)}), \qquad f \mapsto \sum_{j=1}^{m_i} a_j^{(i)}(x) f(x_j^{(i)})$$

based on $X^{(i)}$ and given by certain functions $a_j^{(i)} : I^{(i)} \to \mathbb{R}$ for $j = 1 \dots, m_i$. Following [2,17], these univariate formulas give rise to the tensor product formula

$$\begin{aligned}(\mathcal{U}^{(i_1)} \otimes \cdots \otimes \mathcal{U}^{(i_d)})(f) &= \sum_{j_1=1}^{m_{i_1}} \cdots \sum_{j_d=1}^{m_{i_d}} f(x_{j_1}^{(i_1)}, \dots, x_{j_d}^{(i_d)}) a_{j_1}^{(i_1)} \otimes \cdots \otimes a_{j_d}^{(i_d)} \\ &=: \sum_{1 \le \mathbf{j} \le \mathbf{m}} f(\mathbf{x}_{\mathbf{j}}^{(\mathbf{i})}) \mathbf{a}_{\mathbf{j}}^{(\mathbf{i})},\end{aligned}$$

where we used the vector notation $\mathbf{j} \le \mathbf{m}$ for vectors $\mathbf{j} = (j_1, \dots, j_d)^T$ and $\mathbf{m} = (m_1, \dots, m_d)^T$, which is defined to mean $j_k \le m_k$ for all $1 \le k \le d$. Moreover, for a scalar $v \in \mathbb{R}$ we define $v \le \mathbf{m}$ to be $\mathbf{v} \le \mathbf{m}$ with $\mathbf{v} = (v, \dots, v)^T$. Finally, we used the notation

$$\mathbf{a}_{\mathbf{j}}^{(\mathbf{i})} := a_{j_1}^{(i_1)} \otimes \cdots \otimes a_{j_d}^{(i_d)}$$

for a tensor product function and

$$\mathbf{x}_{\mathbf{j}}^{(\mathbf{i})} = (x_{j_1}^{(i_1)}, \dots, x_{j_d}^{(i_d)})$$

for a vector. It should be clear from the context, which one is meant.

These formulas are the main building blocks for Smolyak's algorithm. For a given $q \in \mathbb{N}$ with $q \ge d$, Smolyak's algorithm is given by

$$\begin{aligned}A(q,d)(f) &:= \sum_{\substack{q-d+1 \le |\mathbf{i}| \le q \\ \mathbf{i} \in \mathbb{N}^d}} (-1)^{q-|\mathbf{i}|} \binom{d-1}{q-|\mathbf{i}|} (\mathcal{U}^{(i_1)} \otimes \cdots \otimes \mathcal{U}^{(i_d)})(f) \\ &= \sum_{\substack{q-d+1 \le |\mathbf{i}| \le q \\ \mathbf{i} \in \mathbb{N}^d}} (-1)^{q-|\mathbf{i}|} \binom{d-1}{q-|\mathbf{i}|} \sum_{1 \le \mathbf{j} \le \mathbf{m}} f(\mathbf{x}_{\mathbf{j}}^{(\mathbf{i})}) \mathbf{a}_{\mathbf{j}}^{(\mathbf{i})}.\end{aligned}$$

Here, we used the notation $|\mathbf{i}| = i_1 + \cdots + i_d$ for $\mathbf{i} \in \mathbb{N}^d$. Note that $q \ge d$ only means $|\mathbf{i}| \ge 1$. Hence, unless $q \ge 2d-1$, some terms in the sum might still be zero since $\mathbf{i} \in \mathbb{N}^d$ means automatically $|\mathbf{i}| \ge d$.

To evaluate $A(q,d)(f)$, we only need to know f at the *sparse grid*

$$H(q,d) := \bigcup_{\substack{q-d+1 \le |\mathbf{i}| \le q \\ \mathbf{i} \in \mathbb{N}^d}} (X^{(i_1)} \times \cdots \times X^{(i_d)}).$$

If the sets are nested, i.e. $X^{(i)} \subseteq X^{(i+1)}$ this reduces to

$$H(q,d) := \bigcup_{|\mathbf{i}|=q} (X^{(i_1)} \times \cdots \times X^{(i_d)}). \tag{5}$$

In this case, we have the following reproduction result from [2, Lemma 2]:

Lemma 1 *Assume that the formulas $\mathcal{U}^{(i)}$ are exact on the linear spaces $\mathcal{V}^{(i)} \subseteq C(I)$ with $I = [-1, 1]$ and $\mathcal{V}^{(i)} \subseteq \mathcal{V}^{(i+1)}$ for $i \in \mathbb{N}$. Then Smolyak's algorithm $A(q, d)$ is exact on*

$$\sum_{\substack{|\mathbf{i}|=q \\ \mathbf{i}\in\mathbb{N}^d}} (\mathcal{V}^{(i_1)} \otimes \cdots \otimes \mathcal{V}^{(i_d)}).$$

To ensure nestedness we again follow [2] and consider the Clenshaw-Curtis or Chebyshev point sets.

Definition 1 We define a sequence of numbers $m_1 = 1$ and

$$m_i = 2^{i-1} + 1, \qquad i > 1. \tag{6}$$

Associated with these numbers we define the Clenshaw-Curtis point sets to be

$$X_{m_i} := X^{(i)} = \left\{ x_j^{(i)} = -\cos\left(\frac{\pi(j-1)}{m_i - 1}\right) : 1 \le j \le m_i \right\}, \qquad i > 1, \tag{7}$$

and $X_{m_1} = X^{(1)} = \{0\}$.

It is easy to see that for the given choices of m_i the sets $X^{(i)}$ are indeed nested, i.e. $X^{(i)} \subseteq X^{(i+1)}$ for all $i \ge 1$.

In what follows, we are mainly interested in polynomials and polynomial interpolation. Let $\pi_k(\mathbb{R}^d)$ denote the set of all d-variate polynomials of degree less than or equal to k and let $\pi_k(\Omega) := \pi_k(\mathbb{R}^d)|_\Omega$ be its restriction to $\Omega \subseteq \mathbb{R}^d$.

Let us assume that the operators $\mathcal{U}^{(i)}$ are exact on $\pi_{m_i-1}(I^{(i)})$, then, Lemma 1 shows that Smolyak's algorithm $A(q, d)$ is exact on the *Smolyak polynomial space*

$$\mathcal{S}_{q,d}(I^d) := \sum_{|\mathbf{i}|=q} \left(\pi_{m_{i_1}-1}(I) \otimes \cdots \otimes \Pi_{m_{i_d}-1}(I) \right).$$

2.2 Polynomial Reproduction

We will heavily rely on the concept of norming sets and polynomial reproductions. The following description is essentially taken from [21, Section 3.2], but originates from [9].

Let V be a finite dimensional vector space with norm $\|\cdot\|_V$ and dual V^*. Let $Z = \{z_1, \ldots, z_N\} \subseteq V^*$ be a finite set of linearly independent functionals. These functionals define the *sampling operator*

$$T : V \to T(V) \subseteq \mathbb{R}^N, \qquad T(v) = (z_1(v), \ldots, z_N(v))^T.$$

If the sampling operator T is injective, then the set of functionals Z is called a *norming set* for V. If $Z \subseteq V^*$ is a norming set for V, then $\|\cdot\|_V$ and $\|T(\cdot)\|_{\mathbb{R}^N}$ are equivalent norms on the finite dimensional space V. We will use the following result (see [21, Theorem 3.4]).

Theorem 1 *Suppose V is a finite dimensional normed linear space and $Z = \{z_1, \ldots, z_N\}$ is a norming set for V with associated sampling operator T. For every $\psi \in V^*$ there exists a vector $u \in \mathbb{R}^N$ such that*

– $\psi(v) = \sum_{j=1}^{N} u_j z_j(v)$ *for every* $v \in V$,
– $\|u\|_{\mathbb{R}^{N^*}} \le \|\psi\|_{V^*} \|T^{-1}\|_{T(V)\to V}$.

The number $\|T^{-1}\|_{T(V)\to V}$ *is called the* norming constant *of the set* Z.

We will apply this result to polynomials and point evaluations in the univariate setting, i.e. we will use $V = \pi_k(I) = \pi_k(\mathbb{R})|_I$ with $I = [-1,1]$ and $Z = \{\delta_{x_1}, \ldots, \delta_{x_n}\}$, where δ_x for $x \in I$ denotes the point evaluation functional, i.e. $\delta_x(p) = p(x)$.

In what follows we will particularly be interested in studying the relationship between the number n of data points and the degree k of the polynomials.

We will start with the well-known case of $k = n-1$, which gives rise to unique interpolation.

2.3 Univariate Polynomial Reproduction without Oversampling

For a given set $X = \{x_1, \ldots, x_k\} \subseteq I = [-1,1]$ we can write an interpolation operator $I_X : C(I) \to \pi_{k-1}(I)$ as

$$I_X f(x) = \sum_{j=1}^{k} f(x_j) L_j(x), \qquad x \in I,$$

using the Lagrange functions

$$L_j(x) = \prod_{\substack{i=1 \\ i\neq j}}^{k} \frac{x - x_i}{x_j - x_i}, \qquad x \in I.$$

Since this operator is exact on $\pi_{k-1}(I)$ and has norm

$$\Lambda_{k-1}(X) := \max_{x\in I} \sum_{j=1}^{k} |L_j(x)|,$$

we have the first example of a polynomial reproduction.

Theorem 2 *A set* $X = X_k = \{x_1, \ldots, x_k\} \subset I = [-1,1]$ *defines a polynomial reproduction for* $\pi_{k-1}(I)$, *i.e. there are functions* $a_j : [-1,1] \to \mathbb{R}$ *for* $1 \le j \le k$ *such that*

$$\sum_{j=1}^{k} a_j(x) p(x_j) = p(x)$$

for all $p \in \pi_{k-1}(I)$ *and* $x \in I$. *Furthermore, we have*

$$\sum_{j=1}^{k} |a_j(x)| \le \Lambda_{k-1}(X_k) := \max_{x\in I} \sum_{j=1}^{k} |L_j(x)|.$$

Proof We simply set $a_j = L_j$.

The number $\Lambda_{k-1}(X)$ is called the *Lebesgue constant* of this interpolation process. It is well studied in univariate polynomial interpolation. In case of the points defined in (7), the Lebesgue constant is known [2, Equation 8] to satisfy

$$\Lambda_{k-1}(X_k) \le \frac{2}{\pi} \log(k-1) + 1. \tag{8}$$

2.4 Univariate Reproduction with Oversampling

From now on, we will concentrate mainly on the sets $X = X^{(i)}$, introduced in (7). Our focus will, however, be on oversampling, i.e. we will use more points $k = m_i$ than usually required for an exact reproduction of polynomials of degree n.

In this section we will concentrate on the sets $X^{(i)}$ because they allow for the smallest amount of oversampling. Similar arguments concerning the oversampling in the spherical distance can be found in [16].

Definition 2 The spherical fill distance for a point set $X \subseteq I = [-1, 1]$ on I is defined by

$$d_{X,I} := \max_{x\in I} \min_{x_j\in X} |\arccos(x) - \arccos(x_j)|.$$

With this spherical fill distance we can formulate and prove the following theorem.

Theorem 3 *A set $X = \{x_1, \dots, x_k\} \subseteq I = [-1, 1]$ defines a norming set $\Delta(X) := \{\delta_{x_j} : x_j \in X\} \subseteq C(I)^*$ for $\pi_n(I)$, if the oversampling condition $C_\Delta := nd_{X,I} < 1$ is satisfied. In this case, the norming constant is given by $C_X := \frac{1}{1-C_\Delta}$.*

Proof We consider the sampling operator $T : \pi_n(I) \to \mathbb{R}^k$ defined by $Tp = (p(x_1), \dots, p(x_k))^T$.

According to Lemma 3 (ii) in Section 7 of Chapter 3 of [5] we immediately have

$$\frac{\|Tp\|_{\ell_\infty(\mathbb{R}^k)}}{\|p\|_{L_\infty(I)}} \geq 1 - nd_{X_k,I} = 1 - C_\Delta > 0$$

for all $p \in \pi_n(I)$. Hence, we have for the norm of the sampling operator T the estimate

$$\|T\| \geq 1 - C_\Delta > 0.$$

This obviously means that T is injective and that we have an inverse $T^{-1} : T(\pi_n(I)) \subseteq \mathbb{R}^k \to \pi_n(I)$ with

$$\|T^{-1}\| \leq \frac{1}{1 - C_\Delta}.$$

Next, we will check when the Clenshaw-Curtis points satisfy the oversampling condition $nd_{X,I} < 1$. We will do this first for slightly more general points.

Lemma 2 *Let $k \in \mathbb{N}$ with $k \geq 2$. On $I = [-1, 1]$, the point set*

$$X_k := \left\{ x_j = -\cos\left(\pi \frac{j-1}{k-1}\right) : 1 \leq j \leq k \right\}$$

has a spherical fill distance

$$d_{X_k,I} := \max_{x\in I} \min_{x_j\in X_k} |\arccos(x) - \arccos(x_j)| = \frac{\pi}{2(k-1)}.$$

Moreover, the points are uniformly distributed with respect to the distance d.

Proof We consider $\arccos : [-1,1] \to [-\pi, 0]$. With this notation we can compute

$$\begin{aligned}\arccos(x_j) &= \arccos\left(-\cos\left(\pi\frac{j-1}{k-1}\right)\right) = \arccos\left(\cos\left(-\pi+\pi\frac{j-1}{k-1}\right)\right) \\ &= \arccos\left(\cos\left(\pi\frac{j-k}{k-1}\right)\right) = \pi\frac{j-k}{k-1},\end{aligned}$$

since $\pi(j-k)/(k-1) \in [-\pi, 0]$. Hence we see that

$$\widetilde{X}_k := \arccos(X_k) = \{\arccos(x_j) : x_j \in X_k\} = \left\{\pi\frac{j-k}{k-1} : 1 \le j \le k\right\} \subseteq [-\pi, 0]$$

is an equidistant grid in $[-\pi, 0]$ having an Euclidean fill distance $h_{\widetilde{X}_k,[-\pi,0]} = d_{X_k,I}$. Thus, we can compute

$$\begin{aligned}d_{X_k,I} &= h_{\widetilde{X}_k,[-\pi,0]} = \max_{x\in[-\pi,0]} \min_{x_j\in\widetilde{X}_k} = \frac{1}{2}\left|x_1 - x_2\right| \\ &= \frac{1}{2}\left|-\pi-\pi\frac{2-k}{k-1}\right| = \frac{\pi}{2}\left|1+\frac{2-k}{k-1}\right| = \frac{\pi}{2}\left|\frac{1}{k-1}\right| = \frac{\pi}{2(k-1)},\end{aligned}$$

which is the stated equality.

We are now going to apply this result to the previously defined meshes $X^{(i)} = X_{m_i}$ from Definition 1. To be more precise, let $\ell > 0$ and consider $X = X^{(i+\ell)}$. Following Lemma 2, this set has a spherical fill distance

$$d_{X^{(i+\ell)},I} = \frac{\pi}{2(m_{i+\ell}-1)} = \frac{\pi}{2(2^{i+\ell-1}+1-1)} = \pi 2^{-(i+\ell)}.$$

According to Theorem 3, the associated functionals $\Delta(X^{(i+\ell)}) := \{\delta_x : x \in X^{(i+\ell)}\}$ form a norming set for $\pi_n(I)$ with $n = m_i - 1$ provided that $C_{\Delta(X^{(i+\ell)})} = n d_{X^{(i+\ell)},I} < 1$. Since we have

$$C_{\Delta(X^{(i+\ell)})} = n d_{X^{(i+\ell)},I} = \frac{(m_i-1)\pi}{2^{i+\ell}} = \pi 2^{i-1-i-\ell} = \pi 2^{-\ell-1},$$

which is less than one for every $\ell \ge 1$, we have indeed a norming set with norming constant

$$\begin{aligned}C_{X^{(i+\ell)}} &= \frac{1}{1-C_{\Delta(X^{(i+\ell)})}} = \frac{1}{1-\frac{\pi}{2^{\ell+1}}} = \frac{2^{\ell+1}}{2^{\ell+1}-\pi} \\ &= 1 + \frac{\pi}{2^{\ell+1}}\frac{1}{1-\frac{\pi}{2^{\ell+1}}} \\ &\le 1 + \frac{\pi}{2^{\ell+1}}\frac{4}{4-\pi} \\ &\le 1 + \frac{5\pi}{2^{\ell+1}}.\end{aligned}$$

In the last step we used that $4/(4-\pi) \approx 4.659792368 \le 5$.

This, together with Theorem 1, immediately gives the following result.

Theorem 4 *Let $X^{(i+\ell)}$ be the set defined in (7) with $i \in \{1,\dots,d\}$ and $\ell \in \mathbb{N}$. Then, the associated functionals $\Delta(X^{(i+\ell)})$ form a norming set for $\pi_{m_i-1}(I)$. In particular, there are functions $a_j^{(i+\ell)} : I \to \mathbb{R}$ for $1 \le j \le m_{i+\ell}$ such that*

$$\sum_{j=1}^{m_{i+\ell}} a_j^{(i+\ell)}(x) p(x_j^{(i+\ell)}) = p(x)$$

for all $p \in \pi_{m_i-1}(I)$ and $x \in I$. Moreover, we have

$$\sum_{j=1}^{m_{i+\ell}} |a_j^{(i+\ell)}(x)| \le \frac{2^{\ell+1}}{2^{\ell+1}-\pi} \le 1 + \frac{5\pi}{2^{\ell+1}}.$$

Note that the norm of the interpolation operator can become arbitrarily close to 1, if we let ℓ tend to infinity. However, this means that we will use an increasing number of data sites, while we keep the polynomial degree fixed.

2.5 Analysis of Smolyak's Algorithm with and without Oversampling

We can combine Lemma 1 with Theorem 3 and Theorem 4, respectively, to derive the following Theorem on Smolyak's algorithm. Note that now Smolyak's algorithm also depends on $\ell \ge 0$ since we will use the univariate grids $X^{(i+\ell)}$ and the associated functions $a^{(i+\ell)}$ to build $A(q,d)$. However, we will not change the notation to indicate this dependence on ℓ.

We will use the notation $\mathbf{m} + \ell := (m_1 + \ell, \dots m_d + \ell)^T$.

Theorem 5 *Define Smolyak's algorithm $A(q,d)$ using the functions $a_j^{(i+\ell)}$ from Theorem 2 or Theorem 4. Then, the algorithm is exact on $\mathcal{S}_{q,d}(I^d)$, i.e. we have*

$$p(\mathbf{x}) = A(q,d)p(\mathbf{x}) = \sum_{\substack{q-d+1 \le |\mathbf{i}| \le q \\ \mathbf{i} \in \mathbb{N}^d}} (-1)^{q-|\mathbf{i}|} \binom{d-1}{q-|\mathbf{i}|} \sum_{\mathbf{j} \le \mathbf{m}+\ell} p(\mathbf{x}_\mathbf{j}^{(\mathbf{i}+\ell)}) \mathbf{a}_\mathbf{j}^{(\mathbf{i}+\ell)}(\mathbf{x})$$

for all $p \in \mathcal{S}_{q,d}(I^d)$ and $\mathbf{x} \in I^d$. Moreover, in the case of Theorem 2, i.e. $\ell = 0$ (no oversampling), we have the bound

$$\|A(q,d)\| \le c(q,d) \left(\frac{2}{\pi} \log(n(q,d) - 1) + 1 \right)^d,$$

where $n(q,d)$ equals the number of points in the sparse grid $H(q,d)$ and

$$c(q,d) = 2^{d-1} \binom{q-1}{d-1}.$$

Moreover, in the case of Theorem 4, i.e. $\ell \ge 1$ (oversampling) we have

$$\|A(q,d)\| \le c(q,d) \left(1 + \frac{5\pi}{2^{\ell+1}} \right)^d.$$

Proof The first part is a direct consequence of Lemma 1 combined with Theorem 2 and Theorem 4, respectively. We apply them with $\ell \geq 0$ and

$$\begin{aligned}\mathcal{V}^{(i)} &= \pi_{m_i-1}(I),\\ \mathcal{U}^{(i)} &: p \mapsto \sum_{j=1}^{m_{i+\ell}} a_j^{(i+\ell)}(x)p(x_j^{(i+\ell)}).\end{aligned}$$

Then, we know by Theorem 2 and Theorem 4, respectively, that $\mathcal{U}^{(i)}$ is exact on $\pi_{m_i-1}(I)$, i.e. $\mathcal{U}^{(i)}(p) = p$ for all $p \in \pi_{m_i-1}(I)$. Hence, Lemma 1 yields the first claim. For bounding the norm of the operator, we use

$$\begin{aligned}&\|A(q,d)\|\\ &= \sup_{\|f\|_{L_\infty(I^d)}=1} \max_{\mathbf{x}\in I^d} \left| \sum_{\substack{q-d+1\leq|\mathbf{i}|\leq q\\ \mathbf{i}\in\mathbb{N}^d}} (-1)^{q-|\mathbf{i}|} \binom{d-1}{q-|\mathbf{i}|} \sum_{\mathbf{j}\leq\mathbf{m}+\ell} f(\mathbf{x}_\mathbf{j}^{(\mathbf{i}+\ell)})\mathbf{a}_\mathbf{j}^{(\mathbf{i}+\ell)}(\mathbf{x}) \right|\\ &\leq \max_{\mathbf{x}\in I^d} \sum_{\substack{q-d+1\leq|\mathbf{i}|\leq q\\ \mathbf{i}\in\mathbb{N}^d}} \binom{d-1}{q-|\mathbf{i}|} \sum_{1\leq\mathbf{j}\leq\mathbf{m}+\ell} |\mathbf{a}_\mathbf{j}^{(\mathbf{i}+\ell)}(\mathbf{x})|.\end{aligned}$$

Next, using A.2 from [17] and the fact that $j \mapsto \binom{j-1}{d-1}$ is non-decreasing, yields

$$\begin{aligned}\sum_{\substack{q-d+1\leq|\mathbf{i}|\leq q\\ \mathbf{i}\in\mathbb{N}^d}} \binom{d-1}{q-|\mathbf{i}|} &= \sum_{j=q-d+1}^{q} \sum_{\substack{|\mathbf{i}|=j\\ \mathbf{i}\in\mathbb{N}^d}} \binom{d-1}{q-j} = \sum_{j=q-d+1}^{q} \binom{d-1}{q-j} \sum_{\substack{|\mathbf{i}|=j\\ \mathbf{i}\in\mathbb{N}^d}} 1\\ &= \sum_{j=q-d+1}^{q} \binom{d-1}{q-j}\binom{j-1}{d-1} \leq \sum_{j=q-d+1}^{q} \binom{d-1}{q-j}\binom{q-1}{d-1}\\ &= \sum_{k=0}^{d-1} \binom{d-1}{k}\binom{q-1}{d-1} = 2^{d-1}\binom{q-1}{d-1}\\ &=: c(q,d).\end{aligned}$$

With this, we can derive the estimate

$$\begin{aligned}&\sum_{q-d+1\leq|\mathbf{i}|\leq q} \binom{d-1}{q-|\mathbf{i}|} \sum_{\mathbf{j}\leq\mathbf{m}+\ell} |\mathbf{a}_\mathbf{j}^{(\mathbf{i}+\ell)}(\mathbf{x})|\\ &= \sum_{q-d+1\leq|\mathbf{i}|\leq q} \binom{d-1}{q-|\mathbf{i}|} \sum_{\mathbf{j}\leq\mathbf{m}+\ell} |a_{j_1}^{(i_1+\ell)} \otimes \cdots \otimes a_{j_d}^{(i_d+\ell)}(\mathbf{x})|\\ &\leq \sum_{q-d+1\leq|\mathbf{i}|\leq q} \binom{d-1}{q-|\mathbf{i}|} \prod_{k=1}^{d} \sum_{j_k=1}^{m_{i_k}+\ell} |a_{j_k}^{(i_k+\ell)}(x_k)|.\end{aligned}$$

In the case of $\ell = 0$ we can use (8) to continue

$$\begin{aligned}
&\sum_{q-d+1\le|\mathbf{i}|\le q}\binom{d-1}{q-|\mathbf{i}|}\prod_{k=1}^{d}\sum_{j_k=1}^{m_{i_k+\ell}}|a_{j_k}^{(i_k+\ell)}(x_k)|\\
&\le \sum_{q-d+1\le|\mathbf{i}|\le q}\binom{d-1}{q-|\mathbf{i}|}\prod_{k=1}^{d}\left(\frac{2}{\pi}\log(m_k-1)+1\right)\\
&\le c(q,d)\left(\frac{2}{\pi}\log(n(q,d)-1)+1\right)^d.
\end{aligned}$$

In the case of $\ell \ge 1$ we can use Theorem 4 to derive

$$\begin{aligned}
&\sum_{q-d+1\le|\mathbf{i}|\le q}\binom{d-1}{q-|\mathbf{i}|}\prod_{k=1}^{d}\sum_{j_k=1}^{m_{i_k+\ell}}|a_{j_k}^{(i_k+\ell)}(x_k)|\\
&\le \left(1+\frac{5\pi}{2^{\ell+1}}\right)^d\sum_{q-d+1\le|\mathbf{i}|\le q}\binom{d-1}{q-|\mathbf{i}|}\le c(q,d)\left(1+\frac{5\pi}{2^{\ell+1}}\right)^d.
\end{aligned}$$

2.6 Cardinality of Sparse Grids

The remarkable point of the last theorem is that in the case of $\ell \ge 1$, the Lebesgue constant is independent of the number of points. However, this comes at the cost of a larger number of points in our sparse grid.

We now want to estimate this number of points in the Smolyak type sparse grid, where we use the univariate point set $X^{(i+\ell)}$ for an integer offset $\ell \ge 0$. To achieve this, We will use a more general lemma from [17, Lemma 3.23] to bound the final number of points $n(q,d)$:

Lemma 3 *If the number of points m_j in the univariate point sets $X^{(j)}$ satisfies the bound*

$$m_j \le F_0\left(F^j-1\right) \tag{9}$$

for all $j \ge 1$ and some $F_0, F \in \mathbb{N}$ then the cardinality $n(q,d)$ of the Smolyak points $H(q,d)$ can be bounded from above by

$$n(q,d) \le F_0^d\,(F-1)^d\sum_{k=0}^{q-d}F^k\tbinom{k+d-1}{d-1}\le F_0^d\left(\frac{F-1}{F}\right)^d F^q\tbinom{q-1}{d-1}\min\left(\frac{F}{F-1},\frac{q}{d}\right) \tag{10}$$

for $q \ge d$.

In order to apply this result, we have $\#X^{(i+\ell)} = 2^{i+\ell-1}+1 = m_{i+\ell}$. Obviously, we have for $i \ge 1$

$$m_{i+\ell} = 2^{i+\ell-1}+1 \le 2^\ell\left(2^i-1\right) \tag{11}$$

and by definition $m_1 = 1$. Hence, we can choose $F_0 = 2^\ell$ and $F = 2$. This yields

$$n\left(q,d\right) \le 2^{\ell d-d+q}\tbinom{q-1}{d-1}\min\left(2,\frac{q}{d}\right). \tag{12}$$

In order to also obtain a lower bound, we can use another result from [17, Korollar 3.28]:

Lemma 4 *If the number of points m_j in the univariate point sets $X^{(j)}$ satisfies the bound*

$$m_j - m_{j-1} \geq F_0\,(F-1)\,F^{j-1} \tag{13}$$

for all $j \geq 1$ and some $F_0, F \in \mathbb{N}$, then the cardinality $n(q,d)$ of the Smolyak points $H_{q,d}$ has the lower bound

$$n\,(q,d) \geq F_0^d\,(F-1)^d \max\Big(\min\,(1,F)^{q-d}\,(F-1)\tbinom{q}{d}, F^{q-d+1}-1\Big) \tag{14}$$

for $q \geq d$.

For our specific data sets $X^{(j+\ell)}$ we have

$$m_j - m_{j-1} = 2^{j+\ell-1} - 2^{j+\ell-2} = 2^{\ell-1}\,(2-1)\,2^{j-1}.$$

Hence we can apply this result with $F_0 = 2^{\ell-1}$ and $F = 2$ and obtain

$$n\,(q,d) \geq 2^{d\ell-d} \max\Big(\tbinom{q}{d}, 2^{q-d+1}-1\Big)\,. \tag{15}$$

Together with the upper bound (12), this yields

$$2^{d\ell-d} \max\Big(\tbinom{q}{d}, 2^{q-d+1}-1\Big) \leq n\,(q,d) \leq 2^{\ell d-d+q}\tbinom{q-1}{d-1}\min\Big(2, \frac{q}{d}\Big)\,.$$

Using the estimate $\binom{q-1}{d-1} \leq \frac{q^{d-1}}{(d-1)!}$, we can simply this as follows.

Corollary 1 *Let $X^{(i+\ell)}$ be the the points from (7) with $1 \leq i \leq d$ and $\ell \geq 0$. Then, the number of points in the associated Smolyiak sparse grid $H(q,d)$, $q \geq d$, can be bounded from below and above by*

$$2^{\ell d-2d+q+1} \leq n\,(q,d) \leq 2^{\ell d-d+q+1}\frac{q^{d-1}}{(d-1)!}.$$

3 Sampling Inequalities based on Smolyak's Method

The basic idea of proving sampling inequalities is based upon the following argument (see [21]).

Suppose a function $f \in C(I^d)$, a polynomial $p \in \mathcal{S}_{q,d}(I^d)$ and a point $\mathbf{x} \in I^d$ are given. Define the coefficients

$$c_{\mathbf{i}} := (-1)^{q-|\mathbf{i}|}\begin{pmatrix} d-1 \\ q-|\mathbf{i}| \end{pmatrix},$$

so that Smolyiak's algorithm can be written as

$$A(q,d)f(\mathbf{x}) = \sum_{q-d+1\leq|\mathbf{i}|\leq q} c_{\mathbf{i}} \sum_{\mathbf{j}\leq\mathbf{m}} f(\mathbf{x}_{\mathbf{j}}^{(\mathbf{i}+\ell)})\mathbf{a}_{\mathbf{j}}^{(\mathbf{i}+\ell)}(\mathbf{x}).$$

Then, we can estimate

$$\begin{aligned}|f(\mathbf{x})| &\le |f(\mathbf{x}) - p(\mathbf{x})| + |p(\mathbf{x})| = |f(\mathbf{x}) - p(\mathbf{x})| + |A(q,d)p(\mathbf{x})| \\ &\le |f(\mathbf{x}) - p(\mathbf{x})| + |A(q,d)(p-f)(\mathbf{x})| + |A(q,d)f(\mathbf{x})| \\ &\le (1 + \|A(q,d)\|)\|f-p\|_{L_\infty(I^d)} + \left| \sum_{q-d+1 \le |\mathbf{i}| \le q} c_{\mathbf{i}} \sum_{\mathbf{j} \le \mathbf{m}} f(\mathbf{x}_j^{(\mathbf{i})}) \mathbf{a}_{\mathbf{j}}^{(\mathbf{i})}(\mathbf{x}) \right| \\ &\le (1 + \|A(q,d)\|)\|f-p\|_{L_\infty(I^d)} + \|A(q,d)\| \|f\|_{\ell_\infty(X)}\end{aligned}$$

Since this holds for all $\mathbf{x} \in I^d$, we actually have a bound on the $L_\infty(I^d)$-norm of f. Moreover, $p \in \mathcal{S}_{q,d}$ can be chosen arbitrarily and hence we can replace the expression $\|f-p\|_{L_\infty(I^d)}$ by the best approximation error.

Definition 3 Let $V \subseteq C(I^d)$ be a subspace and $1 \le p \le \infty$. Then the best approximation error for a given $f \in C(I^d)$ from V measured in the L_p-norm is defined as

$$\mathcal{E}(f;V)_{L_p} := \inf_{p \in V} \|f-p\|_{L_p(I^d)}.$$

Hence, we can summarise our findings so far. Again, we deal with the case of no oversampling ($\ell = 0$) and oversampling ($\ell \ge 1$) together.

Theorem 6 *Let $f \in C(I^d)$ with $I^d = [-1,1]^d$ be given. Assume that $q \ge d$ and $\ell \ge 0$ are integers. Then f satisfies the sampling inequality*

$$\|f\|_{L_\infty(I^d)} \le (1 + \|A(q,d)\|)\mathcal{E}(f;\mathcal{S}_{q,d})_{L_\infty} + \|A(q,d)\| \|f\|_{\ell_\infty(H(q,d))}.$$

Here, $H(q,d)$ is the sparse grid defined by

$$H(q,d) := \bigcup_{|\mathbf{i}|=q} (X^{(i_1+\ell)} \times \cdots \times X^{(i_d+\ell)}),$$

where $X^{(j)}$ is the univariate set defined in (7).

In the case of $\ell = 0$ we have

$$\|A(q,d)\| \le 2^{d-1} \binom{q-1}{d-1} \left(\frac{2}{\pi} \log(n(q,d) - 1) + 1 \right)^d$$

and in the case of $\ell \ge 1$ we have

$$\|A(q,d)\| \le c(q,d,\ell) = 2^{d-1} \binom{q-1}{d-1} \left(1 + \frac{5\pi}{2^{\ell+1}} \right)^d. \tag{16}$$

We are left with estimating the quantity $\mathcal{E}(f;\mathcal{S}_{q,d}(I^d))_{L_p}$. Here, we can invoke results from [2, Theorem 8 & Equation 13] to reduce the problem to univariate estimates. To be more precise, the results of [2] can be summarised as follows.

Lemma 5 *Suppose the univariate operators $\mathcal{U}^{(i)}$, $1 \le i \le d$, satisfy*

$$\|f - \mathcal{U}^{(i)} f\|_{L_\infty(I^{(i)})} \le c_{1,k} (\log m_i) m_i^{-k} \|f\|_{W_\infty^k(I^{(i)})}, \qquad f \in W_\infty^k(I^{(i)}), \tag{17}$$

for $1 \le i \le d$, where m_i is the number of points on which the operator $\mathcal{U}^{(i)}$ is based. Suppose further that the total number of points $n = n(q,d)$ in the sparse grid $H(q,d)$ is bounded by

$$n \le c_d q^{d-1} 2^q.$$

Then, for every $f \in W_\infty^{k;\otimes^d}(I^d)$ we have

$$\begin{aligned} \mathcal{E}(f; \mathcal{S}_{q,d}(I^d))_{L_\infty} &\le \|f - A(q,d) f\|_{L_\infty(I^d)} \\ &\le c_{d,k} n^{-k} (\log n)^{(k+2)(d-1)+1} \|f\|_{W_\infty^{k;\otimes^d}(I^d)} \end{aligned} \tag{18}$$

with $n = n(q,d)$.

If the univariate operators $\mathcal{U}^{(i)}$, $1 \le i \le d$, satisfy instead of (17) even

$$\|f - \mathcal{U}^{(i)} f\|_{L_\infty(I^{(i)})} \le c_{1,k} m_i^{-k} \|f\|_{W_\infty^{k;\otimes^d}(I^d)}$$

then (18) improves to

$$\begin{aligned} \mathcal{E}(f; \mathcal{S}_{q,d}(I^d))_{L_\infty} &\le \|f - A(q,d) f\|_{L_\infty(I^d)} \\ &\le c_{d,k} n^{-k} (\log n)^{(k+1)(d-1)} \|f\|_{W_\infty^{k;\otimes^d}(I^d)} \end{aligned}$$

The following univariate result is well known, see for example [6, Chapter 7, §6, Theorem 6.2].

Lemma 6 *For $f \in W_p^r(I)$ with $I = [-1,1]$, $1 \le p \le \infty$, and $m \in \mathbb{N}$ with $m \ge r$, we have*

$$\mathcal{E}(f; \pi_m(I))_{L_p} \le C m^{-r} \omega(f^{(r)}, \frac{1}{m})_{L_p} \le C m^{-r} \|f\|_{W_p^r(I)}, \tag{19}$$

where ω is the modulus of continuity (see [6] for details).

We can apply this lemma to our situation in the two cases of no oversampling and with oversampling. We start with the case of no oversampling. This means that we look at the univariate grids $X^{(i)}$, i.e. we set $\ell = 0$.

Theorem 7 *Let $X^{(i)}$ be the univariate grids from (7) and let $H(q,d)$ with $q \ge d$ be the corresponding sparse grid with $n = n(q,d)$ points. Then, for every $f \in W_\infty^{r;\otimes^d}(I^d)$ we have*

$$\begin{aligned} &\|f\|_{L_\infty(I^d)} \\ &\le c_{d,r} c(q,d) \left(\frac{2}{\pi} \log(n-1) + 1 \right)^d n^{-r} (\log n)^{(r+2)(d-1)+1} \|f\|_{W_\infty^{r;\otimes^d}(I^d)} \\ &\quad + c(q,d) \left(\frac{2}{\pi} \log(n-1) + 1 \right)^d \|f\|_{\ell_\infty(H(q,d))}, \end{aligned}$$

where $c(q,d) = 2^{d-1} \binom{q-1}{d-1}$.

Proof Corollary 1 yields in this situation $n \le \frac{2^{1-d}}{(d-1)!} q^{d-1} 2^q =: c_d q^{d-1} 2^q$. Moreover, we have

$$\begin{aligned}
\|f - \mathcal{U}^{(i)} f\|_{L_\infty(I)} &\le c(\log m_i) \cdot \mathcal{E}(f; \pi_{m_i-1}(I))_{L_\infty} \\
&\le c(\log m_i) \cdot m_i^{-r} \|f\|_{W^r_\infty(I)},
\end{aligned}$$

using (8) and (19). Hence, (17) is also satisfied and we can invoke Lemma 5 to derive

$$\begin{aligned}
\|f\|_{L_\infty(I^d)} &\le c(q,d) \left(\frac{2}{\pi} \log(n-1) + 1 \right)^d \mathcal{E}(f; \mathcal{S}_{q,d}(I^d))_{L_\infty} \\
&\quad + c(q,d) (\frac{2}{\pi} \log(n-1) + 1)^d \|f\|_{\ell_\infty(H(q,d))} \\
&\le c(q,d) \left(\frac{2}{\pi} \log(n-1) + 1 \right)^d c_{d,r} n^{-r} (\log n)^{(r+2)(d-1)+1} \|f\|_{W^{k;\otimes^d}_\infty(I^d)} \\
&\quad + c(q,d) (\frac{2}{\pi} \log(n-1) + 1)^d \|f\|_{\ell_\infty(H(q,d))}.
\end{aligned}$$

Next, we come to the situation of oversampling, i.e. we look at the sparse grid $H(q,d)$ generated using the univariate grids $X^{(i+\ell)}$ with $\ell \ge 1$. In this situation, the result improves significantly in two ways. First of all, the term in front of the Sobolev norm of f has now a lower power of the $\log n$ term, though the behaviour is still mainly dominated by the n^{-r} term. Secondly, there is now a constant in front of the discrete norm instead of a $(\log n)^d$ term.

Theorem 8 *Let $X^{(i+\ell)}$ be the univariate grids from (7) with $1 \le i \le d$ and $\ell \ge 1$ and let $H(q,d)$ with $q \ge d$ be the corresponding sparse grid with $n = n(q,d)$ points. Then, for every $f \in W^{r;\otimes^d}_\infty(I^d)$ we have*

$$\begin{aligned}
\|f\|_{L_\infty(I^d)} &\le (1 + c(q,d,\ell)) c_{d,r,\ell} n^{-r} (\log n)^{(r+1)(d-1)} \|f\|_{W^{r;\otimes^d}_\infty(I^d)} \\
&\quad + c(q,d,\ell) \|f\|_{\ell_\infty(H(q,d))}.
\end{aligned}$$

where $c(q,d,\ell)$ is the constant from (16).

Proof Corollary 1 yields this time $n \le \frac{2^{\ell d - d + 1}}{(d-1)!} q^{d-1} 2^q =: c_{d,\ell} q^{d-1} 2^q$. Moreover, we have

$$\begin{aligned}
\|f - \mathcal{U}^{(i)} f\|_{L_\infty(I)} &\le c_\ell \cdot \mathcal{E}(f; \pi_{m_i-1}(I))_{L_\infty} \\
&\le c_\ell \cdot m_i^{-r} \|f\|_{W^r_\infty(I)}
\end{aligned}$$

with $c_\ell = 1 + \frac{5\pi}{2^{\ell+1}}$ from Theorem 4, using the same standard approximation result for univariate polynomials as in the proof of Theorem 7. Thus, Lemma 5 yields this time

$$\begin{aligned}
\|f\|_{L_\infty(I^d)} &\le (1 + c(q,d,\ell)) \mathcal{E}(f; \mathcal{S}_{q,d}(I^d))_{L_\infty} + c(q,d,\ell) \|f\|_{\ell_\infty(H(q,d))} \\
&\le (1 + c(q,d,\ell)) c_{d,r,\ell} n^{-r} (\log n)^{(r+1)(d-1)} \|f\|_{W^{r;\otimes^d}_\infty(I^d)} \\
&\quad + c(q,d,\ell) \|f\|_{\ell_\infty(H(q,d))}.
\end{aligned}$$

It is possible to use other norms than the L_∞ norm on either side of the derived inequalities. Here, we will give only one example, where we replace the mixed order Sobolev space $W_\infty^{r;\otimes^d}(I^d)$ by $W_2^{r;\otimes^d}(I^d)$ since we want to demonstrate how the results derived so far can be used to give error estimates for kernel-based interpolation and smoothing approximation in the next section.

To derive this result, we start again with (19) for $p = \infty$,

$$\mathcal{E}(f;\pi_m(I))_{L_\infty} \le Cm^{-r}\omega\left(f^{(r)}, \frac{1}{m}\right)_{L_\infty} \tag{20}$$

but instead of bounding the modulus of continuity

$$\omega\left(f, \frac{1}{m}\right)_{L\infty} = \sup_{\substack{|x-y|\le\frac{1}{m}\\ x,y\in[-1,1]}} |f(x)-f(y)|$$

by the L_∞ norm, we use

$$|f(x)-f(y)| = \left|\int_x^y f'(t)dt\right| \le \sqrt{|x-y|}\|f'\|_{L_2(I)}.$$

Inserting this into (20) with r replaced by $r-1$, we see that

$$\mathcal{E}(f;\pi_m(I))_{L_\infty} \le Cm^{-r+1/2}\|f^{(r)}\|_{L_2(I)},$$

for all $f \in W_2^r(I)$ and $m \ge r$.

We can now proceed exactly as in the proof of Theorems 7 and 8, respectively. However, since we use Lemma 5, which states all results only using the L_∞-norm, it is important to note that the proof in [2, Theorem 8] remains true in the current situation and, for example, (18) now becomes

$$\begin{aligned}\mathcal{E}(f;\mathcal{S}_{q,d}(I^d))_{L_\infty} &\le \|f - A(q,d)f\|_{L_\infty(I^d)}\\ &\le c_{d,r}n^{-r+\frac{1}{2}}(\log n)^{(r+3/2)(d-1)+1}\|f\|_{W_2^{k;\otimes^d}(I^d)}\end{aligned} \tag{21}$$

Theorem 9 *Let $X^{(i+\ell)}$ be the univariate grids from (7) with $1 \le i \le d$ and $\ell \ge 0$ and let $H(q,d)$ with $q \ge d$ be the corresponding sparse grid with $n = n(q,d)$ points. Then, for every $f \in W_2^{r;\otimes^d}(I^d)$ we have in the case of $\ell = 0$ (no oversampling)*

$$\begin{aligned}&\|f\|_{L_\infty(I^d)}\\ &\le c_{d,r}c(q,d)\left(\frac{2}{\pi}\log(n-1)+1\right)^d n^{-r+1/2}(\log n)^{(r+3/2)(d-1)+1}\|f\|_{W_2^{r;\otimes^d}(I^d)}\\ &\quad + c(q,d)\left(\frac{2}{\pi}\log(n-1)+1\right)^d \|f\|_{\ell_\infty(H(q,d))},\end{aligned}$$

and in the case of $\ell \ge 1$ (oversampling)

$$\begin{aligned}\|f\|_{L_\infty(I^d)} \le &(1+c(q,d,\ell))c_{d,r,\ell}n^{-r+1/2}(\log n)^{(r+1/2)(d-1)}\|f\|_{W_2^{r;\otimes^d}(I^d)}\\ &+ c(q,d,\ell)\|f\|_{\ell_\infty(H(q,d))}.\end{aligned}$$

where $c(q,d,\ell)$ is the constant from (16).

4 Interpolation and Smoothing in Mixed-order Sobolev Spaces

Already in the introduction we mentioned that typical applications of sampling inequalities are stable reconstruction processes. We want to show how the newly derived sampling inequalities can be used in this context. To this end, Suppose we are given a sparse grid $H(q,d)$ as constructed and used in the previous sections. Suppose we are given function values $f(\boldsymbol{\xi})$ for $\boldsymbol{\xi} \in H(q,d)$ and know that the generating function f belongs to $W_2^{r;\otimes^d}(\mathbb{R}^d)$ with $r > 1/2$.

Then, a typical way of reconstructing the unknown function f is to solve either the norm-minimal interpolation problem

$$\min \left\{ \|s\|_{W_2^{r;\otimes^d}} : s \in W_2^{r;\otimes^d}(\mathbb{R}^d) \text{ with } s(\boldsymbol{\xi}) = f(\boldsymbol{\xi}), \boldsymbol{\xi} \in H(q,d) \right\} \tag{22}$$

or the penalised least-squares problem

$$\min_{s \in W_2^{r;\otimes^d}(\mathbb{R}^d)} \left\{ \sum_{\boldsymbol{\xi} \in H(d,q)} |f(\boldsymbol{\xi}) - s(\boldsymbol{\xi})|^2 + \lambda \|s\|^2_{W_2^{r;\otimes^d}} \right\}. \tag{23}$$

Here $\lambda \geq 0$ is a smoothing parameter. For $\lambda = 0$ finding the solution to this problem is equivalent to finding the norm-minimal interpolant to f.

Note that $W_2^{r;\otimes^d}(\mathbb{R}^d)$ is, because of the Sobolev embedding theorem, a reproducing kernel Hilbert space provided that $r > 1/2$. Even more, the reproducing kernel of $W_2^{r;\otimes^d}(\mathbb{R}^d)$ is easily determined as the tensor product of the reproducing kernel of $W_2^r(\mathbb{R})$. To see this, we let

$$\widehat{f}(\boldsymbol{\omega}) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} f(\mathbf{x}) e^{-i\mathbf{x}^T\boldsymbol{\omega}} d\mathbf{x}, \qquad \boldsymbol{\omega} \in \mathbb{R}^d,$$

be the usual Fourier transform of $f \in L_1(\mathbb{R}^d)$, which is extended in the usual way to $L_2(\mathbb{R}^d)$. Then (see [21]), a function $\phi : \mathbb{R} \to \mathbb{R}$ is a reproducing kernel of $W_2^r(\mathbb{R})$ if there are $c_1, c_2 > 0$ such that

$$c_1(1+|t|^2)^{-r} \leq \widehat{\phi}(t) \leq c_2(1+|t|^2)^{-r}, \qquad t \in \mathbb{R}.$$

Proposition 1 *Let $\phi : \mathbb{R} \to \mathbb{R}$ be a reproducing kernel of $W_2^r(\mathbb{R})$ with $r > 1/2$. Then, the reproducing kernel of $W_2^{r;\otimes^d}(\mathbb{R}^d)$ is given by the tensor product $\Phi : \mathbb{R}^d \to \mathbb{R}$ with*

$$\Phi(\mathbf{x}) := \prod_{j=1}^d \phi_j(x_j), \qquad \mathbf{x} = (x_1, \ldots, x_d)^T.$$

Proof The Fourier transform of Φ is given by

$$\begin{aligned}\widehat{\Phi}(\boldsymbol{\omega}) &= (2\pi)^{-d/2} \int_{\mathbb{R}^d} \Phi(\mathbf{x}) e^{-i\mathbf{x}^T\boldsymbol{\omega}} d\mathbf{x} = \prod_{j=1}^d \left((2\pi)^{-1/2} \int_{\mathbb{R}} \phi(x_j) e^{-ix_j\omega_j} dx_j \right) \\ &= \prod_{j=1}^d \widehat{\phi}(\omega_j)\end{aligned}$$

and hence behaves like
$$\prod_{j=1}^{d}(1+|\omega_j|^2)^r,$$
i.e. Φ is a reproducing kernel of $W_2^{r;\otimes^d}(\mathbb{R}^d)$.

Note that different reproducing kernels of $W_2^{r;\otimes^d}(\mathbb{R}^d)$ lead to different but equivalent norms on $W_2^{r;\otimes^d}(\mathbb{R}^d)$. Because of this, we fix the norm in (22) and in (23) as the one generated by a given kernel Φ. Having done this, it is well-known that the solutions to (22) and (23) are a linear combination of the kernel. To be more precise:

Proposition 2 *Let $r > 1/2$ and let $\Phi : \mathbb{R}^d \to \mathbb{R}$ be a reproducing kernel of $W_2^{r;\otimes^d}(\mathbb{R}^d)$. Then, the solution s_λ of (22) and (23), respectively, is given by*
$$s_\lambda(\mathbf{x}) = \sum_{\boldsymbol{\xi}\in H(q,d)} \alpha_{\boldsymbol{\xi}}\Phi(\mathbf{x},\boldsymbol{\xi}), \qquad \mathbf{x}\in\mathbb{R}^d,$$
where the coefficients $\boldsymbol{\alpha}\in\mathbb{R}^{n(q,d)}$ are determined by the linear system
$$(A+\lambda I)\boldsymbol{\alpha} = \mathbf{f}.$$
Here, $A = \Phi(\boldsymbol{\xi},\boldsymbol{\eta})_{\mathbf{x},\boldsymbol{\eta}\in H(q,d)}$ and $\mathbf{f} = (f(\boldsymbol{\xi}))_{\boldsymbol{\xi}\in H(q,d)}$.

To apply our sampling inequality to derive error estimates for the first reconstruction process, i.e. the norm minimal interpolant, we note that we have on the one hand $s_0(\boldsymbol{\xi}) = f(\boldsymbol{\xi})$ for all $\boldsymbol{\xi}\in H(q,d)$ and on the other hand $\|s_0\|_{W_2^{r;\otimes^d}(\mathbb{R}^d)} \le \|f\|_{W_2^{r;\otimes^d}}$. This gives:

Corollary 2 *Let $r > 1/2$ and $\Phi : \mathbb{R}^d \to \mathbb{R}$ be a reproducing kernel of $W_2^{r;\otimes^d}(\mathbb{R}^d)$. Let s_0 be the norm-minimal interpolant (22) on the sparse grid $H(q,d)$. Then, in the case of $\ell = 0$ (no oversampling), we have*
$$\|f-s_0\|_{L_\infty(I^d)} \le Cn^{-r+1/2}(\log n)^{(r+3/2)(d-1)+d+1}\|f\|_{W_2^{r;\otimes^d}(\mathbb{R}^d)}$$
and in the case of $\ell \ge 1$ (oversampling), we have
$$\|f-s_0\|_{L_\infty(I^d)} \le Cn^{-r+1/2}(\log n)^{(r+1/2)(d-1)}\|f\|_{W_2^{r;\otimes^d}(\mathbb{R}^d)}$$
Moreover, s_0 coincides with the interpolant generated by Smolyak's algorithm when using univariate RBF interpolation with the kernel ϕ.

Proof The error estimates immediately follow from Theorem 9. Since the norm-minimal interpolant is the unique function from $V := \operatorname{span}\{\Phi(\cdot,\boldsymbol{\xi}) : \boldsymbol{\xi}\in H(q,d)\}$ and since Smolyak's algorithm produces an interpolant from V both must coincide.

To derive error estimates for the second reconstruction process, i.e. the penalised least-squares problem, we follow [22] and note that we have this time $\|s_\lambda\|_{W_2^{r;\otimes^d}(\mathbb{R}^d)} \le \|f\|_{W_2^{r;\otimes^d}(\mathbb{R}^d)}$ and
$$|s_\lambda(\boldsymbol{\xi}) - f(\boldsymbol{\xi})| \le \sqrt{\lambda}\|f\|_{W_2^{r;\otimes^d}(\mathbb{R}^d)}, \qquad \boldsymbol{\xi}\in H(q,d).$$

Corollary 3 *Let $r > 1/2$ and $\Phi : \mathbb{R}^d \to \mathbb{R}$ be a reproducing kernel of $W_2^{r;\otimes^d}(\mathbb{R}^d)$. Let s_λ be the penalised least-squares solution of (23) on the sparse grid $H(q,d)$. Then, in the case of $\ell \geq 0$ (no oversampling), we have*

$$\|f - s_\lambda\|_{L_\infty(I^d)} \leq C \left(n^{-r+1/2} (\log n)^{(r+3/2)(d-1)+d+1} + \sqrt{\lambda} (\log n)^d \right) \|f\|_{W_2^{r;\otimes^d}(\mathbb{R}^d)}$$

and in the case of $\ell \geq 1$ (oversampling), we have

$$\|f - s_0\|_{L_\infty(I^d)} \leq C \left(n^{-r+1/2} (\log n)^{(r+1/2)(d-1)} + \sqrt{\lambda} \right) \|f\|_{W_2^{r;\otimes^d}(\mathbb{R}^d)}$$

Usually, the parameter λ is determined using statistical methods like cross validation, see [20], but the error estimates derived here can also be used to determine λ in a deterministic way.

5 Numerical Example

For the numerical validation of the convergence results, we first need to discuss the sparse grids $H(q,d)$ in more details. Hence, we have calculated the sparse grids for various combinations of q and d. Note that using formula (5) for our nested point sets from Definition 1 leads to a significant number of multiple points, which is problematic for interpolation and also for the computational cost. Disregarding these multiple points, the cardinality $n = n(q,d)$ of $H(q,d)$ is given in Table 1.

Since the condition number of the interpolation matrix mainly depends on the so-called *separation distance*, defined by $\min_{k \neq j} \|\mathbf{x}_k - \mathbf{x}_j\|_2$, Table 2 shows the separation distance for those sparse grids from Table 1. A closer look at that table shows that the separation distance depends actually only on $i := q - d + 1$ and not on q and d separately. This is not surprising at all. As a matter of fact, the separation distance can be computed analytically. It is given by the two closest points in the finest univariate grid employed and this grid is given for those multi-indices $\mathbf{i} \in \mathbb{N}^d$ with $|\mathbf{i}| = q$ of the form $\mathbf{i} = (1, \ldots, 1, q - d + 1)$. Obviously, it does not matter in which positition the number $i = q - d + 1$ appears. Hence, the separation distance of $H(q,d)$ is determined by the separation distance of the univariate set $X^{(i)} = X_{m_i}$ and is simply given by

$$|x_1^{(i)} - x_2^{(i)}| = 1 - \cos(2^{1-i}\pi) = 1 - \cos(2^{d-q}\pi), \qquad i = q - d + 1.$$

Obviously, a similar relation holds in the case of oversampling, i.e. for $\ell \geq 1$.

In our example, we have used a kernel which is a tensor product of univariate Wendland kernels, i.e.,

$$K(\mathbf{x}, \mathbf{y}) = \prod_{j=1}^{3} K_1(x_j, y_j), \qquad \mathbf{x}, \mathbf{y} \in \mathbb{R}^d$$

with K_1 given by

$$K_1(x,y) = (1 - |x-y|)_+^3 (3|x-y| + 1), \qquad x, y \in \mathbb{R}.$$

$q\|d$	2	3	4	5	6	7	8
2	1						
3	5	1					
4	13	7	1				
5	29	25	9	1			
6	65	69	41	11	1		
7	145	177	137	61	13	1	
8	321	441	401	241	85	15	1
9	705	1073	1105	801	389	113	17
10	1537	2561	2929	2433	1457	589	145
11	3329	6017	7537	6993	4865	2465	849
12	7169	13953	18945	19313	15121	9017	3937
13	15361	32001	46721	51713	44689	30241	15713
14	32769	72705	113409	135073	127105	95441	56737
15	69633	163841	271617	345665	350657	287745	190881

Table 1 Number of points $n(q,d)$ of the grid $H(q,d)$ for various space dimensions d and $q \geq d$.

$q\|d$	2	3	4	5	6	7	8
3	1						
4	2.93e-01	1					
5	7.61e-02	2.93e-01	1				
6	1.92e-02	7.61e-02	2.93e-01	1			
7	4.82e-03	1.92e-02	7.61e-02	2.93e-01	1		
8	1.20e-03	4.82e-03	1.92e-02	7.61e-02	2.93e-01	1	
9	3.01e-04	1.20e-03	4.82e-03	1.92e-02	7.61e-02	2.93e-01	1
10	7.53e-05	3.01e-04	1.20e-03	4.82e-03	1.92e-02	7.61e-02	2.93e-01
11	1.88e-05	7.53e-05	3.01e-04	1.20e-03	4.82e-03	1.92e-03	7.61e-02
12	4.71e-06	1.88e-05	7.53e-05	3.01e-04	1.20e-03	4.82e-03	1.92e-03
13	1.18e-06	4.71e-06	1.88e-05	7.53e-05	3.01e-04	1.20e-03	4.82e-03
14	2.94e-07	1.18e-06	4.71e-06	1.88e-05	7.53e-05	3.01e-04	1.20e-03
15	7.35e-08	2.94e-07	1.18e-06	4.71e-06	1.88e-05	7.53e-05	3.01e-04

Table 2 Separation distance for the sparse grids $H(q,d)$ for various dimensions d and $q \geq d+1$.

It is known (see [21]) that K_1 is the reproducing kernel of $W_2^2(\mathbb{R})$. This means that, according to Corollaries 2 and 3, we can expect an error of the form

$$\|f - s_0\|_{L_\infty(I^3)} \leq C\left(n^{-1.5}(\log n)^{4.5d-2.5} + \sqrt{\lambda}(\log n)^{3.5(d-1)}\right)\|f\|_{W_2^{2;\otimes^3}(\mathbb{R}^3)} \quad (24)$$

in the case of $\ell = 0$ (no oversampling) and an error of the form

$$\|f - s_\lambda\|_{L_\infty(I^3)} \leq C\left(n^{-1.5}(\log n)^{2.5(d-1)} + \sqrt{\lambda}\right)\|f\|_{W_2^{2;\otimes^3}(\mathbb{R}^3)} \quad (25)$$

in the case of $\ell \geq 1$, both with $\lambda \geq 0$ and for target functions $f \in W_2^{2;\otimes^d}(\mathbb{R}^d)$. The target function we have used here is given by

$$f(\mathbf{x}) = \prod_{j=1}^{d} |x_j|^{1.6}, \qquad \mathbf{x} \in \mathbb{R}^d, \quad (26)$$

which obviously belongs to $W_2^{2;\otimes^d}(\mathbb{R}^d)$ but not to $W_2^{3;\otimes^d}(\mathbb{R}^d)$. To calculate the error, we used a full grid of univariate step-size $h = 0.01$, which restricts our test cases to $d = 2$ and $d = 3$.

q	N	L_∞	order	N	L_∞	order
	$\ell = 0$			$\ell = 1$		
4	13	0.198646		49	0.0235198	
5	29	0.124942	0.577896	113	0.0159142	0.467505
6	65	0.0235198	2.06917	257	0.00323925	1.93731
7	145	0.0159142	0.486862	577	0.000624243	2.03589
8	321	0.00323926	2.00309	1281	0.000144757	1.83245
9	705	0.000624255	2.09282	2817	4.55627e-05	1.46691
10	1537	0.000144753	1.87523	6145	1.44712e-05	1.47048
11	3329	4.54492e-05	1.49893	13313	4.4915e-06	1.51335
12	7169	1.44186e-05	1.49665	28673	1.06589e-06	1.8748
13	15361	4.51259e-06	1.52434	61441		

Table 3 $d = 2$, $\lambda = 10^{-8}$, test function (26).

q	N	L_∞	order	N	L_∞	order
	$\ell = 0$			$\ell = 1$		
5	25	1		225	0.0453241	
6	69	0.254733	1.34702	593	0.0159435	1.07811
7	177	0.159975	0.493818	1505	0.00351793	1.62256
8	441	0.0453241	1.38152	3713	0.00211576	0.563045
9	1073	0.0159435	1.17501	8961	0.000410079	1.86236
10	2561	0.00351792	1.73711	21247	8.94666e-05	1.7633
11	6017	0.00211579	0.595235	49665	1.87446e-05	1.84096
12	13953	0.000410097	1.95075			
13	32001	8.95064e-05	1.83367			

Table 4 $d = 3$, $\lambda = 10^{-8}$, test function (26).

As mentioned above, we have to expect a conditioning problem with a separation distance too small. Since it is also well-known that the $\sqrt{\lambda}$ term in (24) and (25) is a rather pessimistic estimate, we have used $\lambda = 10^{-8}$ for all of our computations without significantly compromising the interpolation error, as long as the number of data sites does not become too large.

The results on the L_∞-error and the estimated approximation order can be seen in Tables 3 and 4. In the case of $d = 2$ (Table 3), we see that the predicted rates are asymptotically sharp, regarding the $n^{-1.5}$ term since the $\log n$ term is hard to track. In the case of $d = 3$ (Table 4) the numerical approximation order varies more but eventually also confirms the order, see also Figure 1.

6 Conclusion

In this paper we have, for the first time, derived rigorous sampling inequalities for functions from mixed regularity Sobolev spaces on certain sparse grids. We have done this for classical sparse grids built from Chebyshev points with and without oversampling. Such sampling inequalities play a crucial role for deriving error estimates for stable reconstruction processes. As examples, we have shown how these sampling inequalties can be applied for norm-minimal, kernel-based interpolation and penalised least-squares approximation. The derived bounds are explicit with respect to the oversampling and the penalisation parameter.

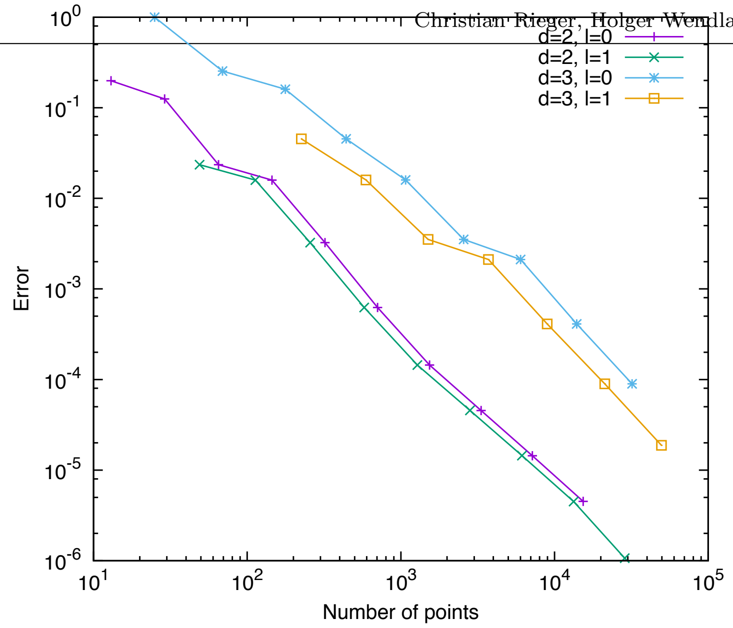

Fig. 1 Error plots for the test function (26).

Acknowledgement

We thank Michael Griebel for helpful discussions. The authors acknowledge partial support by the Deutsche Forschungsgemeinschaft through the Collaborative Research Centers (SFB) 1060 and the Hausdorff Center for Mathematics.

References

1. R. Arcangéli, M. Cruz López de Silanes, and J. J. Torrens. An extension of a bound for functions in Sobolev spaces, with applications to (m, s)-spline interpolation and smoothing. *Numer. Math.*, 107:181–211, 2007.
2. V. Barthelmann, E. Novak, and K. Ritter. High dimensional polynomial interpolation on sparse grids. *Adv. Comput. Math.*, 12:273–288, 2000.
3. H.-J. Bungartz and M. Griebel. Sparse grids. In A. Iserles, editor, *Acta Numerica*, volume 13, pages 1–123. Cambridge University Press, 2004.
4. Glenn Byrenheid, Dinh Dung, Winfried Sickel, and Tino Ullrich. Sampling on energy-norm based sparse grids for the optimal recovery of Sobolev type functions in H. *ArXiv e-prints*, 2014. arXiv:1408.3498 [math.NA].
5. E. W. Cheney. *Introduction to Approximation Theory*. McGraw-Hill Book Company, New York, 1966.
6. R.A. De Vore and G.G. Lorentz. *Constructive Approximation*. Grundlehren der mathematischen Wisenschaften. Springer, Berlin, New York, 1993.
7. Jochen Garcke and Markus Hegland. Fitting multidimensional data using gradient penalties and the sparse grid combination technique. *Computing*, 84:1–25, 2009.
8. M. Griebel. Sparse grids for higher dimensional problems. In L. M. Pardo, A. Pinkus, E. Süli, and M. J. Todd, editors, *Foundations of Computational Mathematics, Santander 2005*, pages 106–161. Cambridge University Press, 2006.
9. K. Jetter, J. Stöckler, and J. Ward. Error estimates for scattered data interpolation on spheres. *Math. Comput.*, 68:733–747, 1999.

10. J. Krebs. Support vector regression for the solution of linear integral equations. *Inverse Problems*, 27(6):065007 (23 pages), 2011.
11. J. Krebs, A. K. Louis, and H. Wendland. Sobolev error estimates and a priori parameter selection for semi-discrete tikhonov regularization. *Journal of Inverse and Ill-Posed Problems*, 17:845–869, 2009.
12. W. R. Madych. An estimate for multivariate interpolation II. *J. Approx. Theory*, 142:116–128, 2006.
13. F. J. Narcowich, J. D. Ward, and H. Wendland. Sobolev bounds on functions with scattered zeros, with applications to radial basis function surface fitting. *Math. Comput.*, 74:643–763, 2005.
14. C. Rieger, R. Schaback, and B. Zwicknagl. Sampling and stability. In *Mathematical Methods for Curves and Surfaces*, volume 5862 of *Lecture Notes in Computer Science*, pages 347–369. Springer, New York, 2010.
15. C. Rieger and B. Zwicknagl. Deterministic error analysis of support vector machines and related regularized kernel methods. *J. Mach. Learn. Res.*, 10:2115–2132, 2009.
16. C. Rieger and B. Zwicknagl. Improved exponential convergence rates by oversampling near the boundary. submitted, 2012.
17. A. Schreiber. *Die Methode von Smolyak bei der multivariaten Interpolation*. PhD thesis, Universität Göttingen, 2000.
18. S. A. Smolyak. Quadrature and interpolation formulas for tensor products of certain classes of functions. *Soviet Math. Dokl.*, 4:240–243, 1963.
19. Tino Ullrich. Smolyak's algorithm, sampling on sparse grids and Sobolev spaces of dominating mixed smoothness. *East Journal on Approximations*, 14(1):1–38, 2008.
20. G. Wahba. *Spline Models for Observational Data*. CBMS-NSF, Regional Conference Series in Applied Mathematics. Siam, Philadelphia, 1990.
21. H. Wendland. *Scattered Data Approximation*. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, Cambridge, UK, 2005.
22. H. Wendland and C. Rieger. Approximate interpolation with applications to selecting smoothing parameters. *Numer. Math.*, 101:643–662, 2005.