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

M. Griebel, J. Hamaekers

Fast Discrete Fourier Transform on Generalized Sparse Grids

INS Preprint No. 1305

March 2013

Fast Discrete Fourier Transform on Generalized Sparse Grids

Michael Griebel and Jan Hamaekers

Abstract In this paper, we present an algorithm for trigonometric interpolation of multivariate functions on generalized sparse grids and study its application for the approximation of functions in periodic Sobolev spaces of dominating mixed smoothness. In particular, we derive estimates for the error and the cost. We construct interpolants with a computational cost complexity which is substantially lower than for the standard full grid case. The associated generalized sparse grid interpolants have the same approximation order as the standard full grid interpolants, provided that certain additional regularity assumptions on the considered functions are fulfilled. Numerical results validate our theoretical findings.

1 Introduction

In many application areas of numerical simulation, like e.g. physics, chemistry, finance and statistics, high-dimensional approximation problems arise. Here, a conventional numerical approach encounters the so-called curse of dimensionality [4], i.e. the rate of convergence with respect to the number of degrees of freedom deteriorates exponentially with the dimension n. For example a conventional discretization on uniform grids with $\mathcal{O}(2^L)$ points in each direction would involve $M = \mathcal{O}(2^{nL})$ degrees of freedom. Moreover, only a convergence rate of the type

$$\|f - f_L^{\mathrm{FG}}\|_{\mathcal{H}^r} \leq c \cdot M^{-\frac{s-r}{n}} \|f\|_{\mathcal{H}^s}$$

Michael Griebel
Institute for Numerical Simulation, University of Bonn, Wegelerstr. 6, 53115 Bonn, Germany, e-mail: griebel@ins.uni-bonn.de

Jan Hamaekers
Fraunhofer Institute for Algorithms and Scientific Computing SCAI, Schloss Birlinghoven, 53754 Sankt Augustin, Germany, e-mail: jan.hamaekers@scai.fraunhofer.de

can be achieved, where $\|\cdot\|_{\mathcal{H}^r}$ is the usual Sobolev norm in $\mathcal{H}^r$, s denotes the isotropic smoothness of f and c is a constant which may depend on n and the underlying domain Ω but not on the discretization parameter L.

So-called sparse grid based approaches have emerged as useful techniques to tackle higher dimensional problems, since they open the possibility to break the curse of dimensionality under certain conditions. They date back to [47]. For example, if f is in a Sobolev space of bounded mixed smoothness $\mathcal{H}^t_{\text{mix}}(\Omega)$, i.e. if the t-th mixed derivatives of f are bounded, and Ω is a product domain, an error estimate of the type

$$\|f - f_L^{\text{SG}}\|_{\mathcal{H}^r} \leq c \cdot 2^{-(t-r)L} L^{n-1} \|f\|_{\mathcal{H}^t_{\text{mix}}}$$

can be achieved using so-called regular sparse grids where $\mathcal{O}(2^L L^{n-1})$ degrees of freedom are involved.[1] Here, the rate of convergence with respect to the number of degrees of freedom does no longer exponentially deteriorate with the number n of dimensions, except for the logarithmic terms L^{n-1}. Moreover, in specific cases, the use of so-called energy-norm based sparse grids [6] may even result in an error estimate of type

$$\|f - f_L^{\text{ESG}}\|_{\mathcal{H}^r} \leq c \cdot 2^{-(t-r)L} \|f\|_{\mathcal{H}^t_{\text{mix}}},$$

where only $\mathcal{O}(2^L)$ degrees of freedom are involved. Hence, compared to the regular sparse grid case even the logarithmic terms L^{n-1} are eliminated. [2]

For the discretization with sparse grids, Galerkin type methods, finite difference approaches and the so-called combination technique have been developed over the last two decades [6]. Furthermore, these approaches were used in the context of moderate higher-dimensional elliptic, parabolic and hyperbolic differential equations. In addition, sparse grid techniques were successfully applied for the solution of integral equations [25], for quadrature [14], for regression [12] and for time series prediction [5]. Moreover, the sparse grid method was supplemented with adaptive refinement schemes [5, 14, 22], was used for the construction of anisotropic sparse tensor product spaces [21, 20] and was applied in the context of weighted mixed spaces [19, 22]. Sparse grid based collocation schemes were for example discussed in [2, 26, 29, 30, 33, 39]. They recently found widespread use in the important field of uncertainty quantification [38]. On a theoretical level, sparse grids are closely related to ANOVA-like decompositions [11, 16, 27] which are well-known from statistics. A detailed survey on sparse grids is for example given in [6].

Note that the adaption of the sparse grid techniques to Fourier based methods is done by means of Fourier polynomials from the hyperbolic cross and hence sparse grid methods are also known under the name hyperbolic cross approximation

---

[1] Here, in case of the best linear approximation, estimates of type $\|f - f_L^{\text{SG}}\|_{\mathcal{H}^r} \lesssim 2^{-(t-r)L} L^{(n-1)/2} \|f\|_{\mathcal{H}^t_{\text{mix}}}$ and even of type $\|f - f_L^{\text{SG}}\|_{\mathcal{H}^r} \lesssim 2^{-(t-r)L} \|f\|_{\mathcal{H}^t_{\text{mix}}}$ could be achieved for certain types of basis sets [25, 34, 49]. This holds, e.g. for wavelets and the Fourier basis, respectively.

[2] The constants in the cost and accuracy estimates still depend on n, though.

[43, 48]. The properties of such approximations of functions in Sobolev spaces on the n-dimensional torus $\mathbb{T}^n$ have been studied by several authors [7, 9, 15, 31, 32, 34, 36, 41, 44, 45, 48]. In particular, spaces of generalized mixed Sobolev smoothness

$$\mathcal{H}^{t,r}_{\mathrm{mix}}(\mathbb{T}^n) := \left\{ f : \sqrt{\sum_{\mathbf{k}\in\mathbb{Z}^n} \prod_{d=1}^{n} (1+|k_d|)^{2t}(1+|\mathbf{k}|_\infty)^{2r}|\hat{f}_{\mathbf{k}}|^2} < \infty \right\}$$

and a specific generalization of the regular sparse grid spaces based on Fourier polynomials $e^{i\mathbf{k}^T\mathbf{x}}$ with frequencies $\mathbf{k}$ from the generalized hyperbolic cross

$$\Gamma_L^T := \left\{ \mathbf{k} \in \mathbb{Z}^n : \prod_{d=1}^{n} (1+k_d) \cdot (1+|\mathbf{k}|_\infty) \le L^{(1-T)} \right\}$$

were introduced in [34], further discussed in [23, 24, 27, 35] and a generalization to Banach spaces is given in [8]. Here, $T \in [-\infty, 1)$ is an additional parameter that controls the mixture of isotropic and mixed smoothness: The case $T = 0$ corresponds to the conventional hyperbolic cross (or regular sparse grid). In that case for example the $\mathcal{H}^r$-error of the best linear approximate f_L^{HC} in the conventional hyperbolic cross discretization space of a function in a periodic Sobolev space of dominated mixed smoothness $\mathcal{H}^{t,r}(\mathbb{T}^n)$ is of order $\mathcal{O}(2^{-tL})$, where $\mathcal{O}(2^L L^{n-1})$ frequencies are involved. Furthermore, the case $T = -\infty$ corresponds to the full grid, the case $T \to 1$ corresponds to a latin hypercube and the case $0 < T < 1$ resembles energy-norm based sparse grids where the order of the amount of included frequencies does not depend on the number of dimensions n, i.e. it is $\mathcal{O}(2^L)$. But let us note here that it is in general not clear, if the approximation error of an *interpolant* exhibits the same convergence rates as that of the best linear approximation.

In this paper, we now mainly deal with trigonometric interpolation on generalized sparse grids and its application for the approximation of multivariate functions in certain periodic Sobolev spaces of bounded mixed smoothness. For functions on the torus, regular sparse grid interpolation methods based on the fast Fourier transform were for example introduced by Hallatschek in [26] and also discussed in [3, 15, 30, 31, 36, 44]. For example, for a function in a Korobov space of mixed smoothness $t > 1$ it is proved in [26] that the approximation error in the maximum norm of its regular sparse grid interpolant is of the order $\mathcal{O}(2^{-L(t-1)}L^{n-1})$ and in [15] a (suboptimal) upper bound estimate of the same order is shown for the approximation error in the $\mathcal{L}^2$ norm. Here, the involved degrees of freedom are of the order $\mathcal{O}(2^L L^{n-1})$ and the computational cost complexity is of the order[3] $\mathcal{O}(2^L L^n)$. Based on the results of [48] it was furthermore shown in [36, 37] that the approximation error in the $\mathcal{H}^r$-norm of the interpolant associated with a regular sparse grid is of the order $\mathcal{O}(2^{-(t-r)L}L^{n-1})$, if the function is in a periodic Sobolev space $\mathcal{H}^t_{\mathrm{mix}}$ with $t > \frac{1}{2}$.

[3] We here do not have $\mathcal{O}(2^L L^{n-1})$ but we have $\mathcal{O}(2^L L^n)$ since one L stems from the computational complexity of the one-dimensional FFT involved.

In this work, we present an extension of the algorithm of Hallatschek given in [26] to the case of interpolation on the generalized sparse grids as introduced in [23]. We will further study its best linear approximation error and will give cost complexity estimates for functions in different variants of periodic Sobolev spaces of dominating mixed smoothness. Moreover, for functions of mixed Sobolev smoothness $\mathcal{H}^t_{\text{mix}}$, we will show estimates for the approximation error of the interpolant in the $\mathcal{H}^r$-norm. To our knowledge this has been done so far only for the regular sparse grid case $T = 0$, but not yet for the case of generalized sparse grids with $0 < T < 1$, which resemble the energy-norm based sparse grids. Note further that the behavior of the approximation error of the interpolant versus the computational complexity of the interpolant is of practical interest. This holds especially with possible applications in the field of uncertainty quantification in mind. Therefore, we give also estimates for its computational complexity. Altogether, it turns out that under specific conditions the order rates of the error complexity and computational complexity are independent of the dimension n of the function. For example, let $f \in \mathcal{H}^2_{\text{mix}}(\mathbb{T}^n)$ and let us measure the approximation error in the $\mathcal{H}^1$-norm, where $\mathbb{T}^n$ denotes the n-dimensional torus. Then, an error of the order $\mathcal{O}(2^{-L})$ and a computational complexity of the order $\mathcal{O}(2^L L)$ can be achieved[4] for interpolants corresponding to a generalized sparse grid with $0 < T < \frac{1}{2}$ for any dimension n.

The remainder of this paper is organized as follows: In section 2 we introduce the fast Fourier transform on general sparse grids with hierarchical bases. In particular, we will recall the conventional Fourier basis representation of periodic functions in subsection 2.1 and the so-called hierarchical Fourier basis representation in subsection 2.2. Furthermore, in subsection 2.3, we will present generalized sparse grids and discuss the construction and application of associated trigonometric interpolation operators and computational complexities. In section 3 we will introduce different variants of periodic Sobolev spaces and will discuss their associated best linear approximation error in subsection 3.2, the approximation error of the trigonometric general sparse grid interpolants in subsection 3.3 and its overall complexities in subsection 3.4. Then, in subsection 3.5, we will give some short remarks on further generalizations of sparse grids like, e.g. periodic Sobolev spaces with finite-order weights and dimension-adaptive approaches. In section 4 we will apply our approach to some test cases. Finally we give some concluding remarks in section 5.

2 Fourier transform on general sparse grids with hierarchical bases

To construct a trigonometric interpolation operator for generalized sparse grids, we will follow the approach of Hallatschek [26]. To this end, we will first recall the conventional Fourier basis and then introduce the so-called hierarchical Fourier basis and its use in the construction of a generalized sparse grid interpolant.

[4] Here, also the L stems from the involved FFT algorithm.

2.1 Fourier basis representation

First, let us shortly recall the usual Fourier basis representation of periodic functions. To this end, let $\mathbb{T}^n$ be the n-torus, which is the n-dimensional cube $\mathbb{T}^n \subset \mathbb{R}^n$, $\mathbb{T} = [0, 2\pi]$, where opposite sides are identified. We then have n-dimensional coordinates $\mathbf{x} := (x_1, \dots, x_n)$, where $x_d \in \mathbb{T}$. We define the basis function associated with a multi index $\mathbf{k} = (k_1, \dots, k_n) \in \mathbb{Z}^n$ by

$$\omega_{\mathbf{k}}(\mathbf{x}) := \left(\bigotimes_{d=1}^{n} \omega_{k_d} \right)(\mathbf{x}) = \prod_{d=1}^{n} \omega_{k_d}(x_d), \quad \omega_k(x) := e^{ikx}. \tag{1}$$

The set $\{\omega_{\mathbf{k}}\}_{\mathbf{k}\in\mathbb{Z}^n}$ is a complete orthogonal system of the space $\mathcal{L}^2(\mathbb{T}^n)$ and hence every $f \in \mathcal{L}^2(\mathbb{T}^n)$ has the unique expansion

$$f(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbb{Z}^n} \hat{f}_{\mathbf{k}} \omega_{\mathbf{k}}(\mathbf{x}), \tag{2}$$

where the Fourier coefficients are given by

$$\hat{f}_{\mathbf{k}} := \frac{1}{(2\pi)^n} \int_{\mathbb{T}^n} \omega^*_{\mathbf{k}}(\mathbf{x}) f(\mathbf{x})\, d\mathbf{x}. \tag{3}$$

Note that it is common to characterize the smoothness classes of a function f by the decay properties of its Fourier coefficients [28]. In this way, we introduce the periodic Sobolev space of isotropic smoothness as

$$\mathcal{H}^r(\mathbb{T}^n) := \left\{ f(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbb{Z}^n} \hat{f}_{\mathbf{k}} \omega_{\mathbf{k}}(\mathbf{x}) : \|f\|_{\mathcal{H}^r} := \sqrt{\sum_{\mathbf{k}\in\mathbb{Z}^n} (1+|\mathbf{k}|_\infty)^{2r} |\hat{f}_{\mathbf{k}}|^2} < \infty \right\}$$

for $r \in \mathbb{R}$.

Let us now define finite-dimensional subspaces of the space $\mathcal{L}^2(\mathbb{T}^n) = \mathcal{H}^0(\mathbb{T}^n)$ for discretization purposes. To this end, we set

$$\sigma : \mathbb{N}_0 \to \mathbb{Z} : j \mapsto \begin{cases} -j/2 & \text{if } j \text{ is even,} \\ (j+1)/2 & \text{if } j \text{ is odd.} \end{cases} \tag{4}$$

For $l \in \mathbb{N}_0$ we introduce the one-dimensional nodal basis

$$\mathcal{B}_l := \{\phi_j\}_{0 \le j \le 2^l - 1} \quad \text{with} \quad \phi_j := \omega_{\sigma(j)} \tag{5}$$

and the corresponding spaces $V_l := \operatorname{span}\{\mathcal{B}_l\}$. For a multi index $\mathbf{l} \in \mathbb{N}_0^n$ we define finite-dimensional spaces by a tensor product construction, i.e. $V_{\mathbf{l}} := \bigotimes_{d=1}^{n} V_{l_d}$. Finally, we introduce the space[5]

[5] Except for the completion with respect to a chosen Sobolev norm, V is just the associated Sobolev space.

$$V := \sum_{\mathbf{l}\in\mathbb{N}^n} V_{\mathbf{l}}.$$

In the following we will shortly recall the common one-dimensional trigonometric interpolation. Let the Fourier series $\sum_{k\in\mathbb{Z}} \hat{f}_k \omega_k$ be pointwise convergent to f. Then, for interpolation points $\mathcal{S}_l := \{m\frac{2\pi}{2^l} : m = 0,\dots,2^l-1\}$ of level $l \in \mathbb{N}_0$, the interpolation operator can be defined by

$$I_l : V \to V_l : f \mapsto I_l f := \sum_{j\in\mathcal{G}_l} \hat{f}_j^{(l)} \phi_j$$

with indices $\mathcal{G}_l := \{0,\dots,2^l-1\}$ and discrete nodal Fourier coefficients

$$\hat{f}_j^{(l)} := 2^{-l} \sum_{x\in\mathcal{S}_l} f(x)\phi_j^*(x). \tag{6}$$

This way, the 2^l interpolation conditions

$$f(x) = I_l f(x) \quad \text{for all } x \in \mathcal{S}_l$$

are fulfilled. In particular, from (6) and (2) one can deduce the well-known aliasing formula

$$\hat{f}_j^{(l)} = \sum_{k\in\mathbb{Z}} \hat{f}_k 2^{-l} \sum_{x\in\mathcal{S}_l} \omega_{\sigma(j)}^*(x)\omega_k(x) = \sum_{m\in\mathbb{Z}} \hat{f}_{\sigma(j)+m2^l}. \tag{7}$$

Next, let us consider the case of multivariate functions. To this end, let the Fourier series $\sum_{\mathbf{k}\in\mathbb{Z}^n} \hat{f}_{\mathbf{k}} \omega_{\mathbf{k}}$ be pointwise convergent to f. Then, according to the tensor product structure of the n-dimensional spaces, we introduce the n-dimensional interpolation operator on full grids as

$$I_{\mathbf{l}} := I_{l_1} \otimes \cdots \otimes I_{l_n} : V \to V_{\mathbf{l}} : f \mapsto I_{\mathbf{l}} f = \sum_{\mathbf{j}\in\mathcal{G}_{\mathbf{l}}} \hat{f}_{\mathbf{j}}^{(\mathbf{l})} \phi_{\mathbf{j}},$$

with

$$\mathcal{G}_{\mathbf{l}} := \mathcal{G}_{l_1} \times \cdots \times \mathcal{G}_{l_n} \subset \mathbb{N}_0^n$$

and multi-dimensional discrete nodal Fourier coefficients

$$\hat{f}_{\mathbf{j}}^{(\mathbf{l})} := 2^{-|\mathbf{l}|_1} \sum_{\mathbf{x}\in\mathcal{S}_{\mathbf{l}}} f(\mathbf{x})\phi_{\mathbf{j}}^*(\mathbf{x}), \tag{8}$$

where

$$\mathcal{S}_{\mathbf{l}} := \mathcal{S}_{l_1} \times \cdots \times \mathcal{S}_{l_n} \subset \mathbb{T}^n.$$

Similar to (7) it holds the multi-dimensional aliasing formula

$$\hat{f}_{\mathbf{j}}^{(\mathbf{l})} = \sum_{\mathbf{m}\in\mathbb{Z}^n} \hat{f}_{\sigma(\mathbf{j})+\mathbf{m}2^{\mathbf{l}}}, \tag{9}$$

where $\sigma(\mathbf{j}) := (\sigma(j_1),\dots,\sigma(j_n))$ and $\mathbf{m}2^{\mathbf{l}} := \left(m_1 2^{l_1},\dots,m_n 2^{l_n}\right)$.

2.2 One-dimensional hierarchical Fourier basis representation

Now we discuss a hierarchical variant of the Fourier basis representation. Let us first consider the one-dimensional case. To this end, we introduce an univariate Fourier hierarchical basis function for $j \in \mathbb{N}_0$ by

$$\psi_j := \begin{cases} \phi_0 & \text{for } j = 0, \\ \phi_j - \phi_{2^l-1-j} & \text{for } 2^{l-1} \leq j \leq 2^l - 1,\, l \geq 1, \end{cases} \tag{10}$$

and we define the one-dimensional hierarchical Fourier basis including basis functions up to level $l \in \mathbb{N}_0$ by

$$\mathcal{B}_l^{\mathrm{h}} := \{\psi_j\}_{0 \leq j \leq 2^l-1}.$$

Let us further introduce the difference spaces

$$W_l := \begin{cases} \operatorname{span}\{\mathcal{B}_0^{\mathrm{h}}\} & \text{for } l = 0, \\ \operatorname{span}\{\mathcal{B}_l^{\mathrm{h}} \setminus \mathcal{B}_{l-1}^{\mathrm{h}}\} & \text{for } l > 0. \end{cases}$$

Note that it holds the relation $V_l = \operatorname{span}\{\mathcal{B}_l\} = \operatorname{span}\{\mathcal{B}_l^{\mathrm{h}}\}$ for all $l \in \mathbb{N}_0$. Thus, we have the direct sum decomposition $V_l = \bigoplus_{v=0}^{l} W_l$. Now, let $l \in \mathbb{N}_0$ and $u \in V_l$. Then, one can easily switch from the hierarchical representation $u = \sum_{0 \leq j \leq 2^l-1} u'_j \psi_j$ to the nodal representation $u = \sum_{0 \leq j \leq 2^l-1} u_j \phi_j$ by a linear transform. For example for $l = 0,1,2,3$, the corresponding *de-hierarchization* matrices read as

$$\begin{pmatrix}1\end{pmatrix}, \quad \begin{pmatrix}1 & -1 \\ 0 & 1\end{pmatrix}, \quad \begin{pmatrix}1 & -1 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1\end{pmatrix}, \quad \begin{pmatrix}1 & -1 & 0 & -1 & 0 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1\end{pmatrix},$$

respectively. For all $l \in \mathbb{N}_0$ the de-hierarchization matrix can be easily inverted and its determinant is equal to one. Here, the corresponding *hierarchization* matrices read as

$$\begin{pmatrix}1\end{pmatrix}, \quad \begin{pmatrix}1 & 1 \\ 0 & 1\end{pmatrix}, \quad \begin{pmatrix}1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1\end{pmatrix}, \quad \begin{pmatrix}1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1\end{pmatrix},$$

for $l = 0,1,2,3$ respectively. Simple algorithms with cost complexity $\mathcal{O}(2^l)$ for hierarchization and de-hierarchization are given in [26].

Let us now define an operator

$$\check{I}_l := (I_l - I_{l-1}) : V \to W_l, \text{ for } l \geq 0,$$

where we set $I_{-1} = 0$. Note that the image of $\breve{I}_l$ is a subspace of W_l. Hence, we define the corresponding hierarchical Fourier coefficients $\breve{f}_j$ by the unique representation

$$\breve{I}_l f = \sum_{0 \le v < l} \sum_{j \in \mathcal{J}_v} (\hat{f}_j^{(l)} - \hat{f}_j^{(l-1)}) \phi_j + \sum_{j \in \mathcal{J}_l} \hat{f}_j^{(l)} \phi_j = \sum_{j \in \mathcal{J}_l} \breve{f}_j \psi_j \tag{11}$$

with

$$\mathcal{J}_v := \begin{cases} \{0\} & \text{for } v = 0, \\ \{2^{v-1}, \ldots, 2^v - 1\} & \text{for } v \ge 1. \end{cases} \tag{12}$$

Moreover, we can write the interpolation operator associated with a level l in the form

$$\begin{aligned} I_l f &= (I_l - I_{l-1} + I_{l-1} - \cdots - I_0 + I_0 - I_{-1}) f \\ &= (\breve{I}_l + \cdots + \breve{I}_0) f \\ &= \sum_{0 \le v \le l} \sum_{j \in \mathcal{J}_v} \breve{f}_j \psi_j = \sum_{0 \le j \le 2^l - 1} \breve{f}_j \psi_j. \end{aligned}$$

In particular, let us note that the following interpolation relation holds:

$$\breve{I}_l f(x) = f(x) \quad \text{for all } x \in \mathcal{S}_l^{\mathrm{h}}$$

and

$$\breve{I}_l f(x) = 0 \quad \text{for all } x \in \mathcal{S}_{l-1},$$

where

$$\mathcal{S}_l^{\mathrm{h}} := \mathcal{S}_l \setminus \mathcal{S}_{l-1},$$

with $\mathcal{S}_{-1} := \emptyset$.

For $l \in \mathbb{N}_0$ it follows by the definitions (10) and (11) the equation

$$\sum_{0 \le v < l} \sum_{j \in \mathcal{J}_v} (\hat{f}_j^{(l)} - \hat{f}_j^{(l-1)}) \phi_j + \sum_{j \in \mathcal{J}_l} \hat{f}_j^{(l)} \phi_j = - \sum_{0 \le v < l} \sum_{j \in \mathcal{J}_v} \breve{f}_{2^l - 1 - j} \phi_j + \sum_{j \in \mathcal{J}_l} \breve{f}_j \phi_j$$

and with it for $j \in \mathcal{J}_l$ that the hierarchical Fourier coefficient $\breve{f}_j$ is equal to the discrete nodal Fourier coefficient $\hat{f}_j^{(l)}$ associated with level l. Hence in the case $l \in \mathbb{N}_0$, $j \in \mathcal{J}_l$ we obtain the relation

$$\breve{f}_j = \hat{f}_j^{(l)} = \sum_{m \in \mathbb{Z}} \hat{f}_{\sigma(j) + m 2^l} \tag{13}$$

with the help of the aliasing formula (7).

Let us remark that the one-dimensional standard Fourier basis representation is sufficient to define multi-dimensional full grids, but the hierarchical Fourier basis representation is indeed necessary for the definition of sparse grids.

2.3 Generalized sparse grids

Now we consider the case of multivariate functions. Here, we use a tensor product ansatz to construct n-dimensional basis functions as well as spaces. This way, we set $\psi_{\mathbf{j}} := \bigotimes_{d=1}^{n} \psi_{j_d}$ and $W_{\mathbf{l}} := \bigotimes_{d=1}^{n} W_{l_d}$ for $\mathbf{l} \in \mathbb{N}_0^n$. In particular, we have the direct sum decomposition

$$V = \bigoplus_{\mathbf{l} \in \mathbb{N}^n} W_{\mathbf{l}}.$$

Moreover, we define $W_{\mathcal{I}} := \bigoplus_{\mathbf{l} \in \mathcal{I}} W_{\mathbf{l}}$ for an index set $\mathcal{I} \subset \mathbb{N}_0^n$. For the general sparse grid construction, we restrict ourselves to index sets, which obey the following condition [14, 26]: An index set $\mathcal{I} \subset \mathbb{N}_0^n$ is called admissible if it holds the relation

$$\{\mathbf{v} \in \mathbb{N}_0^n : \mathbf{v} \leq \mathbf{l}\} \subset \mathcal{I}, \tag{14}$$

for all $\mathbf{l} \in \mathcal{I}$. Here, the inequality $\mathbf{v} \leq \mathbf{w}$ is to be understood componentwise, i.e. $\mathbf{v} \leq \mathbf{w} :\Leftrightarrow v_d \leq w_d$ for all $1 \leq d \leq n$. Now, for an admissible index set $\mathcal{I}$, we define generalized sparse grid spaces by

$$V_{\mathcal{I}} := \sum_{\mathbf{l} \in \mathcal{I}} V_{\mathbf{l}} = \bigoplus_{\mathbf{l} \in \mathcal{I}} W_{\mathbf{l}} = W_{\mathcal{I}}. \tag{15}$$

Due to property (14) of $\mathcal{I}$ we are able to introduce the corresponding general sparse grid trigonometric interpolation operator by

$$I_{\mathcal{I}} := \sum_{\mathbf{l} \in \mathcal{I}} \check{I}_{\mathbf{l}} : V \to V_{\mathcal{I}}, \quad \text{where } \check{I}_{\mathbf{l}} := \check{I}_{l_1} \otimes \cdots \otimes \check{I}_{l_n} : V \to W_{\mathbf{l}}.$$

This way, the associated set of interpolation points is given by

$$\mathcal{S}_{\mathcal{I}} := \bigcup_{\mathbf{l} \in \mathcal{I}} \mathcal{S}_{\mathbf{l}}^{\mathrm{h}},$$

where

$$\mathcal{S}_{\mathbf{l}}^{\mathrm{h}} := \mathcal{S}_{l_1}^{\mathrm{h}} \times \cdots \times \mathcal{S}_{l_n}^{\mathrm{h}}.$$

Let us note that this general sparse grid construction includes generalized sparse grids as introduced in [23], i.e. we may employ for $\mathcal{I}$ the index set

$$\mathcal{I}_L^T := \{\mathbf{l} : |\mathbf{l}|_1 - T|\mathbf{l}|_\infty \leq (1-T)L\}, \quad T < 1. \tag{16}$$

Hence also full grids, i.e. $\mathcal{I}_L^{-\infty} = \{\mathbf{l} : |\mathbf{l}|_\infty \leq L\}$ and conventional sparse grids, i.e. $\mathcal{I}_L := \mathcal{I}_L^0 = \{\mathbf{l} : |\mathbf{l}|_1 \leq L\}$ of level $L \in \mathbb{N}_0$ are covered as special cases. For a function f with a pointwise convergent Fourier series, the multi-dimensional hierarchical coefficients $\check{f}_{\mathbf{j}}$ are given by the unique representation

$$\check{I}_{\mathbf{l}} f = \sum_{\mathbf{j} \in \mathcal{J}_{\mathbf{l}}} \check{f}_{\mathbf{j}} \psi_{\mathbf{j}},$$

where

$$\mathcal{J}_{\mathbf{l}} := \mathcal{J}_{l_1} \times \cdots \times \mathcal{J}_{l_n}.$$

In particular, the hierarchical Fourier series

$$\sum_{\mathbf{l}\in\mathbb{N}_0^n} \sum_{\mathbf{j}\in\mathcal{J}_{\mathbf{l}}} \check{f}_{\mathbf{j}} \psi_{\mathbf{j}} \tag{17}$$

converges pointwise to f on all grids $S_{\mathbf{l}}$, $\mathbf{l} \in \mathbb{N}_0^n$. Furthermore, with the help of the multi-dimensional aliasing formula (9) a relation similar to (13) can easily be deduced, that is, for $\mathbf{l} \in \mathbb{N}_0^n$ and $\mathbf{j} \in \mathcal{J}_{\mathbf{l}}$, it holds

$$\check{f}_{\mathbf{j}}^{(\mathbf{l})} = \hat{f}_{\mathbf{j}}^{(\mathbf{l})} = \sum_{\mathbf{m}\in\mathbb{Z}^n} \hat{f}_{\sigma(\mathbf{j})+\mathbf{m}2^{\mathbf{l}}}. \tag{18}$$

According to definition (15) of the general sparse grid space $V_{\mathcal{I}}$, we can estimate its number of degrees of freedom by

$$|V_{\mathcal{I}}| \lesssim \sum_{\mathbf{l}\in\mathcal{I}} 2^{|\mathbf{l}|_1}. \tag{19}$$

Starting from relation (19) the following complexity estimate is shown in the case of the general index sets $\mathcal{I}_L^T$ of (16) in [23, 24]:

Lemma 1. *Let $L \in \mathbb{N}_0$ and $T < 1$. The number of degrees of freedom of the general sparse grid spaces $V_{\mathcal{I}_L^T}$ with respect to the discretization parameter L is*

$$\left|V_{\mathcal{I}_L^T}\right| \lesssim \sum_{\mathbf{l}\in\mathcal{I}_L^T} 2^{|\mathbf{l}|_1} \lesssim \begin{cases} 2^L & \text{for } 0 < T < 1, \\ 2^L L^{n-1} & \text{for } T = 0, \\ 2^{L\frac{T-1}{T/n-1}} & \text{for } T < 0, \\ 2^{Ln} & \text{for } T = -\infty. \end{cases} \tag{20}$$

Furthermore, analogously to the well-known case of a multi-dimensional discrete Fourier transform, we can utilize the tensor product structure of the underlying spaces and operators to efficiently compute the general sparse grid interpolant $I_{\mathcal{I}} f$ for a $f \in V$. Here, the multi-dimensional transformation is expressed in terms of one-dimensional discrete Fourier transforms, hierarchizations and de-hierarchizations of different size, cf. [26] and Algorithm 1. Note that the application of a fast Fourier transform algorithm for the computation of a one-dimensional discrete Fourier transform of length 2^l results in a computational complexity of order $\mathcal{O}(l2^l)$. Note furthermore that the complexity for a one-dimensional hierarchization or de-hierarchization of length 2^l is of linear order $\mathcal{O}(2^l)$. In this way, one can give an upper estimate of the order $\mathcal{O}(2^{|\mathbf{l}|_1}|\mathbf{l}|_1)$ for the overall computational complexity in the case of a full grid $V_{\mathbf{l}}$.

In Algorithm 1 we give a procedure to apply the general sparse grid interpolation operator $I_{\mathcal{I}}$ associated to an admissible index set $\mathcal{I}$, where we define for $d \in \{1,\dots,n\}$ the set

Algorithm 1 A procedure analog to [26] to apply the general sparse grid interpolation operator $I_{\mathcal{I}}$ for a given admissible index set $\mathcal{I}$ and given interpolation values $\{u_{\mathbf{j}} \in \mathbb{C}\}_{\mathbf{j}\in\mathcal{J}_{\mathbf{l}},\mathbf{l}\in\mathcal{I}}$ associated to the general sparse grid interpolation points $\mathcal{S}_{\mathcal{I}}$. The algorithm works in-place on the given input coefficients, where we use an additional temporary array to perform the involved one-dimensional FFTs.

for $d = 1$ to n **do**
  for all $\mathbf{l} \in \mathcal{M}_d$ **do**
    for all $\mathbf{j} \in \mathcal{J}_{l_1} \times \cdots \times \mathcal{J}_{l_{d-1}} \times \mathcal{J}_0 \times \mathcal{J}_{l_{d+1}} \times \cdots \times \mathcal{J}_{l_n}$ **do**
      One-dimensional FFT for $(u_{j_1,\dots,j_{d-1},0,j_{d+1},\dots,j_n}, \dots, u_{j_1,\dots,j_{d-1},2^{l_d}-1,j_{d+1},\dots,j_n})$
      Hierarchization for $(u_{j_1,\dots,j_{d-1},0,j_{d+1},\dots,j_n}, \dots, u_{j_1,\dots,j_{d-1},2^{l_d}-1,j_{d+1},\dots,j_n})$
    end for
  end for
end for
// At this stage, the hierarchical Fourier coefficients are given in $\{u_{\mathbf{j}}\}_{\mathbf{j}\in\mathcal{J}_{\mathbf{l}},\mathbf{l}\in\mathcal{I}}$.
for $d = n$ to 1 **do**
  for all $\mathbf{l} \in \mathcal{M}_d$ **do**
    for all $\mathbf{j} \in \mathcal{J}_{l_1} \times \cdots \times \mathcal{J}_{l_{d-1}} \times \mathcal{J}_0 \times \mathcal{J}_{l_{d+1}} \times \cdots \times \mathcal{J}_{l_n}$ **do**
      De-hierarchization for $(u_{j_1,\dots,j_{d-1},0,j_{d+1},\dots,j_n}, \dots, u_{j_1,\dots,j_{d-1},2^{l_d}-1,j_{d+1},\dots,j_n})$
    end for
  end for
end for
// Finally, the non-hierarchical sparse grid Fourier coefficients are given in $\{u_{\mathbf{j}}\}_{\mathbf{j}\in\mathcal{J}_{\mathbf{l}},\mathbf{l}\in\mathcal{I}}$.

$$\mathcal{M}_d(\mathcal{I}) := \{\mathbf{l} \in \mathcal{I} : \mathbf{l} + \mathbf{e}_d \notin \mathcal{I}\},$$

with the d-th unit vector $\mathbf{e}_d$. Now, an upper estimate for the resulting overall computational complexity $\mathcal{T}[I_{\mathcal{I}}]$ of Algorithm 1 can be easily deduced in the form

$$\begin{aligned}\mathcal{T}[I_{\mathcal{I}}] &\lesssim \sum_{d=1}^{n} \sum_{\mathbf{l}\in\mathcal{M}_d(\mathcal{I})} 2^{l_d} l_d 2^{|\mathbf{l}|_1 - l_d} = \sum_{d=1}^{n} \sum_{\mathbf{l}\in\mathcal{M}_d(\mathcal{I})} 2^{|\mathbf{l}|_1} l_d \le \sum_{d=1}^{n} l_{\max} \sum_{\mathbf{l}\in\mathcal{M}_d(\mathcal{I})} 2^{|\mathbf{l}|_1} \\ &\le n l_{\max} \sum_{\mathbf{l}\in\mathcal{I}} 2^{|\mathbf{l}|_1}, \end{aligned} \tag{21}$$

where $l_{\max} := \max_{\mathbf{l}\in\mathcal{I}} |\mathbf{l}|_\infty$. Note that the inverse operator $I_{\mathcal{I}}^{-1}$ can easily computed by performing the algorithm in a reverse way [26]. In the case of the general sparse grid index sets $\mathcal{I}_L^T$, relation (21) and Lemma 1 lead directly to the following computational complexity estimate:

Lemma 2. *Let $L \in \mathbb{N}_0$ and $T < 1$. An upper estimate for the computational complexity of the general sparse grid interpolation operator $I_{\mathcal{I}_L^T}$ with respect to the discretization parameter L is given by*

$$\mathcal{T}[I_{\mathcal{I}_L^T}] \lesssim L \sum_{\mathbf{l}\in\mathcal{I}_L^T} 2^{|\mathbf{l}|_1} \lesssim \begin{cases} L2^L & \text{for } 0 < T < 1, \\ L2^L L^{n-1} & \text{for } T = 0, \\ L2^{L\frac{T-1}{T/n-1}} & \text{for } T < 0, \\ L2^{Ln} & \text{for } T = -\infty. \end{cases}$$

Let us remark that the case $T = 0$ is already presented in [26], i.e. $\mathcal{T}[I_{\mathcal{I}_L}] = \mathcal{O}(L^n 2^L)$. Note in particular that both, the asymptotic number of degrees of freedom of V_L^T in Lemma 1 and the asymptotic computational complexity of $I_{\mathcal{I}_L^T}$ in Lemma 2, are not exponentially dependent on the dimension n in the case $0 < T < 1$.[2]

Let us finally note that, alternatively, the interpolation operator $I_{\mathcal{I}}$ can be applied using the so-called combination technique or the blending scheme [3, 13, 34]. For an admissible index set $\mathcal{I}$ it holds

$$I_{\mathcal{I}} f = \sum_{\mathbf{l} \in \mathcal{I}} r_{\mathcal{I}}(\mathbf{l}) I_{\mathbf{l}} f, \tag{22}$$

where

$$r_{\mathcal{I}}(\mathbf{l}) := \sum_{\mathbf{v} \in \{0,1\}^n} (-1)^{|\mathbf{v}|} \chi_{\mathcal{I}}(\mathbf{l} + \mathbf{v}),$$

with the characteristic function

$$\chi_{\mathcal{I}}(\mathbf{l}) := \begin{cases} 1 & \text{for } \mathbf{l} \in \mathcal{I}, \\ 0 & \text{otherwise.} \end{cases}$$

The computational complexity of the combination technique (22) can be estimated by

$$\mathcal{T}[I_{\mathcal{I}}] \lesssim \sum_{\mathbf{l} \in \mathcal{I}, r_{\mathcal{I}}(\mathbf{l}) \neq 0} 2^{|\mathbf{l}|_1} |\mathbf{l}|_1.$$

3 Approximation estimates

In this section, we first define different variants of (periodic) Sobolev spaces on the torus via Fourier series, i.e. we classify functions via the decay of their Fourier coefficients and hence by their smoothness. Then, we give approximation estimates for these spaces. Here, we will first discuss the best linear approximation error and then the approximation error of the interpolant. Based on the derived estimates we further study the resulting error and cost complexities.

3.1 Periodic Sobolev spaces

As already noted in section 2.1 we characterize the smoothness classes of a function f by the decay properties of its Fourier coefficients [28]. To this end, let $w : \mathbb{Z}^n \to \mathbb{R}_+$ be a continuous and positive weight. Then we define

$$\mathcal{H}_w(\mathbb{T}^n) := \left\{ f(\mathbf{x}) = \sum_{\mathbf{k} \in \mathbb{Z}^n} \hat{f}_{\mathbf{k}} \omega_{\mathbf{k}}(\mathbf{x}) \; : \; \|f\|_w := \sqrt{\sum_{\mathbf{k} \in \mathbb{Z}^n} w(\mathbf{k})^2 |\hat{f}_{\mathbf{k}}|^2} < \infty \right\}. \tag{23}$$

Here, e.g. for $r,t \in \mathbb{R}$ the weights

$$w(\mathbf{k}) = \lambda_{\text{iso}}(\mathbf{k})^r, \quad \text{where} \quad \lambda_{\text{iso}}(\mathbf{k}) := 1 + |\mathbf{k}|_\infty,$$

and

$$w(\mathbf{k}) = \lambda_{\text{mix}}(\mathbf{k})^t, \quad \text{where} \quad \lambda_{\text{mix}}(\mathbf{k}) := \prod_{d=1}^{n}(1+|k_d|),$$

result in the conventional isotropic Sobolev spaces $\mathcal{H}^r$ [1] and in the standard Sobolev spaces with dominating mixed smoothness $\mathcal{H}^t_{\text{mix}}$ [42], respectively. A further example is the multiplicative combination of these weights, i.e.

$$w(\mathbf{k}) = \lambda_{\text{iso}}(\mathbf{k})^r \lambda_{\text{mix}}(\mathbf{k})^t,$$

which leads to generalized Sobolev spaces of dominating mixed smoothness [23]

$$\mathcal{H}^{t,r}_{\text{mix}}(\mathbb{T}^n) := \Bigg\{ f(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbb{Z}^n} \hat{f}_{\mathbf{k}} \omega_{\mathbf{k}}(\mathbf{x}) : \\ \|f\|_{\mathcal{H}^{t,r}_{\text{mix}}} := \sqrt{\sum_{\mathbf{k}\in\mathbb{Z}^n} \left(\lambda_{\text{mix}}(\mathbf{k})^t \lambda_{\text{iso}}(\mathbf{k})^r\right)^2 |\hat{f}_{\mathbf{k}}|^2} < \infty \Bigg\}. \quad (24)$$

In particular, these spaces include the conventional spaces as special cases, i.e.

$$\mathcal{H}^r(\mathbb{T}^n) = \mathcal{H}^{0,r}_{\text{mix}}(\mathbb{T}^n) \quad \text{and} \quad \mathcal{H}^t_{\text{mix}}(\mathbb{T}^n) = \mathcal{H}^{t,0}_{\text{mix}}(\mathbb{T}^n),$$

respectively. Hence, the parameter r from equation (24) governs the isotropic smoothness, whereas t governs the mixed smoothness. The spaces $\mathcal{H}^{t,r}_{\text{mix}}$ give us a quite flexible framework for the study of problems in Sobolev spaces.

Moreover, the spaces $\mathcal{H}^{t,r}_{\text{mix}}(\mathbb{T}^n)$ can be generalized to the case of n-dimensional smoothness parameters $\mathbf{t},\mathbf{r} \in \mathbb{R}^n$ with $\mathbf{r} \geq \mathbf{0}$ [24]. To this end, for $\mathbf{t},\mathbf{r} \in \mathbb{R}^n$ with $\mathbf{r} \geq \mathbf{0}$ we set $w(\mathbf{k}) = \lambda^{(\mathbf{t})}_{\text{mix}}(\mathbf{k})\lambda^{(\mathbf{r})}_{\text{iso}}(\mathbf{k})$, where

$$\lambda^{(\mathbf{t})}_{\text{mix}}(\mathbf{k}) := \prod_{d=1}^{n}(1+|k_d|)^{t_d} \quad \text{and} \quad \lambda^{(\mathbf{r})}_{\text{iso}}(\mathbf{k}) := \sum_{d=1}^{n}(1+|k_d|)^{r_d}$$

and introduce the spaces

$$\mathcal{H}^{\mathbf{t},\mathbf{r}}_{\text{mix}}(\mathbb{T}^n) := \Bigg\{ f(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbb{Z}^n} \hat{f}_{\mathbf{k}} \omega_{\mathbf{k}}(\mathbf{x}) : \\ \|f\|_{\mathcal{H}^{\mathbf{t},\mathbf{r}}_{\text{mix}}} := \sqrt{\sum_{\mathbf{k}\in\mathbb{Z}^n} \left(\lambda^{(\mathbf{t})}_{\text{mix}}(\mathbf{k})\lambda^{(\mathbf{r})}_{\text{iso}}(\mathbf{k})\right)^2 |\hat{f}_{\mathbf{k}}|^2} < \infty \Bigg\}. \quad (25)$$

In this way, for $r \geq 0$ the spaces $\mathcal{H}^{t,r}_{\text{mix}}$ are up to norm equivalency[6] special cases of the spaces $\mathcal{H}^{\mathbf{t},\mathbf{r}}_{\text{mix}}$, i.e. $\mathcal{H}^{t,r}_{\text{mix}} = \mathcal{H}^{(t,\ldots,t),(r,\ldots,r)}_{\text{mix}}$. We use the short form $\mathcal{H}^{\mathbf{r}} := \mathcal{H}^{\mathbf{0},\mathbf{r}}_{\text{mix}}$ and $\mathcal{H}^{\mathbf{t}}_{\text{mix}} := \mathcal{H}^{\mathbf{t},\mathbf{0}}_{\text{mix}}$.

Furthermore, following [46, 50], for a set of weights $\Gamma := \{\gamma_u\}_{u\subset\{1,\ldots,n\}}$ with $\gamma_u \geq 0$ and a weight function w we introduce a weighted periodic Sobolev space by

$$\mathcal{H}^{\Gamma}_{w}(\mathbb{T}^n) := \Bigg\{ f(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbb{Z}^n} \hat{f}_{\mathbf{k}} \omega_{\mathbf{k}}(\mathbf{x}) : \\ \|f\|_{\mathcal{H}^{\Gamma}_{w}} := \sqrt{\sum_{u\subset\{1,\ldots,n\}} \frac{1}{\gamma_u} \sum_{\mathbf{l}\in\Omega_u} \sum_{\mathbf{j}\in\mathcal{J}_{\mathbf{l}}} w(\sigma(\mathbf{j}))^2 |\hat{f}_{\sigma(\mathbf{j})}|^2} < \infty \Bigg\}, \quad (26)$$

where

$$\Omega_u := \{\mathbf{l} \in \mathbb{N}^n_0 : l_d = 0 \text{ for all } d \in \{1,\ldots,n\} \setminus u\}.$$

Let us remark that the orthogonal decomposition

$$f = \sum_{u\subset\{1,\ldots,n\}} f_u, \quad \text{with } f_u := \sum_{\mathbf{l}\in\Omega_u} \sum_{\mathbf{j}\in\mathcal{J}_{\mathbf{l}}} \hat{f}_{\sigma(\mathbf{j})} \omega_{\sigma(\mathbf{j})} \quad (27)$$

is well-known in statistics under the name ANOVA (analysis of variance) [10], where f_u in particular depends on the coordinates $\{x_d\}_{d\in u}$ only. Note that for $f \in \mathcal{H}^{\Gamma}_{w}$ the weight γ_u prescribes the importance of the term f_u and hence the importance of different dimensions and of correlations between dimensions. In particular for a weight $\gamma_u \to 0$ the norm $\|f_u\|_{\mathcal{H}_w}$ is forced to be zero. If the size of terms $\|f_u\|_{\mathcal{H}_w}$ decays fast with e.g. $|u|$, then a proper restriction onto certain lower dimensional functions results in a substantial reduction in computational complexity. For example a set of weights $\Gamma_q := \{\gamma_u\}_{u\subset\{1,\ldots,n\}}$ with $\gamma_u = 0$ for all $u \subset \{1,\ldots,n\}$, $|u| > q$ results in a periodic Sobolev space of finite-order q, cf. [46, 50, 51]. Thus, all terms f_u with $|u| > q$ are either not present at all or can be neglected due to the decay with $|u|$. Then, the problem of approximating a n-dimensional function reduces to the problem of approximating q-dimensional functions.

Note further that the introduced periodic Sobolev spaces, i.e. $\mathcal{H}^{t,r}_{\text{mix}}(\mathbb{T}^n)$, $\mathcal{H}^{\mathbf{t},\mathbf{r}}_{\text{mix}}(\mathbb{T}^n)$ and $\mathcal{H}^{\Gamma}_{w}$, can be straightforward generalized to the case of many-particle spaces [17, 18, 19, 27].

[6] For $r \geq 0$ we could also use the weight $\prod_{d=1}^{n}(1+|k_d|)^t\,(\sum_{d=1}^{n}(1+|k_d|)^r)$ instead of $\prod_{d=1}^{n}(1+|k_d|)^t\,(1+|\mathbf{k}|_\infty)^r$ to define the space $\mathcal{H}^{t,r}_{\text{mix}}$. This weight is equal to the special case $\lambda^{(t,\ldots,t)}_{\text{mix}}(\mathbf{k})\lambda^{(r,\ldots,r)}_{\text{iso}}(\mathbf{k})$ and hence also the associated spaces $\mathcal{H}^{t,r}_{\text{mix}}$ and $\mathcal{H}^{(t,\ldots,t),(r,\ldots,r)}_{\text{mix}}$ would be equal. However, many of the given proofs would get more technical and thus for reasons of simplicity we restrict ourselves to the definition (24).

3.2 Best linear approximation error

In the following, we will consider the error of the best linear approximation in finite-dimensional general sparse grid discretization spaces. Here, we will restrict ourselves to some specific Sobolev spaces of dominating mixed smoothness.

For $\mathbf{l} \in \mathbb{N}_0^n$ we define an approximation operator $Q_\mathbf{l}$ with respect to the $\mathcal{L}^2$-norm by

$$Q_\mathbf{l} := Q_{l_1} \otimes \ldots \otimes Q_{l_n} : \mathcal{L}^2(\mathbb{T}^n) \to V_\mathbf{l},$$

where

$$Q_l : \mathcal{L}^2(\mathbb{T}) \to V_l \,:\, f \mapsto \sum_{0 \le j \le 2^l - 1} \hat{f}_{\sigma(j)} \phi_j.$$

For an admissible index set $\mathcal{I}$, as introduced in section 2.3, we define a general sparse grid approximation operator $Q_\mathcal{I} : \mathcal{L}^2(\mathbb{T}^n) \to V_\mathcal{I}$ by

$$Q_\mathcal{I} f := \sum_{\mathbf{l} \in \mathcal{I}} \sum_{\mathbf{j} \in \mathcal{J}_\mathbf{l}} \hat{f}_{\sigma(\mathbf{j})} \omega_{\sigma(\mathbf{j})}.$$

Now let us consider two weight functions w and $\tilde{w}$ with associated Sobolev spaces $\mathcal{H}_w$ and $\mathcal{H}_{\tilde{w}}$ and norms $\|f\|_{\mathcal{H}_w}$ and $\|f\|_{\mathcal{H}_{\tilde{w}}}$, respectively. It should hold $\mathcal{H}_w \subset \mathcal{H}_{\tilde{w}} \subset \mathcal{L}^2$ and thus $w(\mathbf{k}) \lesssim \tilde{w}(\mathbf{k})$. Then, let us consider $f \in \mathcal{H}_w(\mathbb{T}^n) \subset \mathcal{L}^2(\mathbb{T}^n)$ with the unique representation $f = \hat{f}_\mathbf{k} \omega_\mathbf{k}$. Now, if $\max_{\mathbf{l} \in \mathbb{Z}^n \setminus \mathcal{I}, \mathbf{j} \in \mathcal{J}_\mathbf{l}} \frac{\tilde{w}(\sigma(\mathbf{j}))^2}{w(\sigma(\mathbf{j}))^2} < \infty$, we obtain for the best linear approximation in $V_\mathcal{I}$ the estimate

$$\begin{aligned}
\inf_{\tilde{f} \in V_\mathcal{I}} \|f - \tilde{f}\|^2_{\mathcal{H}_{\tilde{w}}} \le \|f - Q_\mathcal{I} f\|^2_{\mathcal{H}_{\tilde{w}}} &= \| \sum_{\mathbf{l} \in \mathbb{Z}^n \setminus \mathcal{I}} \sum_{\mathbf{j} \in \mathcal{J}_\mathbf{l}} \hat{f}_{\sigma(\mathbf{j})} \omega_{\sigma(\mathbf{j})} \|^2_{\mathcal{H}_{\tilde{w}}} \\
&= \sum_{\mathbf{l} \in \mathbb{Z}^n \setminus \mathcal{I}} \sum_{\mathbf{j} \in \mathcal{J}_\mathbf{l}} \tilde{w}(\sigma(\mathbf{j}))^2 |\hat{f}_{\sigma(\mathbf{j})}|^2 \\
&= \sum_{\mathbf{l} \in \mathbb{Z}^n \setminus \mathcal{I}} \sum_{\mathbf{j} \in \mathcal{J}_\mathbf{l}} \frac{\tilde{w}(\sigma(\mathbf{j}))^2}{w(\sigma(\mathbf{j}))^2} |\hat{f}_{\sigma(\mathbf{j})}|^2 w(\sigma(\mathbf{j}))^2 \\
&\le \left(\max_{\mathbf{l} \in \mathbb{Z}^n \setminus \mathcal{I}, \mathbf{j} \in \mathcal{J}_\mathbf{l}} \frac{\tilde{w}(\sigma(\mathbf{j}))^2}{w(\sigma(\mathbf{j}))^2} \right) \sum_{\mathbf{l} \in \mathbb{Z}^n \setminus \mathcal{I}} \sum_{\mathbf{j} \in \mathcal{J}_\mathbf{l}} |\hat{f}_{\sigma(\mathbf{j})}|^2 w(\sigma(\mathbf{j}))^2 \\
&\le \left(\max_{\mathbf{l} \in \mathbb{Z}^n \setminus \mathcal{I}, \mathbf{j} \in \mathcal{J}_\mathbf{l}} \frac{\tilde{w}(\sigma(\mathbf{j}))^2}{w(\sigma(\mathbf{j}))^2} \right) \sum_{\mathbf{l} \in \mathbb{Z}^n} \sum_{\mathbf{j} \in \mathcal{J}_\mathbf{l}} |\hat{f}_{\sigma(\mathbf{j})}|^2 w(\sigma(\mathbf{j}))^2 \\
&= \left(\max_{\mathbf{l} \in \mathbb{Z}^n \setminus \mathcal{I}, \mathbf{j} \in \mathcal{J}_\mathbf{l}} \frac{\tilde{w}(\sigma(\mathbf{j}))^2}{w(\sigma(\mathbf{j}))^2} \right) \|f\|^2_{\mathcal{H}_w}. \qquad (28)
\end{aligned}$$

This general result allows us to derive error estimates for a wide range of situations. We shortly consider two specific cases, namely the pairings $(\mathcal{H}^{t',r'}_{\text{mix}}, \mathcal{H}^{t,r}_{\text{mix}})$ and $(\mathcal{H}^r, \mathcal{H}^\mathbf{t}_{\text{mix}})$.

First, for the linear approximation in general sparse grid spaces $V_{\mathcal{I}_L^T}$ with the index $\mathcal{I}_L^T$ set given in (16) the following error estimate for functions in optimized Sobolev spaces of dominating mixed smoothness $\mathcal{H}_{\text{mix}}^{t,r}$ can be derived:

Lemma 3. *For $L \in \mathbb{N}_0$, $T < 1$, $t' + r' < t + r$, $t - t' \geq 0$ and $f \in \mathcal{H}_{\text{mix}}^{t,r}(\mathbb{T}^n)$ it holds:*

$$\inf_{\tilde{f} \in V_{\mathcal{I}_L^T}} \|f - \tilde{f}\|_{\mathcal{H}_{\text{mix}}^{t',r'}} \leq \|f - Q_{\mathcal{I}_L^T} f\|_{\mathcal{H}_{\text{mix}}^{t',r'}} \lesssim \begin{cases} 2^{L\left((r'-r)-(t-t')+\left(T(t-t')-(r'-r)\right)\frac{n-1}{n-T}\right)} \|f\|_{\mathcal{H}_{\text{mix}}^{t,r}} & \text{for } T \geq \frac{r'-r}{t-t'}, \\ 2^{L\left((r'-r)-(t-t')\right)} \|f\|_{\mathcal{H}_{\text{mix}}^{t,r}} & \text{for } T \leq \frac{r'-r}{t-t'}. \end{cases}$$

Proof. According to (28), the estimation of

$$\max_{\mathbf{l} \in \mathbb{Z}^n \setminus \mathcal{I}, \mathbf{j} \in \mathcal{J}_\mathbf{l}} \frac{\lambda_{\text{mix}}(\sigma(\mathbf{j}))^{t'} \lambda_{\text{iso}}(\sigma(\mathbf{j}))^{r'}}{\lambda_{\text{mix}}(\sigma(\mathbf{j}))^{t} \lambda_{\text{iso}}(\sigma(\mathbf{j}))^{r}}$$

leads to the desired result, see also [24, 27]. □

Second, for $\mathbf{t} \geq \mathbf{1}$ we can define an anisotropic admissible index set by

$$\mathcal{I}_L^{\mathbf{t}} := \{\mathbf{l} \in \mathbb{N}_0^n : \sum_{d=1}^n t_d l_d \leq L\},$$

see also [21, 20]. Then, the following error estimate can be derived analogously to Lemma 3:

Lemma 4. *For $L \in \mathbb{N}_0$, $\mathbf{t} > \mathbf{0}$, $f \in \mathcal{H}_{\text{mix}}^{\mathbf{t}}$, $|\mathbf{t}|_{min} - r > 0$ and with $\mathbf{t}' := \frac{\mathbf{t}}{|t|_{min}}$, where $|\mathbf{t}|_{min} := \min_{d=1}^n t_d$, it holds:*

$$\inf_{\tilde{f} \in V_{\mathcal{I}_L^{\mathbf{t}'}}} \|f - \tilde{f}\|_{\mathcal{H}^r} \leq \|f - Q_{\mathcal{I}_L^{\mathbf{t}'}} f\|_{\mathcal{H}_{\text{mix}}^{\mathbf{t}}} \lesssim 2^{-(|\mathbf{t}|_{min} - r)} \|f\|_{\mathcal{H}_{\text{mix}}^{\mathbf{t}}}.$$

Note that, for e.g. $t_1 = |\mathbf{t}|_{min} < t_2 \leq \ldots \leq t_n$, the number of degrees of freedom $|V_{\mathcal{I}_L^{\mathbf{t}'}}|$ is of order $\mathcal{O}(2^L)$, which then results in an overall complexity rate which is independent of the number of dimensions n.

So far we have considered the best linear approximation of a function. However, its coefficients are given by Fourier integrals (3), which can only be evaluated by analytic formulae in special cases. In practice, one possibility to compute an approximation to the best linear approximation, is the numerical calculation of the interpolant of the function. Here, however, it is in general not clear if the associated approximation error exhibits the same convergence rate as that of the best linear approximation. This issue will be discussed in the next section.

3.3 Approximation error of interpolant

In the following we will consider the error of the approximation by trigonometric *interpolation*. To this end, let us first recall the following two lemmata:

Lemma 5. *For $L \in \mathbb{N}_0$, $f \in \mathcal{H}^s$, $s > \frac{n}{2}$ and $0 \le r < s$ it holds:*

$$\|f - I_{\mathcal{I}_L^{-\infty}} f\|_{\mathcal{H}^r} \lesssim 2^{-(s-r)L} \|f\|_{\mathcal{H}^s}. \tag{29}$$

Lemma 6. *For $L \in \mathbb{N}_0$, $f \in \mathcal{H}^t_{\text{mix}}$, $t > \frac{1}{2}$ and $0 \le r < t$ it holds:*

$$\|f - I_{\mathcal{I}_L^0} f\|_{\mathcal{H}^r} \lesssim 2^{-(t-r)L} L^{n-1} \|f\|_{\mathcal{H}^t_{\text{mix}}}. \tag{30}$$

Let us remark that analogous lemmata are given in [36, 37] based on the works of [40] and [48], respectively. We give proofs based on the estimation of the aliasing error in the appendix.

Next, we will extend Lemma 6 to the case of general sparse grids. To this end, we will estimate the hierarchical surplus. Let $f \in \mathcal{H}_w$ obey a pointwise convergent (hierarchical) Fourier series. Then, the relation

$$\begin{aligned}
\|f - I_{\mathcal{I}} f\|_{\mathcal{H}_{\tilde{w}}} &= \|\sum_{\mathbf{l} \in \mathbb{N}_0^n} \sum_{\mathbf{j} \in \mathcal{J}_\mathbf{l}} \breve{f}_\mathbf{j} \psi_\mathbf{j} - \sum_{\mathbf{l} \in \mathcal{I}} \sum_{\mathbf{j} \in \mathcal{J}_\mathbf{l}} \breve{f}_\mathbf{j} \psi_\mathbf{j}\|_{\mathcal{H}_{\tilde{w}}} \\
&= \|\sum_{\mathbf{l} \in \mathbb{N}_0^n \setminus \mathcal{I}} \sum_{\mathbf{j} \in \mathcal{J}_\mathbf{l}} \breve{f}_\mathbf{j} \psi_\mathbf{j}\|_{\mathcal{H}_{\tilde{w}}} \\
&\le \sum_{\mathbf{l} \in \mathbb{N}_0^n \setminus \mathcal{I}} \|\sum_{\mathbf{j} \in \mathcal{J}_\mathbf{l}} \breve{f}_\mathbf{j} \psi_\mathbf{j}\|_{\mathcal{H}_{\tilde{w}}}
\end{aligned}$$

holds. By definition of the hierarchical basis we obtain

$$\begin{aligned}
\|\sum_{\mathbf{j} \in \mathcal{J}_\mathbf{l}} \breve{f}_\mathbf{j} \psi_\mathbf{j}\|^2_{\mathcal{H}_{\tilde{w}}} &= \|\sum_{\mathbf{j} \in \mathcal{J}_\mathbf{l}} \sum_{\mathbf{v} \in \{0,1\}^n} \breve{f}_\mathbf{j} \bigotimes_{d=1}^n \phi_{\mu^{l_d}_{v_d}(j_d)}\|^2_{\mathcal{H}_{\tilde{w}}} \\
&= \sum_{\mathbf{j} \in \mathcal{J}_\mathbf{l}} \sum_{\mathbf{v} \in \{0,1\}^n, \mathbf{l}-\mathbf{v} \ge \mathbf{0}} |\breve{f}_\mathbf{j}|^2 \tilde{w}(\sigma(\mu^\mathbf{l}_\mathbf{v}(\mathbf{j})))^2 \\
&= \sum_{\mathbf{j} \in \mathcal{J}_\mathbf{l}} \sum_{\mathbf{v} \in \{0,1\}^n, \mathbf{l}-\mathbf{v} \ge \mathbf{0}} \left| \sum_{\mathbf{m} \in \mathbb{Z}^n} \hat{f}_{\sigma(\mathbf{j}) + \mathbf{m} 2^\mathbf{l}} \right|^2 \tilde{w}(\sigma(\mu^\mathbf{l}_\mathbf{v}(\mathbf{j})))^2 \\
&= \sum_{\mathbf{j} \in \mathcal{J}_\mathbf{l}} \sum_{\mathbf{v} \in \{0,1\}^n, \mathbf{l}-\mathbf{v} \ge \mathbf{0}} \left| \sum_{\mathbf{m} \in \mathbb{Z}^n} \hat{f}_{\sigma(\mathbf{j}) + \mathbf{m} 2^\mathbf{l}} \frac{w(\sigma(\mathbf{j}) + \mathbf{m} 2^\mathbf{l})}{w(\sigma(\mathbf{j}) + \mathbf{m} 2^\mathbf{l})} \right|^2 \tilde{w}(\sigma(\mu^\mathbf{l}_\mathbf{v}(\mathbf{j})))^2,
\end{aligned}$$

where $\mu^l_0(j) = j$,

$$\mu^l_1(j) = \begin{cases} -1 & \text{if } l \le 0, \\ 2^l - 1 - j & \text{if } l \ge 1, \end{cases}$$

$\mu_{\mathbf{v}}^{\mathbf{l}} = (\mu_{v_1}^{l_1}, \ldots, \mu_{v_n}^{l_n})$ and $\phi_{-1} = 0$. With Cauchy-Schwarz it follows

$$\left| \sum_{\mathbf{m}\in\mathbb{Z}^n} \hat{f}_{\sigma(\mathbf{j})+\mathbf{m}2^{\mathbf{l}}} \frac{w(\sigma(\mathbf{j})+\mathbf{m}2^{\mathbf{l}})}{w(\sigma(\mathbf{j})+\mathbf{m}2^{\mathbf{l}})} \right|^2 \leq \left(\sum_{\mathbf{m}\in\mathbb{Z}^n} \left| \hat{f}_{\sigma(\mathbf{j})+\mathbf{m}2^{\mathbf{l}}} w(\sigma(\mathbf{j})+\mathbf{m}2^{\mathbf{l}}) \right|^2 \right) \left(\sum_{\mathbf{m}\in\mathbb{Z}^n} w(\sigma(\mathbf{j})+\mathbf{m}2^{\mathbf{l}})^{-2} \right) \quad (31)$$

and hence it holds

$$\| \sum_{\mathbf{j}\in\mathcal{J}_{\mathbf{l}}} \breve{f}_{\mathbf{j}} \psi_{\mathbf{j}} \|_{\mathcal{H}_{\tilde{w}}}^2 \leq \sum_{\mathbf{j}\in\mathcal{J}_{\mathbf{l}}} \sum_{\mathbf{v}\in\{0,1\}^n, \mathbf{l}-\mathbf{v}\geq\mathbf{0}} \left(\sum_{\mathbf{m}\in\mathbb{Z}^n} \left| \hat{f}_{\sigma(\mathbf{j})+\mathbf{m}2^{\mathbf{l}}} \right|^2 \left| w(\sigma(\mathbf{j})+\mathbf{m}2^{\mathbf{l}}) \right|^2 \right) \times \times \left(\sum_{\mathbf{m}\in\mathbb{Z}^n} w(\sigma(\mathbf{j})+\mathbf{m}2^{\mathbf{l}})^{-2} \right) \tilde{w}(\sigma(\mu_{\mathbf{v}}^{\mathbf{l}}(\mathbf{j})))^2.$$

Now, let us assume that there is a function $g : \mathbb{N}_0^n \to \mathbb{R}$ such that it holds

$$\tilde{w}(\sigma(\mu_{\mathbf{v}}^{\mathbf{l}}(\mathbf{j})))^2 \sum_{\mathbf{m}\in\mathbb{Z}^n} \left| w(\sigma(\mathbf{j})+\mathbf{m}2^{\mathbf{l}}) \right|^{-2} \leq C^2 g(\mathbf{l})^2 \quad (32)$$

for all $\mathbf{j} \in \mathcal{J}_{\mathbf{l}}$ and $\mathbf{v} \in \{0,1\}^n, \mathbf{l}-\mathbf{v} \geq \mathbf{0}$ with a constant C independent of $\mathbf{j}$ and $\mathbf{v}$. Then, with $|\{0,1\}^n| = 2^n$, we have

$$\| \sum_{\mathbf{j}\in\mathcal{J}_{\mathbf{l}}} \breve{f}_{\mathbf{j}} \psi_{\mathbf{j}} \|_{\mathcal{H}_{\tilde{w}}} \leq 2^n C g(\mathbf{l}) \left(\sum_{\mathbf{j}\in\mathcal{J}_{\mathbf{l}}} \sum_{\mathbf{m}\in\mathbb{Z}^n} |\hat{f}_{\sigma(\mathbf{j})+\mathbf{m}2^{\mathbf{l}}}|^2 w(\sigma(\mathbf{j})+\mathbf{m}2^{\mathbf{l}})^2 \right)^{\frac{1}{2}} \lesssim g(\mathbf{l}) \|f\|_{\mathcal{H}_w}$$

and hence

$$\|f - I_{\mathcal{I}} f\|_{\mathcal{H}_{\tilde{w}}} \lesssim \sum_{\mathbf{l}\in\mathbb{N}_0^n \setminus \mathcal{I}} g(\mathbf{l}) \|f\|_{\mathcal{H}_w}. \quad (33)$$

Let us now consider the approximation error in the $\mathcal{H}^r$-norm for approximating $f \in \mathcal{H}_{\text{mix}}^t$ in the sparse grid space $V_{\mathcal{I}_L^T}$ by interpolation. To this end, let us first recall the following upper bound:

Lemma 7. *For $L \in \mathbb{N}_0$, $T < 1$, $r < t$ and $t \geq 0$ it holds:*

$$\sum_{\mathbf{l}\in\mathbb{N}_0^n \setminus \mathcal{I}_L^T} 2^{-t|l|_1 + r|l|_\infty} \lesssim \begin{cases} 2^{-\left((t-r)+(Tt-r)\frac{n-1}{n-T}\right)L} L^{n-1} & \text{for } T \geq \frac{r}{t}, \\ 2^{-(t-r)L} \|f\|_{\mathcal{H}_{\text{mix}}^t} & \text{for } T < \frac{r}{t}. \end{cases}$$

Proof. A proof is given in Theorem 4 in [34]. □

Now, we can give the following lemma:

Lemma 8. *Let $L \in \mathbb{N}_0$, $T < 1$, $r < t$, $t > \frac{1}{2}$ and $f \in \mathcal{H}^t_{\text{mix}}$ with a pointwise convergent Fourier series. Then it holds:*

$$\|f - I_{\mathcal{I}_L^T} f\|_{\mathcal{H}^r} \lesssim \begin{cases} 2^{-\left((t-r)+(Tt-r)\frac{n-1}{n-T}\right)L} L^{n-1} \|f\|_{\mathcal{H}^t_{\text{mix}}} & \text{for } T \geq \frac{r}{t}, \\ 2^{-(t-r)L} \|f\|_{\mathcal{H}^t_{\text{mix}}} & \text{for } T < \frac{r}{t}. \end{cases} \tag{34}$$

Proof. For $t > \frac{1}{2}$, $\mathbf{j} \in \mathcal{J}_{\mathbf{l}}$ and $\mathbf{v} \in \{0,1\}^n$ with $\mathbf{l} - \mathbf{v} \geq \mathbf{0}$ it follows the relation

$$\begin{aligned} \sum_{\mathbf{m} \in \mathbb{Z}^n} \prod_{d=1}^n (1 + |\sigma(j) + m_d 2^{l_d}|)^{-2t} &\lesssim \sum_{\mathbf{m} \in \mathbb{Z}^n} \prod_{d=1}^n \left(2^{l_d}(1 + |m_d|)\right)^{-2t} \\ &\lesssim 2^{-2t|l|_1} \sum_{\mathbf{m} \in \mathbb{Z}^n} \prod_{d=1}^n (1 + |m_d|)^{-2t} \\ &\lesssim 2^{-2t|l|_1} \end{aligned}$$

and hence

$$(1 + |\sigma(\mu_{\mathbf{v}}^{\mathbf{l}}(\mathbf{j}))|_\infty)^{2r} \sum_{\mathbf{m} \in \mathbb{Z}^n} \prod_{d=1}^n (1 + |\sigma(j) + m_d 2^{l_d}|)^{-2t} \lesssim 2^{-t|l|_1 + r|l|_\infty}.$$

According to (32) and (33) this yields

$$\|f - I_{\mathcal{I}_L^T} f\|_{\mathcal{H}^r} \lesssim \sum_{\mathbf{l} \in \mathbb{N}_0^n \setminus \mathcal{I}_L^T} 2^{-t|l|_1 + r|l|_\infty} \|f\|_{\mathcal{H}^t_{\text{mix}}}$$

and with Lemma 7 we obtain the desired result. □

Let us remark that for regular sparse grids, i.e. $T = 0$, there is a difference in the error behavior between the best approximation and the approximation by interpolation. That is, in the $\mathcal{L}^2$-norm error estimate for the interpolant resulting from Lemma 8 with $t > \frac{1}{2}$, $r = 0$ and $T = 0$, there is a logarithmic factor present, i.e. L^{n-1}. In contrast, for the best linear approximation error in the $\mathcal{L}^2$-norm, there is no logarithmic term L^{n-1} involved according to Lemma 3 with $t > 0$, $t' = r' = r = 0$ and $T = 0$.

3.4 Convergence rates with respect to the cost

Now, we cast the estimates on the degrees of freedom and the associated error of approximation by interpolation into a form which measures the error with respect to the involved degrees of freedom. In the following, we will restrict ourselves to special cases, where the rates are independent of the dimension:

Lemma 9. *Let $L \in \mathbb{N}_0$, $0 < r < t$, $t > \frac{1}{2}$, $0 < T < \frac{r}{t}$, and $f \in \mathcal{H}^t_{\text{mix}}$ with a pointwise convergent Fourier series. Then it holds:*

$$\|f - I_{\mathcal{I}_L^T} f\|_{\mathcal{H}^r} \lesssim M^{-(t-r)} \|f\|_{\mathcal{H}^t_{\text{mix}}},$$

with respect to the involved number of degrees of freedom $M := |V_{\mathcal{I}_L^T}|$.

Proof. This is a simple consequence of the Lemmata 1 and 8. First, we use the relation (20), that is

$$M = |V_{\mathcal{I}_L^T}| \leq c_1(n) \cdot 2^L$$

for $0 < T < \frac{r}{t}$, which results in $2^{-L} \leq c_1(n) M^{-1}$. We now plug this into (34), i.e. into the relation

$$\|f - I_{\mathcal{I}_L^T} f\|_{\mathcal{H}^r} \leq c_2(n) \cdot 2^{-L(t-r)} \cdot \|f\|_{\mathcal{H}^t_{\text{mix}}}$$

and arrive at the desired result with the order constant $C(n) = c_1(n)^{t-r} \cdot c_2(n)$. □

Analogously, we can measure the error with respect to the computational complexity, which results in the following upper estimate:

Lemma 10. *For* $0 < r < t$, $t > \frac{1}{2}$, $0 < T < \frac{r}{t}$, *and* $f \in \mathcal{H}^t_{\text{mix}}$ *with a pointwise convergent Fourier series, it holds:*

$$\|f - I_{\mathcal{I}_L^T} f\|_{\mathcal{H}^r} \lesssim R^{-(t-r)} \log(R)^{t-r} \|f\|_{\mathcal{H}^t_{\text{mix}}},$$

with respect to the involved computational costs $R := \mathcal{T}[I_{\mathcal{I}_L^T}]$.

Proof. This is a simple consequence of the Lemmata 2 and 8. Analogously to the proof of Lemma 9 the relation $\|f - I_{\mathcal{I}_L^T} f\|_{\mathcal{H}^r} \lesssim R^{-(t-r)} L^{t-r} \|f\|_{\mathcal{H}^t_{mix}}$, can been shown, which yields the desired result. □

Note that in [18] a result analogous to Lemma 9 is shown in the case of measuring the best linear approximation error with respect to the involved degrees of freedom.

Finally, let us discuss shortly two cases of regular sparse grids with involved logarithmic terms. First, again a simple consequence of the Lemmata 2 and 8 is that for $L \in \mathbb{N}_0$, $t > \frac{1}{2}$ and $f \in \mathcal{H}^t_{\text{mix}}$ with a pointwise convergent Fourier series it holds the relation

$$\|f - I_{\mathcal{I}_L^0} f\|_{\mathcal{L}^2} \lesssim M^{-t} L^{(t+1)(n-1)} \|f\|_{\mathcal{H}^t_{\text{mix}}} \lesssim M^{-t} \log(M)^{(t+1)(n-1)} \|f\|_{\mathcal{H}^t_{\text{mix}}}. \quad (35)$$

Second, for the case $\frac{1}{2} < t - 1$ the relation

$$\|f - I_{\mathcal{I}_L^0} f\|_{\mathcal{H}^1} \lesssim M^{-(t-1)} L^{(t-1)(n-1)} \|f\|_{\mathcal{H}^t_{\text{mix}}} \lesssim M^{-(t-1)} \log(M)^{(t-1)(n-1)} \|f\|_{\mathcal{H}^t_{\text{mix}}} \quad (36)$$

can be derived.

3.5 Further generalizations of sparse grids

In the following we will give some brief remarks on periodic Sobolev spaces with finite-order weights and dimension-adaptive approaches.

3.5.1 Finite-order spaces

First, we consider so-called periodic Sobolev spaces with finite-order weights. These spaces are a special case of the weighted periodic Sobolev space $\mathcal{H}_w^{\Gamma}$, which we introduced in section 3.1 by definition (26). Let us recall from section 3.1 that a set of weights $\Gamma_q = \{\gamma_u\}_{u \subset \{1,\ldots,n\}}$ is denoted to be of finite order $q \in \mathbb{N}_0$, if it holds $\gamma_u = 0$ for all $\gamma_u \in \Gamma$ with $|u| > q$. Note that there are $\binom{n}{q} \lesssim n^q$ possible subsets $u \subset 1, \ldots, n$ of order q. Therefore, the problem of the approximation of a n-dimensional function $f \in \mathcal{H}_w^{\Gamma_q}(\mathbb{T}^n)$ is reduced to the problem of the approximation of $\mathcal{O}(n^q)$ functions of dimension q. Hence, in that case the curse of dimensionality can be broken and in particular for $w(\mathbf{k}) = \lambda_{\text{iso}}(\mathbf{k})^r \lambda_{\text{mix}}(\mathbf{k})^t$ the previous lemmata can be straightforwardly adapted, compare also [27].

3.5.2 Dimension-adaptive approach

So far we considered admissible index sets $\mathcal{I}_T^L$ and their associated generalized sparse grid spaces $V_{\mathcal{I}_T^L}$, which are chosen such that the corresponding a-priori estimated approximation error for *all* functions in a specific function *class* of dominating mixed smoothness is as small as possible for a given amount of degrees of freedom [6]. The goal of a so-called dimension-adaptive approach is to find an admissible index set such that the corresponding approximation error for a *single* given function is as small as possible for a prescribed amount of degrees of freedom. To this end, a scheme similar to that given in [14] for dimension-adaptive quadrature could be applied.

Such a scheme starts with an a-priori chosen small admissible index set, e.g. $\mathcal{I}^{(0)} = \{\mathbf{0}\}$. The idea is to extend the index set successively such that the index sets remain admissible and that an error reduction as large as possible is achieved. To this end, a so called *error indicator* is computed for each index $\mathbf{l} \in \mathbb{N}_0^n$ and its associated subspace $W_{\mathbf{l}}$. In the case of approximation by interpolation we use the hierarchical surplus to derive an error indicator [29], e.g. $\eta_{\mathbf{l}} = |\sum_{\mathbf{j} \in \mathcal{J}_{\mathbf{l}}} \check{f}_{\mathbf{j}} \Psi_{\mathbf{j}}|$ with appropriate norm. For further details of the dimension-adaptive sparse grid approximation algorithm we refer to [5, 14, 29].

Let us note furthermore that in [16] a relation between the dimension-adaptive sparse grid algorithm and the concept of ANOVA-like decompositions was established. There, it was also shown that general sparse grids correspond to just a hierarchical approximation of the single terms in the ANOVA decomposition (27) see also [11, 22, 38].

Note finally that various locally adaptive sparse grid approaches exist [6, 12, 29] which are based on hierarchical multilevel sparse grid techniques. But, together with fast discrete Fourier transforms, they are not easy to apply at all.

4 Numerical experiments and results

We implemented the generalized sparse grid trigonometric interpolation operator $I_{\mathcal{I}}$ for general admissible index sets $\mathcal{I} \subset \mathbb{N}_0^n$ according to Algorithm 1 in a software library called *HCFFT*. This library also includes the functionality for the application of dimension-adaptive approaches. In addition to the fast discrete Fourier transform, which we deal with in this paper, it includes actually the following variants: fast discrete sine transform, fast discrete cosine transform and fast discrete Chebyshev transform.

In the following, we present the results of some numerical calculations performed by the *HCFFT* library. We restrict ourselves to the case of the FFT based application of the interpolation operator $\mathcal{I}_L^T$. Here, we in particular study the dependence of the convergence rates on the number of dimensions for the regular sparse grid case $T = 0$ and the energy-norm like sparse grid case $T > 0$. To this end, we consider the approximation by interpolation of functions in the periodic Sobolev spaces of dominating mixed smoothness $\mathcal{H}^t_{\text{mix}}(\mathbb{T}^n)$. As test cases we use the functions

$$G_p : \mathbb{T}^n \to \mathbb{R} : \mathbf{x} \mapsto \bigotimes_{d=1}^{n} g_p(x_d)$$

with

$$g_p : \mathbb{T} \to \mathbb{R} : x \mapsto N_p \cdot (2 + \operatorname{sgn}(x - \pi) \cdot \sin(x)^p)$$

for $p = 1, 2, 3, 4$. Here, sgn denotes the sign function, i.e.

$$\operatorname{sgn}(x) := \begin{cases} -1 & x < 0, \\ 0 & x = 0, \\ 1 & x > 0 \end{cases}$$

and N_p denotes a normalization constant such that $\|g_p\|_{\mathcal{L}^2} = 1$. Note that for $\varepsilon > 0$ we have $g_p \in \mathcal{H}^{\frac{1}{2}+p-\varepsilon}(\mathbb{T})$ and thus $G_p \in \mathcal{H}^{\frac{1}{2}+p-\varepsilon}_{\text{mix}}(\mathbb{T}^n)$. In particular, the $\mathcal{L}^2$- and $\mathcal{H}^1$-error can be computed by analytic formulae and the relative $\mathcal{L}^2$-error is equal to the absolute $\mathcal{L}^2$-error, i.e. $\|G_p - I_{\mathcal{I}_L^T} G_p\|_{\mathcal{L}^2} / \|G_p\|_{\mathcal{L}^2} = \|G_p - I_{\mathcal{I}_L^T} G_p\|_{\mathcal{L}^2}$. Let us note these test functions are of simple product form, but the decay behavior of its Fourier coefficients reflects that of the considered Sobolev spaces of dominating mixed smoothness. The numerical results for more complicated functions of non-product structure from these Sobolev spaces were basically the same.

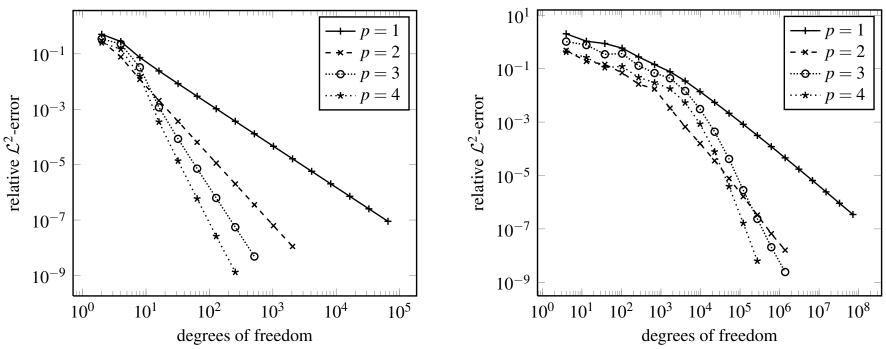

Fig. 1 Convergence behavior for approximating the functions $G_p \in \mathcal{H}_{\text{mix}}^{\frac{1+2p}{2}-\varepsilon}$ by trigonometric interpolation on regular spare grids, i.e. $\|G_p - I_{\mathcal{I}_L^0} G_p\|_{\mathcal{L}^2}$ versus $|V_{\mathcal{I}_L^0}|$. *Left:* Case of $n=1$. *Right:* Case of $n=3$.

For validation we first performed numerical calculations in the one dimensional case for G_p with $p = 1,2,3,4$. We show the measured error versus the number of degrees of freedom in Figure 1. To estimate the respective convergence rates, we computed a linear least square fit to the results of the three largest levels. This way, we obtained rates of values about 1.50, 2.50, 3.51 and 4.40, respectively, which coincide with the theoretically expected rates in the one-dimensional case, cf. Lemmata 1 and 8. Then, we performed calculations for the three dimensional case. The values $p = 1,2,3,4$ result in numerically measured convergence rates of about 1.25, 1.87, 2.83 and 3.90, respectively. Moreover, for the approximation of the test functions G_2

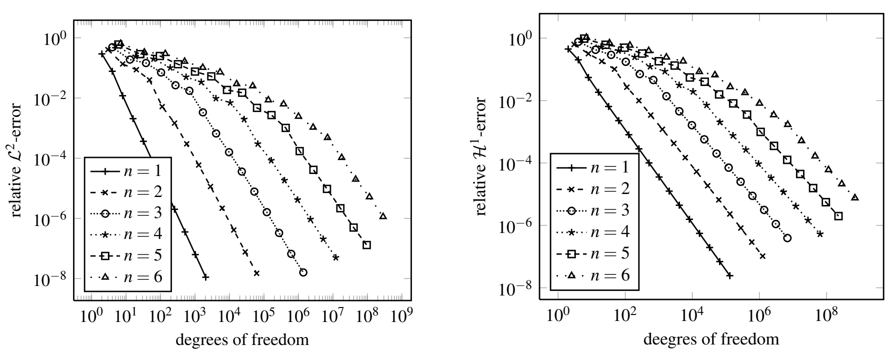

Fig. 2 Convergence behavior for approximating the function $G_2 \in \mathcal{H}_{\text{mix}}^{\frac{5}{2}-\varepsilon}$ by trigonometric interpolation on regular spare grids. *Left:* Case of relative/absolute $\mathcal{L}^2$-error, i.e. $\|G_2 - I_{\mathcal{I}_L^0} G_2\|_{\mathcal{L}^2}$. *Right:* Case of relative $\mathcal{H}^1$-error, i.e. $\|G_2 - I_{\mathcal{I}_L^0} G_2\|_{\mathcal{H}^1} / \|G_2\|_{\mathcal{H}^1}$.

for up to six dimensions by trigonometric interpolation on regular sparse grids, we observe that the rates indeed decrease with the number of dimensions, see Figure 2.

For example in case of the $\mathcal{L}^2$-error the rates deteriorate from a value of 2.50 for $n = 1$ to a value of 1.56 for $n = 6$. All calculated rates are given in Table 1. Note that this decrease in the rates with respect to the number of dimensions is to be expected from theory, since the cost and error estimates in Lemmata 1 and 8 involve dimension-dependent logarithmic terms for the regular sparse grid case, i.e. for $T = 0$. We additionally give in Table 1 the computed rates associated to the relative errors

Table 1 Numerically measured convergence rates with respect to the number of degrees of freedom according to the relative $\mathcal{L}^2$-norm error and the relative $\mathcal{H}^1$-norm for the approximation of the function $G_2 \in \mathcal{H}^{5/2-\varepsilon}_{\text{mix}}$ by trigonometric interpolation on regular spare grids, i.e. $T = 0$. In addition, we present the rates according to the relative error divided by the respective logarithmic term versus the number of degrees of freedom, see also estimates (35) and (36).

	$n=1$	$n=2$	$n=3$	$n=4$	$n=5$	$n=6$
$\mathcal{L}^2$-norm	2.50	2.17	1.87	1.73	1.60	1.56
$\mathcal{L}^2$-norm / $L^{(\frac{5}{2}+1)(n-1)}$	2.50	2.55	2.49	2.55	2.60	2.76
$\mathcal{H}^1$-norm	1.50	1.38	1.30	1.23	1.19	1.15
$\mathcal{H}^1$-norm / $L^{(\frac{5}{2}-1)(n-1)}$	1.50	1.57	1.51	1.48	1.55	1.44

divided by the respective logarithmic term versus the number of degrees of freedom. Here, the derived values fit quite well to the rates which could be expected from theory, that is, 2.5 and 1.5 for the error measured in the $\mathcal{L}^2$-norm and the $\mathcal{H}^1$-norm, respectively.

Note that according to Lemma 9, we can get rid of the logarithmic terms in some special cases. For example, if we measure the error of the approximation of G_2 by the general sparse grid interpolant $I_{\mathcal{I}_L^T} G_2$ in the $\mathcal{H}^1$-norm, then Lemma 9 leads with $r = 1$, $t = \frac{5}{2} - \varepsilon$ to a convergence rate of $\frac{3}{2} - \varepsilon$ for $0 < T < \frac{2}{3+2\varepsilon}$. Hence, we performed numerical calculations for the generalized sparse grids with $T = \frac{1}{8}$ and $T = \frac{1}{4}$. The obtained errors are plotted in Figure 3. The results show that the rates are substantially improved compared to the regular sparse grid case. We give all measured rates in Table 2. Note that we still observe a slight decrease of the rates with the number of dimensions. This is surely a consequence of the fact that we are still in the pre-asymptotic regime for the higher-dimensional cases. Note furthermore that the constant involved in the complexity estimate in Lemma 9 probably depends exponentially on the number n of dimensions. This explains the offset of the convergence with rising n in Figure 3.

In [34] it is noted that the involved order constant in the convergence rate estimate for the case $0 < T < 1$ is typically increasing with n and T and it is in particular larger than in the case of regular sparse grids with $T = 0$. In contrast, under certain assumptions, the convergence rate is superior in the case $0 < T < 1$ to that of the regular sparse grid with $T = 0$. Hence, in the pre-asymptotic regime, the effects of constants and order rates counterbalance each other a bit in practice. For example, let

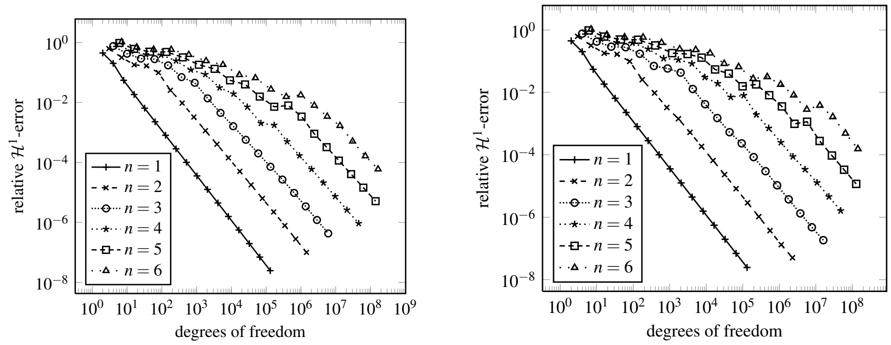

Fig. 3 Convergence behavior for approximating the function $G_2 \in \mathcal{H}_{\text{mix}}^{5/2-\varepsilon}$ by the general sparse grid interpolation operator $I_{\mathcal{I}_L^T}$ with respect to the relative $\mathcal{H}^1$-error. *Left:* Case of $T = \frac{1}{8}$. *Right:* Case of $T = \frac{1}{4}$.

Table 2 Numerically measured convergence rates with respect to the number of degrees of freedom for the approximation of the function $G_2 \in \mathcal{H}_{\text{mix}}^{5/2-\varepsilon}$ by trigonometric interpolation on generalized spare grids with $0 < T < \frac{2}{3}$.

error	T	$n=1$	$n=2$	$n=3$	$n=4$	$n=5$	$n=6$
$\mathcal{H}^1$	$\frac{1}{8}$	1.50	1.44	1.39	1.29	1.28	1.27
$\mathcal{H}^1$	$\frac{1}{4}$	1.50	1.41	1.36	1.39	1.37	1.49

us consider the $\mathcal{H}^1$-error of the interpolant $I_{\mathcal{I}_L^T} G_1$ for $T = 0, \frac{1}{8}, \frac{1}{4}$ and $n = 3, 4$. The associated computed rates are given in Table 3. Here, a break-even point can be seen from our numerical results depicted in Figure 4, i.e. for $n = 4$ the computed $\mathcal{H}^1$-error is slightly smaller in the case $T = \frac{1}{4}$ than in the cases $T = 0$ and $T = \frac{1}{8}$ for a number of involved degrees of freedom greater than about $|V_{\mathcal{I}_L^T}| \approx 10^6$. A similar effect is also present, albeit barely visible, for $n = 3$ and $|V_{\mathcal{I}_L^T}| \approx 10^5$. Nevertheless, in any case, the various rates are nearly the same anyway and these differences are quite small.

Table 3 Numerically measured convergence rates with respect to the number of degrees of freedom of the approximation of the function $G_1 \in \mathcal{H}_{\text{mix}}^{3/2-\varepsilon}$ by trigonometric interpolation on regular and generalized spare grids.

error	T	$n=3$	$n=4$
$\mathcal{H}^1$	0.0	0.45	0.42
$\mathcal{H}^1$	$\frac{1}{8}$	0.47	0.44
$\mathcal{H}^1$	$\frac{1}{4}$	0.49	0.47

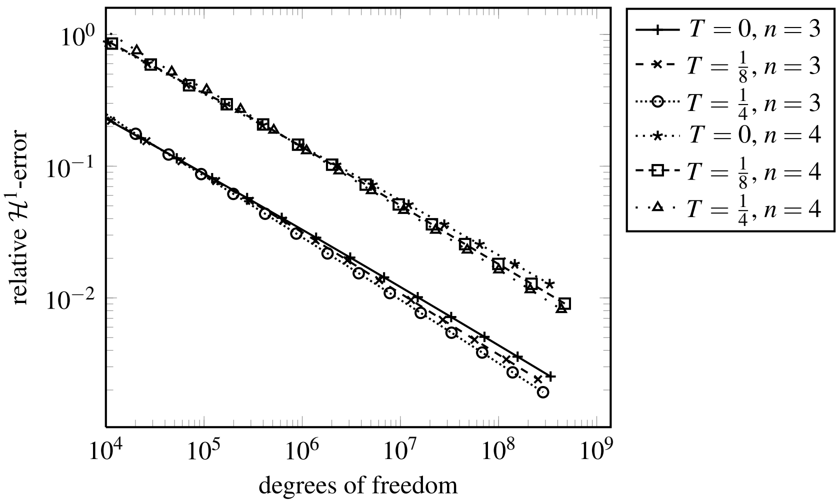

Fig. 4 Convergence behavior for the approximation of the function $G_1 \in \mathcal{H}_{\mathrm{mix}}^{3/2-\varepsilon}$ by trigonometric interpolation on generalized spare grids, i.e. $\|G_1 - I_{\mathcal{I}_L^T} G_1\|_{\mathcal{L}^2}$ versus $|V_{\mathcal{I}_L^T}|$, where the error is measured in the relative $\mathcal{H}^1$-error, i.e. $\|G_1 - I_{\mathcal{I}_L^T} G_1\|_{\mathcal{H}^1} / \|G_1\|_{\mathcal{H}^1}$.

5 Concluding Remarks

In this article, we discussed several variants of periodic Sobolev spaces of dominating mixed smoothness and we constructed the general sparse grid discretization spaces $V_{\mathcal{I}_L^T}$. We gave estimates for their number of degrees of freedom and the best linear approximation error for multivariate functions in $\mathcal{H}_{\mathrm{mix}}^{t,r}(\mathbb{T}^n)$ and $\mathcal{H}_{\mathrm{mix}}^{\mathbf{t}}(\mathbb{T}^n)$. In addition, we presented an algorithm for the general sparse grid interpolation based on the fast discrete Fourier transform and showed its computational cost and the resulting error estimates for the general sparse grid interpolant $I_{\mathcal{I}_L^T}$ of functions in $\mathcal{H}_{\mathrm{mix}}^{t}(\mathbb{T}^n)$. Specifically, we identified smoothness assumptions that make it possible to choose $I_{\mathcal{I}_L^T}$ in such a way that the number of degrees of freedom is $\mathcal{O}(2^L)$ compared to $\mathcal{O}(2^L L^{n-1})$ and $\mathcal{O}(2^{nL})$ for the regular sparse grid and full grid spaces, respectively, while keeping the optimal order of approximation. For this case, we also showed that the *asymptotic* computational cost complexities rates are independent of the number of dimensions. The constants involved in the $\mathcal{O}$-notation may still depend exponentially on n however.

Let us finally note that we mainly discussed the sparse grid interpolation operator $I_{\mathcal{I}_L^T}$ in the present paper. However, our implemented software library HCFFT allows us to deal with discretization spaces associated with arbitrary admissible index sets and in particular also features dimension-adaptive methods. Furthermore, discrete cosine, discrete sine and discrete Chebyshev based interpolation can be applied. We presently work on its extension to polynomial families which are commonly used in the area of uncertainty quantification [38]. We will discuss these approaches and its applications in a forthcoming paper.

Acknowledgements This work was supported in part by the Collaborative Research Centre 1060 of the Deutsche Forschungsgemeinschaft.

Appendix

In the following, we will give proofs for Lemmata 5 and 6 based on the estimation of the aliasing error.

Let $f \in \mathcal{H}_w$ obey a pointwise convergent Fourier series. Then, it holds the relation

$$\begin{aligned}\|f - I_{\mathcal{I}} f\|_{\mathcal{H}_{\tilde{w}}} &= \|f - I_{\mathcal{I}} f + Q_{\mathcal{I}} f - Q_{\mathcal{I}} f\|_{\mathcal{H}_{\tilde{w}}} \\ &\leq \|f - Q_{\mathcal{I}} f\|_{\mathcal{H}_{\tilde{w}}} + \|I_{\mathcal{I}} f - Q_{\mathcal{I}} f\|_{\mathcal{H}_{\tilde{w}}}.\end{aligned} \tag{37}$$

For the first term of the right hand side, an upper bound can be obtained according to (28). For the second term, it holds with (22) the relation[7]

$$\begin{aligned}\|I_{\mathcal{I}} f - Q_{\mathcal{I}} f\|_{\mathcal{H}_{\tilde{w}}} &= \|\sum_{\mathbf{l} \in \mathcal{I}} r_{\mathcal{I}}(\mathbf{l}) I_{\mathbf{l}} f - Q_{\mathcal{I}} f\|_{\mathcal{H}_{\tilde{w}}} = \|\sum_{\mathbf{l} \in \mathcal{I}} r_{\mathcal{I}}(\mathbf{l})(I_{\mathbf{l}} - Q_{\mathbf{l}}) f\|_{\mathcal{H}_{\tilde{w}}} \\ &\lesssim \sum_{\mathbf{l} \in \mathcal{I}, r_{\mathcal{I}}(\mathbf{l}) \neq 0} \|(I_{\mathbf{l}} - Q_{\mathbf{l}}) f\|_{\mathcal{H}_{\tilde{w}}}.\end{aligned}$$

With the help of the aliasing formula (9) and the Cauchy-Schwarz inequality, we obtain for $\mathbf{l} \in \mathbb{N}_0^n$ the relation

$$\begin{aligned}\|I_{\mathbf{l}} f - Q_{\mathbf{l}} f\|^2_{\mathcal{H}_{\tilde{w}}} &= \|\sum_{\mathbf{j} \in \mathcal{J}_{\mathbf{l}}} (\hat{f}^{(\mathbf{l})}_{\mathbf{j}} - \hat{f}_{\sigma(\mathbf{j})}) \phi_{\mathbf{j}}\|^2_{\mathcal{H}_{\tilde{w}}} \\ &= \sum_{\mathbf{j} \in \mathcal{J}_{\mathbf{l}}} |\hat{f}^{(\mathbf{l})}_{\mathbf{j}} - \hat{f}_{\sigma(\mathbf{j})}|^2 \tilde{w}(\sigma(\mathbf{j}))^2 \\ &= \sum_{\mathbf{j} \in \mathcal{J}_{\mathbf{l}}} |\sum_{\mathbf{m} \in \mathbb{Z}^n \setminus \{\mathbf{0}\}} \hat{f}_{\sigma(\mathbf{j}) + \mathbf{m} 2^{\mathbf{l}}}|^2 \tilde{w}(\sigma(\mathbf{j}))^2 \\ &= \sum_{\mathbf{j} \in \mathcal{J}_{\mathbf{l}}} |\sum_{\mathbf{m} \in \mathbb{Z}^n \setminus \{\mathbf{0}\}} \hat{f}_{\sigma(\mathbf{j}) + \mathbf{m} 2^{\mathbf{l}}} w(\sigma(\mathbf{j}) + \mathbf{m} 2^{\mathbf{l}})|^2 \frac{\tilde{w}(\sigma(\mathbf{j}))^2}{w(\sigma(\mathbf{j}) + \mathbf{m} 2^{\mathbf{l}})^2} \\ &\leq \sum_{\mathbf{j} \in \mathcal{J}_{\mathbf{l}}} \left(\sum_{\mathbf{m} \in \mathbb{Z}^n \setminus \{\mathbf{0}\}} |\hat{f}_{\sigma(\mathbf{j}) + \mathbf{m} 2^{\mathbf{l}}}|^2 |w(\sigma(\mathbf{j}) + \mathbf{m} 2^{\mathbf{l}})|^2 \right) \times \\ &\qquad \times \left(\sum_{\mathbf{m} \in \mathbb{Z}^n \setminus \{\mathbf{0}\}} |w(\sigma(\mathbf{j}) + \mathbf{m} 2^{\mathbf{l}})|^{-2} \right) \tilde{w}(\sigma(\mathbf{j}))^2\end{aligned} \tag{38}$$

[7] Note that analogously to (22) for $I_{\mathcal{I}}$, it holds $Q_{\mathcal{I}} f = \sum_{\mathbf{l} \in \mathcal{I}} r_{\mathcal{I}}(\mathbf{l}) Q_{\mathbf{l}} f$.

Let us assume that there is a function $F:\mathbb{N}_0^n \to \mathbb{R}$ such that it holds

$$\tilde{w}(\sigma(\mathbf{j}))^2 \sum_{\mathbf{m}\in\mathbb{Z}^n\setminus\{\mathbf{0}\}} |w(\sigma(\mathbf{j})+\mathbf{m}2^{\mathbf{l}})|^{-2} \le cF(\mathbf{l})^2 \tag{39}$$

for all $\mathbf{j}\in\mathcal{J}_{\mathbf{l}}$ with a constant c independent of $\mathbf{j}$. Then, (38) yields

$$\begin{aligned}\|(I_{\mathbf{l}}-Q_{\mathbf{l}})f\|^2_{\mathcal{H}_{\tilde{w}}} &\le cF(\mathbf{l})^2 \sum_{\mathbf{j}\in\mathcal{J}_{\mathbf{l}}} \left(\sum_{\mathbf{m}\in\mathbb{Z}^n\setminus\{\mathbf{0}\}} |\hat{f}_{\sigma(\mathbf{j})+\mathbf{m}2^{\mathbf{l}}}|^2 |w(\sigma(\mathbf{j})+\mathbf{m}2^{\mathbf{l}})|^2\right)\\ &\lesssim F(\mathbf{l})^2\|f\|^2_{\mathcal{H}_w},\end{aligned}$$

and altogether we obtain

$$\begin{aligned}\|I_{\mathcal{I}}f-Q_{\mathcal{I}}f\|_{\mathcal{H}_{\tilde{w}}} &\lesssim \sum_{\mathbf{l}\in\mathcal{I},r_{\mathcal{I}}(\mathbf{l})\neq 0} F(\mathbf{l})\|f\|_{\mathcal{H}_w}\\ &\lesssim \left(\max_{\mathbf{l}\in\mathcal{I},r_{\mathcal{I}}(\mathbf{l})\neq 0} F(\mathbf{l})\right)\left(\sum_{\mathbf{l}\in\mathcal{I},r_{\mathcal{I}}(\mathbf{l})\neq 0} 1\right)\|f\|_{\mathcal{H}_w}.\end{aligned} \tag{40}$$

Let us now consider the approximation error in the $\mathcal{H}^r$-norm for interpolating $f\in\mathcal{H}^s$, $s>\frac{n}{2}$ in the full grid space $V_{\mathcal{I}_L^{-\infty}}$ and $0\le r<s$. According to (39), we may estimate

$$\begin{aligned}\sum_{\mathbf{m}\in\mathbb{Z}^n\setminus\{\mathbf{0}\}} (1+|\sigma(\mathbf{j})+\mathbf{m}2^{\mathbf{l}})|_\infty)^{-2s} &\lesssim 2^{-2s|\mathbf{l}|_{\min}} \sum_{\mathbf{m}\in\mathbb{Z}^n\setminus\{\mathbf{0}\}} (1+|\mathbf{m}|_\infty)^{-2s}\\ &\lesssim 2^{-2s|\mathbf{l}|_{\min}} \sum_{m\in\mathbb{N}} ((m^n-(m-1)^n)|m|^{-2s}\\ &\lesssim 2^{-2s|\mathbf{l}|_{\min}} \sum_{m\in\mathbb{N}} m^{-2s+n-1}\\ &\lesssim 2^{-2s|\mathbf{l}|_{\min}}\end{aligned}$$

for all $\mathbf{j}\in\mathcal{J}_{\mathbf{l}}$. With (37), Lemma 3 and (40) we finally obtain

$$\|I_{\mathcal{I}_L^{-\infty}}f-f\|_{\mathcal{H}^r} \lesssim 2^{-(s-r)L}\|f\|_{\mathcal{H}^s}, \tag{41}$$

which proves Lemma 5.

Now, we consider the approximation error in the $\mathcal{H}^r$-norm for interpolating $f\in\mathcal{H}^t_{\text{mix}}$, $t>\frac{1}{2}$ in the sparse grid space $V_{\mathcal{I}_L}$ and $0\le r<t$. Here, according to (39), we may estimate

$$\begin{aligned}\sum_{\mathbf{m}\in\mathbb{Z}^n\setminus\{\mathbf{0}\}} \prod_{d=1}^n (1+|\sigma(j_d)+m_d2^{l_d}|)^{-2t} &\lesssim 2^{-2t|\mathbf{l}|_1} \sum_{\mathbf{m}\in\mathbb{Z}^n\setminus\{\mathbf{0}\}} \prod_{d=1}^n (1+|m_d|)^{-2t}\\ &\lesssim 2^{-2t|\mathbf{l}|_1}.\end{aligned}$$

With (37), Lemma 3, the identity

$$I_{\mathcal{I}_L} = \sum_{d=0}^{n-1} (-1)^d \binom{n-1}{d} \sum_{|\mathbf{l}|_1 = L-d} I_{\mathbf{l}}$$

and (40), we finally obtain

$$\begin{aligned} \|I_{\mathcal{I}_L} f - f\|_{\mathcal{H}^r} &\lesssim 2^{-(t-r)L} \|f\|_{\mathcal{H}^t_{\text{mix}}} \left(\sum_{\mathbf{l} \in \mathcal{I}_L, r_{\mathcal{I}_L}(\mathbf{l}) \neq 0} 1 \right) \\ &\lesssim 2^{-(t-r)L} L^{n-1} \|f\|_{\mathcal{H}^t_{\text{mix}}}, \end{aligned} \tag{42}$$

which proves Lemma 6. This is in particular a special case of Lemma 8. Let us finally remark that the estimates (41) and (42) are also shown in [36, 37] based on the works of [40] and [48], respectively.

References

1. R. Adams. *Sobolev Spaces*. Academic Press, 1975.
2. V. Barthelmann, E. Novak, and K. Ritter. High dimensional polynomial interpolation on sparse grids. *Advances in Computational Mathematics*, 12:273–288, 2000.
3. G. Baszenski and F. Delvos. A discrete Fourier transform scheme for Boolean sums of trigonometric operators. *International Series of Numerical Mathematics*, 90:15–24, 1989.
4. R. Bellmann. *Adaptive Control Processes: A Guided Tour*. Princeton University Press, 1961.
5. B. Bohn and M. Griebel. An adaptive sparse grid approach for time series predictions. In J. Garcke and M. Griebel, editors, *Sparse grids and applications*, volume 88 of *Lecture Notes in Computational Science and Engineering*, pages 1–30. Springer, 2012.
6. H. Bungartz and M. Griebel. Sparse grids. *Acta Numerica*, 13:1–123, 2004.
7. D. Dũng. Best multivariate approximations by trigonometric polynomials with frequencies from hyperbolic crosses. *Journal of Approximation Theory*, 91(2):205–225, 1997.
8. D. Dũng. Sampling recovery and cubature on sparse grids. ArXiv e-prints, 2012. arXiv:1211.4319v1 [math.NA].
9. D. Dũng and T. Ullrich. N-widths and ε-dimensions for high-dimensional approximations. *Foundations of Computational Mathematics*, 13:965–1003, 2013.
10. B. Efron and C. Stein. The Jackknife estimate of variance. *The Annals of Statistics*, 9(3):586–596, 1981.
11. C. Feuersänger. *Sparse Grid Methods for Higher Dimensional Approximation*. Dissertation, Institute for Numerical Simulation, University of Bonn, 2010.
12. C. Feuersänger and M. Griebel. Principal manifold learning by sparse grids. *Computing*, 85(4):267–299, 2009.
13. J. Garcke. *Maschinelles Lernen durch Funktionsrekonstruktion mit verallgemeinerten dünnen Gittern*. Dissertation, Institute for Numerical Simulation, University of Bonn, 2004.
14. T. Gerstner and M. Griebel. Dimension-adaptive tensor-product quadrature. *Computing*, 71(1):65–87, 2003.
15. V. Gradinaru. Fourier transform on sparse grids: Code design and the time dependent Schrödinger equation. *Computing*, 80:1–22, 2007.
16. M. Griebel. Sparse grids and related approximation schemes for higher dimensional problems. In L. Pardo, A. Pinkus, E. Suli, and M. Todd, editors, *Foundations of Computational Mathematics (FoCM05), Santander*, pages 106–161. Cambridge University Press, 2006.

17. M. Griebel and J. Hamaekers. A wavelet based sparse grid method for the electronic Schrödinger equation. In M. Sanz-Solé, J. Soria, J. Varona, and J. Verdera, editors, *Proceedings of the International Congress of Mathematicians*, volume III, pages 1473–1506, Madrid, Spain, August 22–30 2006. European Mathematical Society.
18. M. Griebel and J. Hamaekers. Sparse grids for the Schrödinger equation. *Mathematical Modelling and Numerical Analysis*, 41(2):215–247, 2007.
19. M. Griebel and J. Hamaekers. Tensor product multiscale many-particle spaces with finite-order weights for the electronic Schrödinger equation. *Zeitschrift für Physikalische Chemie*, 224:527–543, 2010.
20. M. Griebel and H. Harbrecht. A note on the construction of L-fold sparse tensor product spaces. *Constructive Approximation*, 38(2):1–17, 2013.
21. M. Griebel and H. Harbrecht. On the construction of sparse tensor product spaces. *Mathematics of Computations*, 82(268):975–994, 2013.
22. M. Griebel and M. Holtz. Dimension-wise integration of high-dimensional functions with applications to finance. *Journal of Complexity*, 26:455–489, 2010.
23. M. Griebel and S. Knapek. Optimized tensor-product approximation spaces. *Constructive Approximation*, 16(4):525–540, 2000.
24. M. Griebel and S. Knapek. Optimized general sparse grid approximation spaces for operator equations. *Mathematics of Computation*, 78:2223–2257, 2009.
25. M. Griebel, P. Oswald, and T. Schiekofer. Sparse grids for boundary integral equations. *Numerische Mathematik*, 83(2):279–312, 1999.
26. K. Hallatschek. Fourier-transform on sparse grids with hierarchical bases. *Numerische Mathematik*, 63(1):83–97, 1992.
27. J. Hamaekers. *Sparse Grids for the Electronic Schrödinger Equation: Construction and Application of Sparse Tensor Product Multiscale Many-Particle Spaces with Finite-Order Weights for Schrödinger's Equation*. Südwestdeutscher Verlag für Hochschulschriften, Saarbrücken, 2010.
28. D. Haroske and H. Triebel. *Distributions, Sobolev Spaces, Elliptic Equations*. European Mathematical Society, 2007.
29. J. Jakeman and S. Roberts. Local and dimension adaptive sparse grid interpolation and quadrature. ArXiv e-prints, 2011. arXiv:1110.0010v1 [math.NA].
30. Y. Jiang and Y. Xu. Fast discrete algorithms for sparse Fourier expansions of high dimensional functions. *Journal of Complexity*, 26(1):51–81, 2010.
31. L. Kämmerer and S. Kunis. On the stability of the hyperbolic cross discrete Fourier transform. *Numerische Mathematik*, 117:581–600, 2011.
32. L. Kämmerer, S. Kunis, and D. Potts. Interpolation lattices for hyperbolic cross trigonometric polynomials. *Journal of Complexity*, 28(1):76–92, 2012.
33. A. Klimke and B. Wohlmuth. Algorithm 847: spinterp: Piecewise multilinear hierarchical sparse grid interpolation in MATLAB. *ACM Transactions on Mathematical Software*, 31(4):561–579, 2005.
34. S. Knapek. *Approximation und Kompression mit Tensorprodukt-Multiskalenräumen*. Dissertation, University of Bonn, 2000.
35. S. Knapek. Hyperbolic cross approximation of integral operators with smooth kernel. 2000. Technical Report 665, SFB 256, University of Bonn.
36. F. Kupka. *Sparse Grid Spectral Methods for the Numerical Solution of Partial Differential Equations with Periodic Boundary Conditions*. PhD thesis, University of Wien, 1997.
37. F. Kupka. Sparse grid spectral methods and some results from approximation theory. In C. Lai, P. Bjørstad, M. Cross, and O. Widlund, editors, *Proceedings of the 11th International Conference on Domain Decomposition Methods in Greenwich*, pages 57–64, England, 1999.
38. O. Le Maître and O. Knio. *Spectral Methods for Uncertainty Quantification*, volume XVI of *Scientific Computation*. Springer, 2010.
39. X. Ma and N. Zabaras. An adaptive hierarchical sparse grid collocation algorithm for the solution of stochastic differential equations. *Journal of Computational Physics*, 228(8):3084–3113, 2009.

40. J. Pasciak. Spectral and pseudospectral methods for advection equations. *Math. Comput.*, 35:1081–1092, 1980.
41. H. Schmeisser and W. Sickel. Spaces of functions of mixed smoothness and approximation from hyperbolic crosses. *Journal of Approximation Theory*, 128(2):115–150, 2004.
42. H. Schmeisser and H. Triebel. *Fourier Analysis and Functions Spaces*. John Wiley, 1987.
43. J. Shen and L. Wang. Sparse spectral approximations of high-dimensional problems based on hyperbolic cross. *SIAM Journal on Numerical Analysis*, 48(3):1087–1109, 2010.
44. W. Sickel and F. Sprengel. Interpolation on sparse grids and tensor products of Nikol'skij–Besov spaces. *Journal of Computational Analysis and Applications*, 1:263–288, 1999.
45. W. Sickel and T. Ullrich. Tensor products of Sobolev–Besov spaces and applications to approximation from the hyperbolic cross. *Journal of Approximation Theory*, 161(2):748–786, 2009.
46. I. Sloan, X. Wang, and H. Woźniakowski. Finite-order weights imply tractability of multivariate integration. *Journal of Complexity*, 20(1):46–74, 2004.
47. S. Smolyak. Quadrature and interpolation formulas for tensor products of certain classes of functions. *Soviet Math. Dokl.*, 4:240–243, 1963. Russian original in Dokl. Akad. Nauk SSSR, 148 (1963), pp. 1042–1045.
48. V. Temlyakov. *Approximation of Periodic Functions*. Nova Science, New York, 1993.
49. T. von Petersdorff and C. Schwab. Numerical solution of parabolic equations in high dimensions. *Mathematical Modelling and Numerical Analysis*, 38:93–127, 2004.
50. G. Wasilkowski and H. Woźniakowski. Finite-order weights imply tractability of linear multivariate problems. *Journal of Approximation Theory*, 130(1):57–77, 2004.
51. A. Werschulz and H. Woźniakowski. Tractability of quasilinear problems I: General results. *Journal of Approximation Theory*, 145(2):266–285, 2007.