SPARSE REPRESENTATION OF MULTIVARIATE FUNCTIONS BASED ON DISCRETE POINT EVALUATIONS

Dissertation

zur

Erlangung des Doktorgrades (Dr. rer. nat)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Glenn Byrenheid

aus

Sömmerda

Bonn 2018

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn

1. Gutachter: AR Dr. Tino Ullrich

2. Gutachter: Prof. Dr. Michael Griebel

Tag der Promotion: 09.11.2018

Erscheinungsjahr: 2019

Zusammenfassung und Danksagung

Funktionen liefern einen der wichtigsten Bausteine innerhalb von Modellbeschreibungen der Wirklichkeit. Zentraler Gegenstand dieser Arbeit ist die Diskretisierung bzw. Approximation hochdimensionaler Funktionen mit dominierend gemischter Glattheit, welchen u.a. eine wichtige Bedeutung innerhalb der Quantenchemie zugeschrieben wird [128]. Wir möchten solche Funktionen mittels der Kenntnis einzelner diskreter Funktionswerte bestmöglich approximieren.

Im ersten Teil der Arbeit führen wir die Skalen der Besov-Triebel-Lizorkin Räume ein, wiederholen elementare Eigenschaften und diskutieren später benötigte Charakterisierungsmöglichkeiten ebendieser. Im Anschluss verlagert sich unser Fokus im vierten Kapitel auf das von G.Faber [38] 1908 eingeführte System von Hutfunktionen. Triebel legte mit seinem 2010 erschienenen Werk [120] den Grundstein, Besov-Räume dominierend gemischter Glattheit mit Hilfe des Abfallverhaltens von Koeffizienten der zugehörigen Faber-Schauder Entwicklungen zu beschreiben. Diese Theorie greifen wir auf und erweitern sie um die Sobolev-Triebel-Lizorkin Räume. Neu entwickelte Techniken erlauben es uns, die in [120] als Vermutung formulierten Charakterisierungsaussagen zu belegen.

Schneidet man für eine hinreichend glatte Funktion eine solche Entwicklung über Dilatationen und Translationen von Hutfunktionen endlich ab, so erhalten wir eine Approximation der Funktion, deren Koeffizienten auf Abtastwerten basieren. Im 5. Kapitel nutzen wir die gewonnenen Charakterisierungsresultate um Dünngitterapproximation [129] auf Basis des Faber-Schauder Systems zu untersuchen. Wir stellen fest, dass der untersuchte Algorithmus der asymptotisch Optimale für Abtastwerte, gewonnen auf Dünngitterpunkten, ist. Wir messen Approximationsfehler zum einen in der Norm von Lebesgue-Räumen $L_q([0,1]^d)$ als auch in der Energie-Norm $H^1([0,1]^d)$. Für letztere beweisen wir Resultate auf Basis eines schärferen Optimalitätskriteriums, dem sogenannten worst-case Fehler für Standard Information [84, 85, 86].

Im sechsten Kapitel verändert sich unser Fokus. Während wir in den bisherigen Kapiteln weitestgehend lineare Approximationsmethoden betrachtet haben, verlassen wir diese nun. Wir fragen nach der besten Approximation einer Funktionenklasse mittels Linearkombinationen von m translatierten und dilatierten Hutfunktionen. Dies nennt man beste m-Term Approximation, bezüglich des Faber-Schauder Systems. Interpretiert werden kann diese Quantität unter anderem als ein Maßstab für die Komprimierbarkeit von Funktionenklassen. Betrachten wir dieses Problem näher, so ist es vorerst gar nicht mehr notwendig über den Begriff der Information zu reden. Wir bewerten Linearkombinationen von Hutfunktionen anhand ihrer Größe und des zugehörigen Approximationsfehlers. Abhängig vom Fehlerkriterium im Zielraum liefern nichtlineare Approximationsmethoden dünnere (sparse) Darstellungen bei geringeren Approximationsfehlern, als dies vergleichsweise für lineare Dünngitterapproximatenmethoden der Fall ist. Faber-Schauder Charakterisierungen erlauben uns, Probleme aus kompliziert zu handhabenden Funktionenräumen auf einfachere Folgenräume zu übertragen. Als besonders interessant erweisen sich so genannte kleine Glattheiten, bei denen die asymptotische Approximationsrate nicht von der zugrundeliegenden Dimension des Funktionenraumes abhängt. Wir präsentieren konstruktive nichtlineare Verfahren, die es in einem geeigneten Sinne erlauben, als Abtastalgorithmen interpretiert zu werden.

Im letzten Teil der Arbeit wechseln wir in den klassischen Fall mit periodischen Randbedingungen. Wir beweisen neue trigonometrische Charakterisierungen, die weder in der Glattheit nach oben, noch in der Integrierbarkeit der Modellfunktionen nach unten beschränkt sind. Schließlich nutzen wir bestehende Resultate zum Verhalten diverser s-Zahlen [90], um die Optimalität für die L_q-Approximation mittels Dünngittersampling im Sinne des worst-case Fehlers für Standard Information zu bewerten und einen Vergleich zur Approximation mittels allgemeinerer linearer Information herzustellen.

Danksagung

An dieser Stelle möchte ich all denen Personen danken, die Anteil an der Entstehung dieser Arbeit haben.

Allen voran geht mein außerordentlicher Dank an meinen Doktorvater Tino Ullrich. Ich bedanke mich für die Überlassung eines Themas innerhalb eines sehr aktiven und spannenden Teilgebietes der Mathematik. Ausgezeichnete Betreuung mit zahlreichen Diskussionsstunden vor der Tafel in der Villa Maria legten den Grundstein zu ertragreichen Resultaten.

Weiterhin möchte ich mich bei Michael Griebel für die Übernahme des Zweitgutachtens bedanken. Auch danke ich Winfried Sickel, der mich bereits als jungen Studenten vom Themengebiet der Approximationstheorie begeistern konnte und im weiteren Verlauf stets unterstützt hat. Ein weiteres ganz großes Danke geht an Dorothee D. Haroske. Ich möchte mich unter anderem bedanken für die zur zügigen Fertigstellung dieser Arbeit bereitgestellten Freiheiten, ohne die diese Arbeit in ihrer heutigen Form nicht mehr möglich gewesen wäre. Ebenso hat Van Kien Nguyen mit sorgfältigem und unermüdlichem Korrekturlesen zum Erfolg dieser Arbeit beigetragen. Dafür möchte ich ihm meinen tiefsten Dank aussprechen.

Sebastian Mayer, Robert J. Kunsch, Toni Volkmer, Lutz Kämmerer und vielen weiteren danke ich für die angenehme Zeit, Diskussionen, gemeinsame Projekte und Tagungen welche ich mit ihnen während meiner Promotionszeit genießen durfte.

Zu guter letzt danke ich meiner Familie, allen voran Andrea und Thomas. Dank ihrer unerschöpflichen Unterstützung sind mathematische Problemstellungen die einzigsten, mit welchen ich mich vom ersten Studientag bis heute ernsthaft beschäftigen musste.

Bonn, im Juli 2018 Glenn Byrenheid

Contents

Chapter 1

Introduction

Many applications in engineering, science, and statistics require inter- or extrapolation from data. Generic examples are computer-based simulations [52], data mining [58], or forecasting [92]. Indispensable foundation in every such situation is a mathematical or statistical model. It allows to represent the underlying real-world phenomenon or at least some simplification thereof in a way suitable for computation and mathematical analysis. The model formulation very often involves multivariate functions

$$f(x_1, ..., x_d), \quad \boldsymbol{x} = (x_1, ..., x_d) \in \Omega \subseteq \mathbb{R}^d,$$

where the dimension d may be very large. The problem of inter- and extrapolation then is to find a function which fits the given data in a suitable sense. Problems of this kind are so versatile that several mathematical disciplines are devoted to them. Each uses its own language. In approximation theory and numerical analysis, the terms function identification, function recovery, and function reconstruction are common. Statisticians speak of regression, function estimation or function fitting. To learn a function is a widely used phrase in machine learning and statistical learning theory.

1.1 Functions with bounded mixed derivative or difference

A practically highly relevant model assumption is based on a bounded mixed derivative. The most classical space of functions with bounded mixed derivatives is the Sobolev space H^r_{mix}. The space consists of L_2-functions f such that certain weak derivatives $D^\gamma f = \partial^{\gamma_1}_{x_1} \cdots \partial^{\gamma_d}_{x_d} f$ are bounded in L_2. The most natural norm on H^r_{mix} is the classical Sobolev norm with dominating mixed smoothness,

$$\|f\|^2_{H^r_{\text{mix}}} := \sum_{\substack{0 \le \gamma_i \le r \\ i=1,\dots,d}} \|D^\gamma f\|^2_2 < \infty.$$

Note that this space can be defined via other equivalent norms.

As we will see in this thesis forms of dominating mixed smoothness fit very well to the application of sparse grid techniques [8]. Sparse grids are nowadays widely applied to tackle high-dimensional approximation and recovery problems. Modern examples include multivariate density estimation [51], reconstruction of manifolds [40], and uncertainty quantification [13, 82].

The electronic Schrödinger equation. This is the most prominent example where the regularity theory provides bounded mixed smoothness properties [128]. Numerical solutions of the electronic Schrödinger equation are of growing interest in computational chemistry. These allow to deduce chemical properties of molecules from computer simulations with high scientific validity. In the Born-Oppenheimer model [52], the wave functions depend on the spatial positions of the molecule's electrons $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)} \in \mathbb{R}^3$. The spatial positions of the molecule's nuclei $\boldsymbol{b}^{(1)}, \ldots, \boldsymbol{b}^{(K)} \in \mathbb{R}^3$ are fixed parameters. Thus the dimension $d = 3N$ of the wave function's domain increases with the number of electrons. This leads to very high-dimensional recovery problems already for molecules of moderate size. The wave functions are determined by an eigenvalue problem which represents the stationary Schrödinger equation,

$$Hf = \lambda f \tag{1.1.1}$$

with the electronic Hamilton operator

$$H = -\frac{1}{2}\sum_{i=1}^{N}\Delta_i - \sum_{i=1}^{N}\sum_{k=1}^{K}\frac{Z_k}{\|\boldsymbol{x}^{(i)} - \boldsymbol{b}^{(k)}\|} + \sum_{i\neq j}\frac{1}{\|\boldsymbol{x}^{(i)} - \boldsymbol{x}^{(j)}\|}.$$

Some of the physically admissible eigenfunctions (solutions of (1.1.1)) do not possess C^∞-regularity, but all have mixed hybrid type regularity $H^{t,s}_{mix}$, where $t > 0$ governs the mixed smoothness and $s > 0$ the isotropic regularity [128].

Quasi-Monte Carlo methods. This methods are a second way to naturally take benefit of a bounded mixed derivative. We consider the problem of approximating the integral of a sufficiently often differentiable $d-$variate function f by an average over function values taken at nodes $X_n = \{\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n\}$ which are chosen in advance. Hlawka-Zaremba [62] showed the following identity

$$\frac{1}{n}\sum_{k=1}^{n} f(\boldsymbol{t}_k) - \int_{[0,1]^d} f(\boldsymbol{x})d\boldsymbol{x} = \sum_{u\subset[d]}(-1)^{|u|}\int_{[0,1]^{|u|}}\mathrm{disc}(X_n, \boldsymbol{x}_u)\frac{\partial^{|u|}f(\boldsymbol{x}_u, 1)}{\partial \boldsymbol{x}_u}d\boldsymbol{x}_u$$

where

$$\mathrm{disc}(X_n, \boldsymbol{x}) := x_1 \cdot \cdots \cdot x_d - \frac{1}{N}\sum_{i=1}^{N}\chi_{[\boldsymbol{0},\boldsymbol{x}]}(\boldsymbol{t}_i), \quad \boldsymbol{x} \in [0,1]^d\,,$$

is the discrepancy function. This leads to the well-known L_2-version of the Koksma-Hlawka inequality

$$\Big|\frac{1}{n}\sum_{k=1}^{n} f(\boldsymbol{x}_k) - \int_{[0,1]^d} f(\boldsymbol{x})d\boldsymbol{x}\Big| \leq \|f\|_d \cdot \mathrm{disc}_d(X_n)\,,$$

where

$$\|f\|_d = \Big(\sum_{u \subset [d]} \int_{[0,1]^{|u|}} \Big| \frac{\partial^{|u|} f}{\partial \boldsymbol{x}_u}(\boldsymbol{x}_u, 1) \Big|^2 \, d\boldsymbol{x}_u \Big)^{1/2}$$

and

$$\mathrm{disc}_d(X_n) = \Big(\sum_{u \subset [d]} \int_{[0,1]^u} \mathrm{disc}(X_n, \boldsymbol{x}_u)^2 d\boldsymbol{x}_u \Big)^{\frac{1}{2}}.$$

Clearly, the summands on the right hand side access all mixed derivatives of order 1 in each direction. Assuming this regularity of the function it remains to study the behavior of the discrepancy function for the convergence analysis.

Best m-term approximation. Finally a third way to motivate the concept of bounded mixed derivatives/differences is best m-term approximation. Starting with a univariate wavelet system $\{\psi_{j,k}\}$ with sufficiently many vanishing moments and smoothness we consider its d-variate tensorization over all scales represented by the following dictionary

$$\Psi = \{\psi_{\boldsymbol{j},\boldsymbol{k}} = \psi_{j_1,k_1} \otimes \cdots \otimes \psi_{j_d,k_d} \ : \ \boldsymbol{j} = (j_1, ..., j_d) \in \mathbb{N}_0^d, \boldsymbol{k} = (k_1, ..., k_d) \in \mathbb{Z}^d\} \, .$$

Now we ask for the best m-term approximation space $\mathcal{A}_q^\alpha(L_p(\mathbb{R}^d))$ defined by

$$\mathcal{A}_q^\alpha(L_p(\mathbb{R}^d)) = \Big\{ f \in L_p(\mathbb{R}^d) \ : \ \Big(\sum_{m=0}^{\infty} [m^\alpha \sigma_m(f, \Psi)_p]^q \frac{1}{m} \Big)^{1/q} < \infty \Big\}.$$

Based on classical results for ℓ_p spaces, see Pietsch [89], DeVore [18] and also Temlyakov [114], it was shown in [105], [54] that in the special case $q < p = 2$ and $\alpha = 1/q - 1/2$ this approximation space can be identified as follows

$$\mathcal{A}_q^\alpha(L_2(\mathbb{R}^d)) = S_{q,q}^{1/q-1/2} B(\mathbb{R}^d) \, ,$$

which represents a dominating mixed counterpart of an isotropic result by DeVore and Popov [20]. The space on the right-hand side represents a space with bounded mixed difference (Besov space with dominating mixed smoothness, see Chapter 3 below).

Best m-term approximation (also called sparse approximation) will also play a role in this thesis. We consider indeed a wavelet type dictionary, the tensorized Faber-Schauder system. In contrast to classical wavelet systems the Faber-Schauder system does not provide vanishing moments, which causes severe technical difficulties. However due to coefficient functionals based on discrete point evaluations this system is highly relevant for applications.

1.2 Sparse grid approximation using the hierarchical Faber basis

Approximation of univariate functions. Already in 1909 G. Faber proved in [38] that every univariate continuous function f on $[0, 1]$ can be represented (uniform convergence) as a superposition of hat functions $v_{j,k}$ (see Definition 4.1) in the following way

$$f = f(0)v_{-1,0} + f(1)v_{-1,1} - \frac{1}{2}\sum_{j\in\mathbb{N}_0}\sum_{k=0}^{2^j-1}\Delta^2_{2^{-(j+1)}}f(2^{-j}k)v_{j,k}. \tag{1.2.1}$$

The required information of the function f to compute such a series expansion is only a discrete set of function values taken at the nodes $\{2^{-j}k \ : \ j\in\mathbb{N}_0, k\in\mathbb{Z}\}$. Due to limited storage and computing resources in real life applications we may work with a truncated series

$$F^1_M f = \sum_{j=-1}^{M}\sum_{k\in D_j} d_{j,k}(f)v_{j,k}. \tag{1.2.2}$$

This requires a priori knowledge of the truncation error. For functions belonging to the unit ball of the isotropic smoothness class $W^1_2([0,1])$ it is well-known [120, Theorem 4.11] that

$$\sup_{\|f|W^1_2([0,1])\|\le 1}\|f - F^1_M f|L_2([0,1]^d)\| \asymp 2^{-M}$$

holds, where $n = 2^M+1$ is the number of sample points $\{k/2^M : 0\le k\le 2^M\}$ used by $F^1_M f$.

Approximation of d-variate functions. According to [106] the Sobolev space of dominating mixed smoothness $S^1_2W(\mathbb{R}^d)$ can be written as the tensor product of univariate isotropic Sobolev spaces

$$S^1_2W(\mathbb{R}^d) = W^1_2(\mathbb{R})\otimes\cdots\otimes W^1_2(\mathbb{R}),$$

which, in particular, contains linear combinations of functions of the form (rank-1-tensor)

$$f(\boldsymbol{x}) := \prod_{i=1}^{d} f_i(x_i), \quad f_i\in W^1_2(\mathbb{R}), \quad \boldsymbol{x}\in\mathbb{R}^d.$$

Dealing with tensorized hat functions

$$v_{\boldsymbol{j},\boldsymbol{k}}(\boldsymbol{x}) := \prod_{i=1}^{d} v_{j_i,k_i}(x_i), \quad \boldsymbol{x}\in[0,1]^d,$$

one may use the operator

$$G^d_M f := \sum_{|\boldsymbol{j}|_\infty\le M}\sum_{\boldsymbol{k}\in D_{\boldsymbol{j}}} d_{\boldsymbol{j},\boldsymbol{k}}(f)v_{\boldsymbol{j},\boldsymbol{k}}$$

to approximate the d-variate function f using samples on a so-called full grid $\mathcal{G}^{full}_M := \{\boldsymbol{k}2^{-M} : 0\le k_i\le 2^M, i=1,\ldots,d\}$ with cardinality $|\mathcal{G}^{full}_M|\asymp 2^{Md}$. With techniques presented in this thesis it is not hard to show that

$$\|f - G^d_M f\|_2 \lesssim 2^{-M} \tag{1.2.3}$$

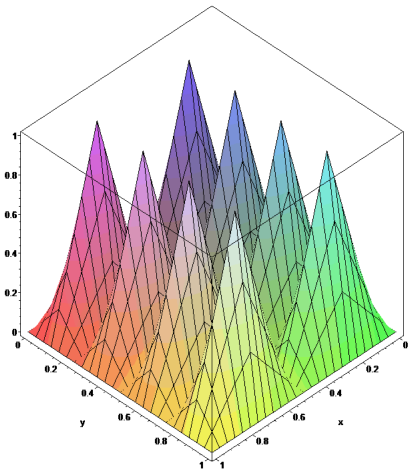


Figure 1.1: Tensorized hat functions

holds for f belonging to the unit ball of $S^1_2 W([0,1]^d)$. Consequently, the asymptotic rate of convergence (in terms of the number of sampling nodes) becomes worse $(1/d)$ with increasing problem dimensions d.

Smolyak's algorithm. An approach which overcomes this issue to some extent goes back to 1963 and started with Smolyak [108] who considered uniform approximation of multivariate functions with mixed smoothness on the basis of function values. He introduced an influential construction which is nowadays known as Smolyak's algorithm

$$T_M[L]f := \sum_{\substack{\boldsymbol{j}\in\mathbb{N}_0^d \\ |\boldsymbol{j}|_1\le M}} (L_{j_1} - L_{j_1-1}) \otimes \ldots \otimes (L_{j_d} - L_{j_d-1})f \quad , \quad M \in \mathbb{N}\,, \tag{1.2.4}$$

where the $(L_j)_{j\in\mathbb{N}_0}$ represent univariate approximation operators (put $L_{-1} := 0$). For more historical comments see [33, Sect. 5]. When applied to F^1_N this construction yields a powerful sampling (interpolation) operator for the multivariate case taking points from a so-called *sparse grid*. We obtain

$$F^d_M f(\boldsymbol{x}) := T_M[F^1]f(\boldsymbol{x}) = \sum_{|\boldsymbol{j}|_1\le M} \sum_{\boldsymbol{k}\in D_{\boldsymbol{j}}} d_{\boldsymbol{j},\boldsymbol{k}}(f) v_{\boldsymbol{j},\boldsymbol{k}}(\boldsymbol{x})\,,$$

which means we take all Faber-Schauder levels with $|\boldsymbol{j}|_1 \le M$ instead of $|\boldsymbol{j}|_\infty \le M$. This operator samples on a sparse grid $\mathcal{G}^{sparse}_M := \{(2^{-j_1}k_1, \ldots, 2^{-j_d}k_d) : |\boldsymbol{j}|_1 \le M, 0 \le k_i \le 2^{j_i}, i = 1, \ldots, d\}$. The notion sparse grid is due to Zenger [129] and comes from the fact that $|\mathcal{G}^{sparse}_M| \asymp 2^M M^{d-1}$ instead of $|\mathcal{G}^{full}_M| \asymp 2^{Md}$.

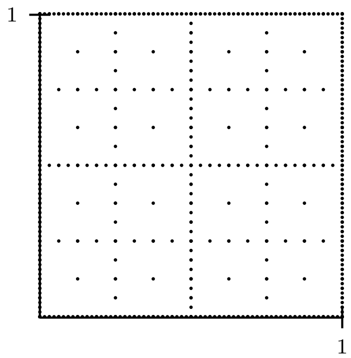


Figure 1.2: Sparse grid $d = 2$, $M = 6$

This construction allows to prove the approximation rate

$$\sup_{\|f|S^1_2W([0,1]^d)\|\leq 1} \|f - F^d_M f|L_2([0,1]^d)\| \lesssim 2^{-M} M^{\frac{d-1}{2}}.$$

Compared to (1.2.3) the convergence rate in M behaves similar, even a little worse. However, the sparse grid contains a significantly less number of sampling nodes n. In fact, taking this into account, the error bound can be written as follows

$$\sup_{\|f|S^1_2W([0,1]^d)\|\leq 1} \|f - F^d_M f|L_2([0,1]^d)\| \lesssim n^{-1} \log^{d-1} n (\log^{d-1} n)^{\frac{1}{2}}$$

whereas

$$\sup_{\|f|S^1_2W([0,1]^d)\|\leq 1} \|f - G^d_M f|L_2([0,1]^d)\| \asymp n^{-\frac{1}{d}}.$$

Hence, using the Smolyak approach, the asymptotic rates depend only in the comparatively small logarithm of n on d. Considering a trigonometric sampling operator the rate stated here was first observed by Sickel [102] ($d = 2$). Later it was extended by Sickel, Ullrich [103, 104] to $d > 2$. For more general approximation problems effects like this have been discovered much earlier in the former Soviet Union, [112, 41]. In the context of the Faber-Schauder system a similar result was discovered in [8] for H^2_{mix} with a slightly worse log-exponent. The result above can be found as a special case in [29], [107] and [120]. Let us additionally mention the following references dealing with sparse grids in an applied context [87, 8, 47, 43].

New sparse grid error bounds in L_q. The focus in the current thesis is on sparse grid approximation with the hierarchical Faber basis in Sobolev spaces $S^r_pW([0,1]^d)$ (including $p \neq 2$) where we measure approximation errors in spaces $L_q([0,1]^d)$ with $1 < p < q < \infty$. Our main result reads as follows

$$\sup_{\|f|S^r_pW([0,1]^d)\|\leq 1} \|f - F^d_M f|L_q([0,1]^d)\| \asymp 2^{-M(r-\frac{1}{p}+\frac{1}{q})}$$

for $1 < p < q < \infty$, $\frac{1}{p} < r < 2 + \frac{1}{p} - \frac{1}{q}$. The limiting case $r = 2 + 1/p - 1/q$ can be incorporated to the expense of an additional logarithmic term M^{d-1}. We show that our error analysis is optimal in the asymptotic sense and that all algorithms using samples from a sparse grid can not beat this rate. Moreover, if $1 < p < q \leq 2$ or $2 \leq p < q < \infty$ then the operator F_M^d is asymptotically optimal among all sampling algorithms. Of special interest is the important case $q = \infty$. Here we prove

$$\sup_{\|f|S_p^rW([0,1]^d)\|\leq 1} \|f - F_M^d f|L_\infty([0,1]^d)\| \asymp 2^{-M(r-\frac{1}{p})} M^{(d-1)(1-\frac{1}{p})}$$

for $1 < p < \infty$ and $\frac{1}{p} < r < 2 + \frac{1}{p}$. This improves on the rates for Faber-Schauder approximation stated in [120] significantly, which were obtained via embeddings from Besov spaces. Up to now Sobolev spaces $S_p^rW([0,1]^d)$ with fractional smoothness r were hard to handle directly in this context. We transferred the method of sparse representations or sampling representations, originally introduced by Dinh Dũng [29, 25, 26, 9] to Sobolev spaces, which allows for proving sharp estimates.

Reconstruction guarantees in the energy norm. Our second main interest in this thesis are error bounds in the energy norm $H^1([0,1]^d)$. The interest in this setting is motivated by the numerical solution of PDEs using Galerkin methods. Assume we have a PDE in variational notation

$$a(u,v) = (f,v), \quad \text{for all } v \in H^1, \tag{1.2.5}$$

with

$$a(u,v) \leq \lambda \|u|H^1\| \|v|H^1\| \quad \text{and} \quad a(u,u) \geq \mu \|u|H^1\|^2.$$

In order to get an approximate numerical solution we can consider the same problem on a finite dimensional subspace $V_h \subset H^1$

$$a(u_h,v) = (f,v), \quad \text{for all } v \in V_h. \tag{1.2.6}$$

The Lax-Milgram theorem [73] yields that the problems (1.2.5) and (1.2.6) have unique solutions u^* and u_h^*, which by Céa's lemma [12], satisfy the inequality

$$\|u^* - u_h^*|H^1\| \leq \frac{\lambda}{\mu} \inf_{v\in V_h} \|u^* - v|H^1\|.$$

One can bound the $H^1([0,1]^d)$ discretization error by best approximations from the discretization subspace. The error of best approximation for the embedding

$$S_2^rW(\mathbb{T}^d) \to H^1(\mathbb{T}^d)$$

including the explicit dimensional dependence of the constants were investigated in [34]. We follow the approach in [8] and consider sampling approximation for this embeddings in the non-periodic case and show for a sampling operator E_M^d which samples functions f on a energy sparse grid $\mathcal{G}_M^{energy} := \{(2^{-j_1}k_1, \ldots, 2^{-j_d}k_d) : r|\boldsymbol{j}|_1 - |\boldsymbol{j}|_\infty \leq M, 0 \leq k_i \leq 2^{j_i}, i = 1, \ldots, d\}$ with cardinality $|\mathcal{G}_M^{energy}| \asymp 2^{\frac{M}{r-1}}$ the asymptotical rate

$$\sup_{\|f|S_2^rW[0,1^d]\|\leq 1} \|f - E_M^d f|H^1([0,1]^d)\| \asymp 2^{-M},$$

whenever $1/2 < r \leq 2$. In other words, with n samples we produce an asymptotical convergence rate of $n^{-(r-1)}$. This result has been first stated by Bungartz and Griebel [8] (in case $p = 2$) but their proof contained some problematic arguments.

In addition, we show that this rate is optimal. Indeed, there is no better algorithm (in the sense of (1.4.2) below) using n samples having a better convergence rate than E_M^d. Actually we can prove even more, namely there is no algorithm using n pieces of linear information of f providing a better convergence rate. In fact, optimal approximation is realized by sampling.

1.3 Constructive m-term approximation with the Faber-Schauder dictionary

Whereas for Smolyak sparse grid and energy sparse grid operators the (linear) information map is fixed in advance for the whole class of functions we will also consider a different approach. We are interested in approximation methods based on non-linear algorithms especially with (adaptively) chosen samples. A possible way is to consider best m-term approximation (or sparse approximation) with respect to a given dictionary. For a given countable set $\mathcal{D} \subset Y$, called dictionary, the algorithms we consider map to a finite linear combination of elements contained in this dictionary. A first obvious question is the convergence rate of such linear combinations measured in terms of elements contained in this linear combination. This can be measured by the following benchmark quantity

$$\sigma_m(\boldsymbol{F}, \mathcal{D})_Y := \sup_{\|f|\boldsymbol{F}\| \leq 1} \inf_{\substack{(b_j)_{j\in\mathbb{N}} \subset \mathcal{D} \\ \lambda, (\lambda_j)_{j\in\mathbb{N}} \subset \mathbb{C}}} \Big\| f - \sum_{j=1}^{m} \lambda_j b_j \Big| Y \Big\|, \tag{1.3.1}$$

called best m-term approximation. For general notions and results on non-linear approximation we refer to the survey [18]. In our case $\boldsymbol{F}$ will be a Besov-Triebel-Lizorkin-Sobolev space, Y a Lebesgue space and $\mathcal{D}$ the set containing all translated and dilated Faber-Schauder hat functions. Then this quantity describes the (nonlinear) approximability of functions f belonging to the unit ball of a Besov-Triebel-Lizorkin-Sobolev space by linear combinations of m hat functions. From a numerical point of view these quantities can be interpreted as benchmark results for data compression issues which use only the function values used to compute the Faber-Schauder coefficients. A possible strategy storing a function f in a computer is provided by decomposing this function into (infinitely many) "simple" functions like wavelets or in our case hat functions belonging to the Faber-Schauder dictionary and storing only a finite number of the corresponding coefficients (e.g. the m biggest ones, all coefficients that are bigger than a certain threshold,...). This means for a fixed dictionary decomposition the best m-term approximation width serves as a benchmark quantity for the minimal error of the approximation of f by a function build on the compressed data. In the first sections it was already motivated that the Faber-Schauder dictionary $\mathbb{F}^d$ is a proven object in numerical analysis. In [44] and [45] it is used for (compressed) representation of topographic and landscape data. A main advantage compared to most wavelet type dictionaries is the simple structure of the single hat functions. It allows to write a continuous multivariate function $f \in C([0,1]^d)$ as a series

$$f = \sum_{\boldsymbol{j} \in \mathbb{N}_{-1}^d} \sum_{\boldsymbol{k} \in D_{\boldsymbol{j}}} d_{\boldsymbol{j},\boldsymbol{k}}(f) v_{\boldsymbol{j},\boldsymbol{k}}$$

with (conditional) convergence in $C([0,1]^d)$, where the coefficients $d_{\boldsymbol{j},\boldsymbol{k}}(f)$ are tensorized 2nd order differences (see (4.1.2)). Indeed, each coefficient can be computed exactly by the knowledge of at most 3^d function values of f at certain points in $[0,1]^d$. In contrast, (hyperbolic) wavelet representations (cf. [19]) require the evaluation of $L_2(\mathbb{R}^d)$ inner-products, which means integrating and extending products of f over $\mathbb{R}^d$. Assuming the real number model, in general this can be done only approximatively by numerical integration, whereas it can be computed exactly in case of the Faber Schauder dictionary. Nevertheless from a combinatorial point of view sequence spaces used to compute results for wavelet type dictionaries are very similar to the discretization spaces for the Faber-Schauder dictionary. For this reason our results are related to results known for Daubechies wavelets [54, 55, 56], Dirichlet kernels [3, 4], de la Vallée Poussin kernels [27, 28], Meyer wavelets [5] and Haar wavelets [109, 110]. A more detailed overview is given in [33, Section 7.2]. A second very popular type of dictionary studied in literature is the trigonometric system $\mathcal{T}^d = \{e^{i\boldsymbol{k}\boldsymbol{x}} : \boldsymbol{k} \in \mathbb{Z}^d\}$. Here we refer to [115, 112, 95, 63, 6] and [33, Section 7.5]. Studying $\sigma_m(S^r_pW(\mathbb{T}^d), \mathcal{T}^d)_{L_q(\mathbb{T}^d)}$ in case $1 < p < q \le 2$ it turns out that this system is less powerful compared to wavelet type dictionaries.

The case of large smoothness. Based on the Faber-Schauder representations we provide a constructive procedure depending on the parameters r, p, θ and d below which computes a m-term approximation realizing the following rates.

$$\begin{aligned} \sigma_m(S^r_{p,\theta}F([0,1]^d), \mathbb{F}^d)_{L_\infty([0,1]^d)} &= \sup_{\|f|S^r_{p,\theta}F([0,1]^d)\| \le 1} \sigma_m(f, \mathbb{F}_d)_{L_\infty} \\ &\lesssim m^{-r}(\log^{d-1} m)^{r+(1-\frac{1}{\theta})} \end{aligned} \tag{1.3.2}$$

in case of $0 < p, \theta \le \infty$ and $\max\{\frac{1}{p}, \frac{1}{\theta}\} < r < 2$. Furthermore,

$$\sigma_m(S^r_{p,\theta}B([0,1]^d), \mathbb{F}^d)_{L_\infty([0,1]^d)} \lesssim m^{-r}(\log^{d-1} m)^{r+(1-\frac{1}{\theta})} \tag{1.3.3}$$

in case $\max\{\frac{1}{p}, \frac{1}{\theta} - 1\} < r < 2$. Note, that in case $\theta = 2$ and $1 < p < \infty$ we may identify

$$S^r_{p,\theta}F([0,1]^d) = S^r_pW([0,1]^d)$$

in the sense of equivalent norms which gives the important special case

$$\sigma_m(S^r_pW(\mathbb{R}^d), \mathbb{F}^d)_{L_\infty([0,1]^d)} \lesssim m^{-r}(\log^{d-1} m)^{r+1/2}\,, \tag{1.3.4}$$

whenever $1/p < r < 2$. Surprisingly, we are able to extend the result to the limiting case $r = 2$ on the expense of a additional logarithm and obtain

$$\sigma_m(S^2_pW([0,1]^d), \mathbb{F}^d)_{L_\infty} \lesssim m^{-2}(\log^{d-1} m)^3\,. \tag{1.3.5}$$

The above mentioned procedure is constructive. Indeed, for a given function f and a desired accuracy the following level-wise greedy strategy works. We take a prescribed (finite) number of samples of the function at dyadic grid points. From this data we compute a finite number of Faber-Schauder coefficients of f. Following a levelwise greedy selection strategy we store the most important ones to build the approximating m-term. The fact that only function values of f are used allows to interpret the presented algorithm as a non-linear sampling algorithm. In [88] so called spatially adaptive sparse grids were considered. The output of our algorithms allows an interpretation as an approximant that contains samples generated on such a adaptively refined sparse grid.

New results in the small-smoothness regime. Considering the class of Besov functions in the quasi-Banach fine index range $0 < \theta < 1$ the approximation rates stated above can be complemented by the following surprising result

$$\sigma_m(S^r_{p,\theta}B([0,1]^d), \mathbb{F}^d)_{L_\infty([0,1]^d)} \asymp m^{-r}$$

in cases $\frac{1}{p} < r < \min\{2, \frac{1}{\theta} - 1\}$ or $\frac{1}{p} < r = \frac{1}{\theta} - 1 < 2$ without a d-dependent logarithmic term in the rate. Note that this result is sharp and is not even known for wavelet dictionaries. To our knowledge this is one of the first known sharp results concerning non-linear L_∞-approximation in the case of spaces with dominating mixed smoothness. Furthermore, the asymptotic approximation rate coincides with that of the univariate case, where one approximates $B^r_{p,\theta}([0,1])$ functions in $L_\infty([0,1])$ by the univariate Faber-Schauder dictionary.

1.4 Optimal sampling recovery of multivariate functions with higher regularity

The limited regularity of hat functions are responsible for the fact, that the convergence rates can not exceed 2. These limitations do not apply to the periodic setting which has been intensively studied in the former Soviet Union. We provide new trigonometric characterizations that are able to overcome the regularity restrictions. Additionally we focus towards information based complexity issues, i.e. sharp lower bounds. We ask for the worst case error of the best possible approximation of a function f while having standard information at n sampling nodes. A generalized quantity is provided by (linear) sampling widths for a class $\boldsymbol{F} \hookrightarrow C([0,1]^d)$ into a (quasi)-Banach space Y, which measure the minimal worst-case error for the (linear) sampling recovery problem with n points. To be more precise, we compare the performance of a optimal sampling algorithm with the linear sampling widths

$$\varrho^{\text{lin}}_n(\boldsymbol{F}, Y) := \inf_{X_n} \inf_{\Psi_n} \sup_{\|f|\boldsymbol{F}\| \le 1} \Big\| f - \sum_{i=1}^n f(\boldsymbol{x}^i)\psi_i(\cdot) \Big| Y \Big\| \quad , \quad n \in \mathbb{N}\,, \tag{1.4.1}$$

where the sampling nodes $X_n := \{\boldsymbol{x}^i\}_{i=1}^n \subset [0,1]^d$ and associated (continuous) functions $\Psi_n := \{\psi_i\}_{i=1}^n$ determine a linear sampling recovery algorithm which is fixed in advance for a class $\boldsymbol{F}$ of multivariate functions on $[0,1]^d$. Here the error is measured for instance in $Y = L_q([0,1]^d)$. Let us emphasize that in (1.4.1) we restrict to linear recovery algorithms, whereas we admit general recovery algorithms $\varphi : \mathbb{C}^n \to L_q$ in the definition of the (non-linear) sampling widths

$$\varrho_n(\boldsymbol{F}, Y) := \inf_{\varphi, X_n} \sup_{\|f|\boldsymbol{F}\| \le 1} \|f - \varphi(X_n(f))|Y\|, \tag{1.4.2}$$

which is also denoted as the worst-case error for standard information, see [84, Sect. 4.1]. Here $X_n(f) := (f(x_1), f(x_2), \ldots, f(x_n))$ denotes a linear information mapping and $\varphi : \mathbb{C}^n \to Y$ a (non-linear) reconstruction map. This quantities are bounded from below by Gelfand n-widths

$$c_n(\boldsymbol{F}, Y) := \inf_{\substack{B\colon \boldsymbol{F} \to \mathbb{C}^n \\ \text{linear}}} \sup_{\substack{\|f|\boldsymbol{F}\| \le 1 \\ f \in \ker B}} \|f|Y\| \tag{1.4.3}$$

This widths describe the maximal distance of 2 functions $f, g \in \boldsymbol{F}$ in Y for that a non-trivial information mapping I_n exists with

$$I_n(f) = I_n(g).$$

That means the information mapping does not see the difference between f and g. We investigate the optimal sampling recovery problem for the embedding

$$id : S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d) \to L_q(\mathbb{T}^d)\,, \tag{1.4.4}$$

where $0 < p < q \le \infty$, $0 < \theta \le \infty$ and $\boldsymbol{r} > 1/p$. Without loss of generality we assume

$$r = r_1 = \ldots = r_\mu < r_{\mu+1} \le \ldots \le r_d < \infty \quad , \quad \mu \le d. \tag{1.4.5}$$

One of the main results in this section is the sharp rate of convergence

$$\varrho_n^{\text{lin}}(S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d), L_q(\mathbb{T}^d)) \asymp \Big(\frac{(\log n)^{\mu-1}}{n}\Big)^{r-1/p+1/q} \quad , \quad n \in \mathbb{N}\,, \tag{1.4.6}$$

whenever $1 < p < q \le 2$, $1 \le \theta \le \infty$ or $2 \le p < q < \infty$, $2 \le \theta \le \infty$ and $r > 1/p$, see Corollary 8.8 below. Our main contribution is the constructive upper bound which holds true whenever $0 < p < q < \infty$, $0 < \theta \le \infty$ and $r > 1/p$. This is complemented by (see Theorem 8.4)

$$\varrho_n^{\text{lin}}(S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d), L_\infty(\mathbb{T}^d)) \lesssim \Big(\frac{(\log n)^{\mu-1}}{n}\Big)^{r-1/p} (\log n)^{(\mu-1)(1-1/p)_+} \quad , \quad n \in \mathbb{N}\,. \tag{1.4.7}$$

The upper bounds are realized by an explicit family of interpolation operators T_m^L using $n \asymp 2^m m^{\mu-1}$ function values on a (anisotropic) Smolyak grid, where the parameter $L \in \mathbb{N}$ refers to the polynomial decay of the univariate fundamental interpolant ($L = 1$ Dirichlet kernel, $L = 2$ de la Vallée Poussin type kernels, $L > 2$ higher order kernels). It turned out that, for the sampling recovery problem (1.4.4) and the upper bounds in (1.4.6), (1.4.7), (1.4.8), (1.4.9), the condition $L > 1/q$ is sufficient, which means that Smolyak's algorithm (1.2.4) applied to the classical trigonometric interpolation (based on the Dirichlet kernel (1.5.3)) does the job. For $\theta = p = 2$ in (1.4.7) this has been already observed in [9, Rem. 6.12].

Let us emphasize the important special case ($\theta = 2$), where it holds the identification $S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d) = S^{\boldsymbol{r}}_pW(\mathbb{T}^d)$ with the space of functions with bounded mixed derivative. As a corollary from (1.4.6) we obtain the new sharp rate of convergence

$$\varrho_n^{\mathrm{lin}}(S^{\boldsymbol{r}}_pW(\mathbb{T}^d), L_q(\mathbb{T}^d)) \asymp \Big(\frac{(\log n)^{\mu-1}}{n}\Big)^{r-1/p+1/q}, \quad n \in \mathbb{N}, \qquad (1.4.8)$$

in case $1 < p < q \le 2$ or $2 \le p < q < \infty$ and $\boldsymbol{r} > 1/p$ which was unknown before. The upper bound is achieved with sparse grid interpolation based on classical univariate trigonometric interpolation. In particular, this improves on the bounds stated by Triebel in [120, Thm. 4.15, Cor. 4.16] in case $r = 1$. The parameter domain where (1.4.8) holds is shown in the left diagram, where the parameters α and β refer to the following rate of convergence

$$\varrho_n(\boldsymbol{F}, L_q) \asymp \Big(\frac{(\log n)^{\mu-1}}{n}\Big)^{\alpha}(\log n)^{(\mu-1)\beta}.$$

The precise statements can be found in Sections 8.2, 8.3.

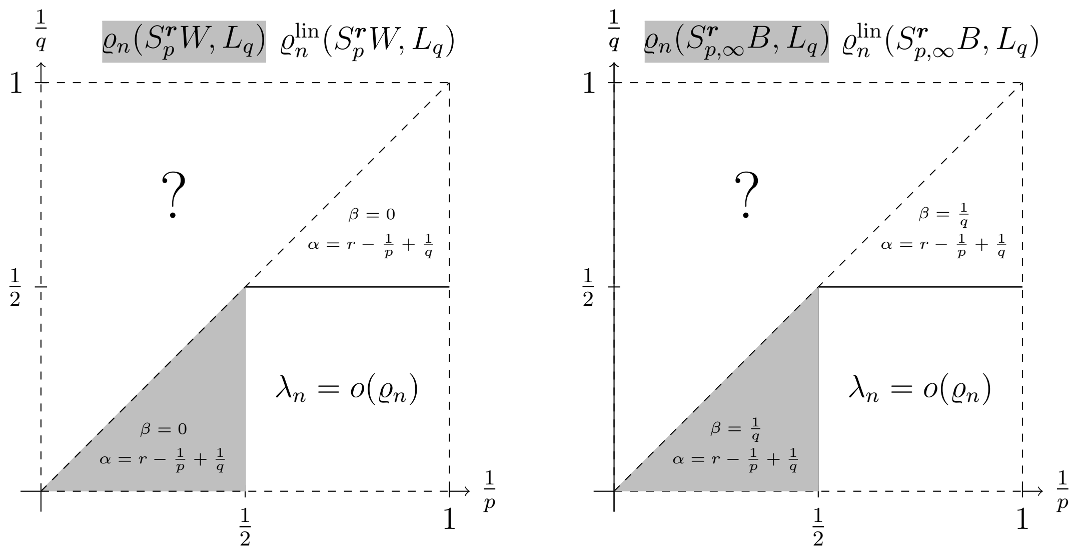


Figure 1.3: Linear and non-linear sampling widths.

We mainly contributed to the upper bounds in the left figure. Most of the results illustrated in the right figure for Hölder-Nikolskij spaces $S^{\boldsymbol{r}}_{p,\infty}B(\mathbb{T}^d)$ of mixed smoothness are well-known. Note that Open Problem 5.3 in [33] refers to the lower triangle in the right Figure 1.3. A new approach of Malykhin and Ryutin [76] settled this question for linear sampling recovery, cf. Corollary 8.12. We observed that their method also bounds the Gelfand widths from below, cf. Theorem 8.19 and Corollary 8.20.

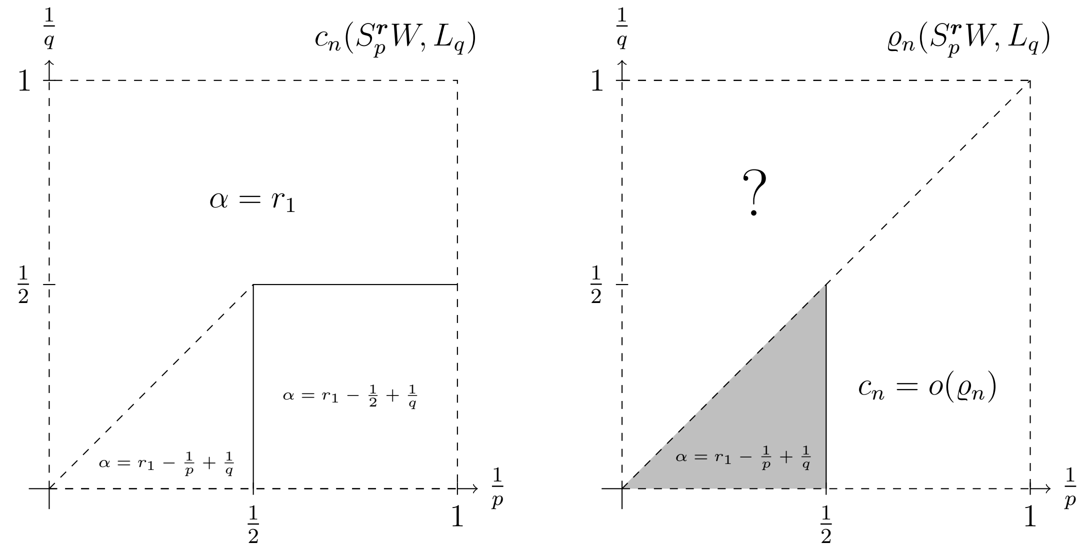


Figure 1.4: The parameter α refers to the sharp rate $(\frac{(\log n)^{\mu-1}}{n})^{\alpha}$.

This yields the optimal order also for the non-linear sampling widths (1.4.2), which is illustrated by the shaded lower triangles in Figure 1.3. The matching bound in the right upper triangle for $S^r_{p,\infty}B(\mathbb{T}^d)$ were obtained by Dinh Dũng [25, 26]. The necessary benchmark results on linear widths were obtained by Galeev [41, 42], Romanyuk [94, 96], and the recent paper by Malykhin and Ryutin [76]. Note, that all sharp upper bounds can be realized by Smolyak type operators (1.2.4), i.e. via linear interpolation on sparse grids based on univariate Dirichlet interpolation, (1.5.2), (1.5.3). What concerns Besov spaces with bounded mixed difference $S^{\boldsymbol{r}}_{p,\theta}B(\mathbb{T}^d)$ it is known that

$$\varrho_n^{\text{lin}}(S^{\boldsymbol{r}}_{p,\theta}B(\mathbb{T}^d), L_q(\mathbb{T}^d)) \asymp \Big(\frac{(\log n)^{\mu-1}}{n}\Big)^{r-1/p+1/q}(\log n)^{(\mu-1)(1/q-1/\theta)_+} \quad , \quad n \in \mathbb{N}\,, \tag{1.4.9}$$

if $1 < p < q \leq 2$, $1 \leq \theta < \infty$ and $\boldsymbol{r} > 1/p$, see [33, Thm. 4.47, 5.15] and the references therein. With our method we can show the upper bound in case $0 < p < q \leq \infty$, $0 < \theta \leq \infty$ and $r > 1/p$, see Theorem 8.6, with interpolation operators providing $L > 1/q$. Comparing to (1.4.6) there is a extra log-term in (1.4.9) in case of "large" $\theta > q$. There are still many open cases in this framework which actually lack the suitable lower bounds. Let us refer to the works by Temlyakov [113, 117] and the more recent papers Sickel, Ullrich [103, 104, 124], Dinh Dũng [29, 30], [9], as well as [33] and the references therein for upper bounds in case $p \geq q$ and the question-marked region. We emphasize that our technique allows to reproduce all those results, including the upper bound in [113], within a few lines of proof.

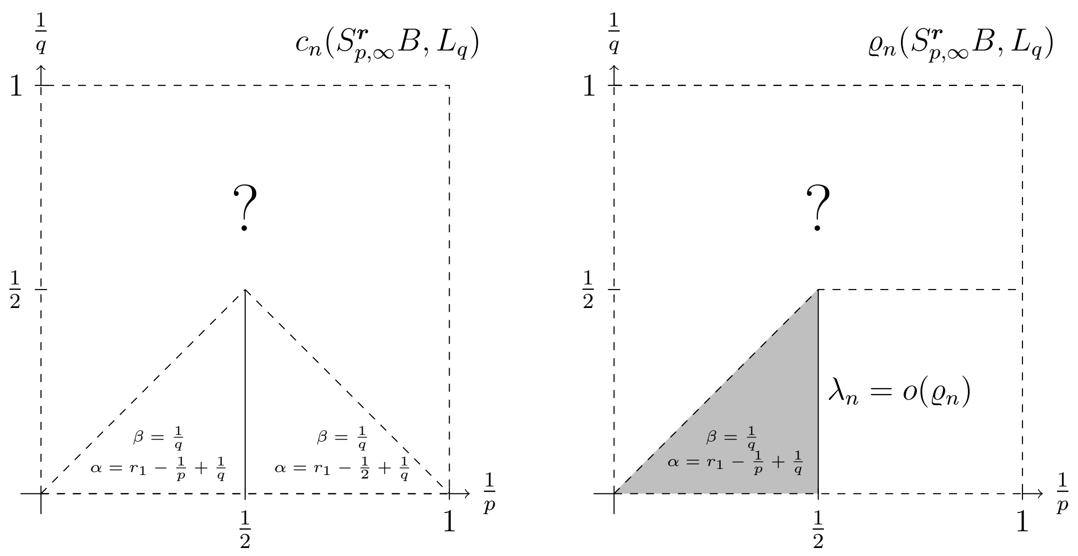


Figure 1.5: The parameters α and β refer to the sharp rate $(\frac{(\log n)^{\mu-1}}{n})^{\alpha}(\log^{\mu-1} n)^{\beta}$.

In Open Problem 18 in [84, Sect. 4.2.4] the authors conjecture the equivalence $\varrho_n^{\text{lin}} \asymp \varrho_n$ for all parameters $1 < p, q < \infty$ in case of isotropic Sobolev spaces $W_p^r(\Omega)$ on bounded Lipschitz domains Ω, see also Novak, Triebel [83] and Heinrich [59, Thms. 5.2, 5.3]. In the present paper we consider mixed smoothness periodic Sobolev embeddings. In our case, the conjecture is true if $2 \leq p < q < \infty$ for both Sobolev and Hölder-Nikolskij spaces, see the shaded regions in the diagrams above. In all other cases it is not known. Our results also support the above conjecture in the mixed smoothness setting. A similar statement as in [84, Rem. 4.18], namely the equivalence $\lambda_n \asymp \varrho_n^{\text{lin}}$ if $p < q$ are on the same side of 2 and $\lambda_n = o(\varrho_n^{\text{lin}})$ if $p < 2 < q$ is also true in our case.

1.5 Characterization in terms of discrete function evaluations

In Definition 3.18 below we introduce periodic Besov-Lizorkin-Triebel spaces of mixed smoothness via Fourier analytic building blocks $\delta_{\boldsymbol{j}}[f]$ generated by a dyadic decomposition of unity. In this thesis we aim for function space characterizations where we replace the building blocks $\delta_{\boldsymbol{j}}[f]$ by the blocks

$$q_{\boldsymbol{j}}[f] = (I_{j_1} - I_{j_1-1}) \otimes ... \otimes (I_{j_d} - I_{j_d-1})f \quad , \quad \boldsymbol{j} \in \mathbb{N}_0^d\,, \tag{1.5.1}$$

used in the classical Smolyak algorithm (see (1.2.4) above). Here the operators $(I_j)_j$ are univariate interpolation operators

$$I_j^L[f] = \sum_{u=0}^{2^j-1} f\Big(\frac{2\pi u}{2^j}\Big) K_{\pi,j}^L\Big(\,\cdot - \frac{2\pi u}{2^j}\Big)\,. \tag{1.5.2}$$

In a way we replace the usual convolution by a discrete one such that the building blocks $q_{\boldsymbol{j}}^L[f]$ are constructed out of $\asymp 2^{|\boldsymbol{j}|_1}$ function values. The parameter $L \in \mathbb{N}$ refers to the decay of the fundamental interpolant $K_{\pi,j}^L$, which represents a suitable trigonometric polynomial of degree 2^j and will be explicitly constructed in Section 7. In case $L = 1$ we have the classical univariate nested trigonometric interpolation, where $K_{\pi,j}^1 := 2^{-j}\mathcal{D}_j^1$ with $\mathcal{D}_0^1 \equiv 1$ and

$$\mathcal{D}_j^1(x) := \mathcal{D}_{2^{j-1}}(x) - e^{i2^{j-1}x} = e^{-i(2^{j-1}-1)x}\frac{e^{i2^jx}-1}{e^{ix}-1} \quad , \quad x \in \mathbb{T}\,, \tag{1.5.3}$$

for $j \in \mathbb{N}$. The parameter $L = 2$ refers to de la Vallée Poussin type operators and $L > 2$ to higher order kernels.

We will prove the following characterization for periodic Sobolev spaces of mixed smoothness if $\boldsymbol{r} > \max\{1/p, 1/2\}$ and $1 < p < \infty$

$$\|f|S_p^{\boldsymbol{r}}W(\mathbb{T}^d)\| \asymp \Big\|\Big(\sum_{\boldsymbol{j}\in\mathbb{N}_0^d} 2^{2\boldsymbol{j}\cdot\boldsymbol{r}}|q_{\boldsymbol{j}}^L[f](\cdot)|^2\Big)^{1/2}\Big|L_p(\mathbb{T}^d)\Big\|\,, \tag{1.5.4}$$

where we may use $L \geq 1$, i.e. Dirichlet type characterizations are admitted. This result provides a powerful tool to deal with Sobolev embeddings $S_p^{\boldsymbol{r}}W(\mathbb{T}^d)$ in $L_q(\mathbb{T}^d)$. Analyzing Smolyak's algorithm (1.2.4) in this context has been a technical issue in the past. With (1.5.4) and its counterpart for Triebel-Lizorkin spaces (1.5.5) it becomes a straight-forward computation. Up to certain regularity restrictions this principle works also in the non-periodic case for Faber-Schauder characterizations.

For Triebel-Lizorkin spaces we obtain the representation (see Theorem 7.14)

$$\|f|S_{p,\theta}^{\boldsymbol{r}}F(\mathbb{T}^d)\| \asymp \Big\|\Big(\sum_{\boldsymbol{j}\in\mathbb{N}_0^d} 2^{\boldsymbol{r}\cdot\boldsymbol{j}\theta}|q_{\boldsymbol{j}}^L[f](\cdot)|^\theta\Big)^{1/\theta}\Big|L_p(\mathbb{T}^d)\Big\| \tag{1.5.5}$$

in case $0 < p < \infty$, $0 < \theta \leq \infty$, $\boldsymbol{r} > \max\{1/p, 1/\theta\}$ and $L > \max\{1/p, 1/\theta\}$ (except in the case $\theta = \infty$ where $L \geq 2$). Note, that we encounter the well-known (and infamous) condition $\boldsymbol{r}, L > \max\{1/p, 1/\theta\}$ (see also (1.5.4) for $\theta = 2$), which is relevant if $p > \theta$. However, this condition is most likely optimal for the respective sampling characterization. Note, that when replacing the classical smooth dyadic decomposition of unity (see Def. 3.1) in the definition of the spaces (see Def. 3.2) by a non-smooth variant like de la Vallée Poussin means, we would encounter the same condition on L, which may not be improved as the recent findings in [14, 100, 101] indicate. In addition, note, that in case of quasi-Banach spaces, where $\min\{p, \theta\} < 1$, we need to use sampling kernels (1.5.2) of higher order L as the condition $L > \max\{1/p, 1/\theta\}$ indicates. Surprisingly, the de la Vallée Poussin type kernels work well for the characterization (1.5.5) if $1/2 < p, \theta < \infty$.

1.6 Structure of this thesis

This thesis is structured as follows. In Chapter 2 we provide the basic preliminaries. In Chapter 3 we introduce Besov-Triebel-Lizorkin spaces with dominating mixed smoothness, state embedding results and provide equivalent ways of characterizing this function spaces. Chapter 4 deals with the tensorized Faber-Schauder system as a basis in $C(\mathbb{R}^d)$. Later in this chapter we provide characterizations of Besov-Triebel-Lizorkin smoothness spaces by decreasing properties of the corresponding Faber-Schauder coefficients. In Chapter 5 we use this characterizations to study sparse grid approximation for Sobolev spaces. We consider optimality by computing sharp bounds for sparse grid sampling widths. Additionally we deal with energy norm sampling and prove optimality in terms of the worst case error for standard information. In Chapter 6 we change the point of view and consider best m-term approximation, first in sequence spaces and later for Besov-Triebel-Lizorkin functions with respect to the Faber-Schauder dictionary. We provide a new constructive approximation strategy dealing with the case of small smoothness. Chapter 7 and 8 generalize the ideas obtained in Chapter 4 and 5 for higher smoothness and broader function classes in the periodic context. We provide a new class of periodic sampling kernels with arbitrary fast decreasing properties. We do a intensive optimality consideration for sampling recovery using results for s-numbers. We compare optimality for linear methods with optimality for non-linear methods.

1.7 Contributions of this thesis

This thesis is concerned with the representation and approximation of functions with dominating mixed smoothness by sampling values. We supplemented to the picture of several types of quantities measuring the performance of sampling approximation in Sobolev-Triebel-Lizorkin and Hölder-Nikolskij spaces. Whereas classical theory is mostly done in the periodic case the first part of this thesis presents results on the d-variate unit cube. The contributions of this thesis can be summarized as follows.

- We extend the theory of sampling representations introduced by Dinh Dũng [25, 29] to technically difficult to handle Sobolev-Triebel-Lizorkin spaces $S^r_{p,\theta}F([0,1]^d)$. We prove the main parts of Conjecture 3.20 in [120] concerning the Faber-Schauder system.

- We study numerically important sparse grid approximation as an application of Faber-Schauder sampling representations. This includes considerations for energy norm sampling. Especially we provide a proof where we can prevent using some critical arguments that were stated in [8, Theorem 3.8].

- We provide the exact rate for $\varrho_n(S^r_pW(\mathbb{T}^d), L_q(\mathbb{T}^d))$ in case $1 < p < q < 2$, $r > \frac{1}{p}$.

- We study the Faber-Schauder dictionary in the context of best m-term approximation and achieve optimal rates in the case of small smoothness. We supplement a new constructive approximation strategy.

- We present a new scale of trigonometric sampling kernels that can handle $S^r_{p,\theta}F(\mathbb{T}^d)$ for arbitrary small integrability, fine index parameters $p, \theta > 0$ and arbitrary large smoothness $r > \max\{\frac{1}{p}, \frac{1}{\theta}\}$.

Chapter 2

Preliminaries

2.1 Notation

As usual $\mathbb{N}$ denotes the natural numbers, $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$, $\mathbb{Z}$ denotes the integers, $\mathbb{R}$ the real numbers, and $\mathbb{C}$ the complex numbers. The letter d is always reserved for the underlying dimension in $\mathbb{R}^d, \mathbb{Z}^d$ etc. With $\mathbb{T}^d$ we denote the torus represented by the interval $[-\pi, \pi]^d$, where opposite points are identified. Elements $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{r} \in \mathbb{R}^d$ are always typesetted in bold face. We denote with $\boldsymbol{x} \cdot \boldsymbol{y}$ the usual Euclidean inner product in $\mathbb{R}^d$. For $a \in \mathbb{R}$ we denote $a_+ := \max\{a, 0\}$ and $a_- := \min\{a, 0\}$. For $0 < p \leq \infty$ and $\boldsymbol{x} \in \mathbb{R}^d$ we denote $|\boldsymbol{x}|_p := (\sum_{i=1}^d |x_i|^p)^{1/p}$ with the usual modification in the case $p = \infty$. By $\boldsymbol{x} = (x_1, \ldots, x_d) > 0$ we mean that each coordinate is positive. For $\boldsymbol{j} \in \mathbb{N}_0^d$ we use the notation $\boldsymbol{2^j} = (2^{j_1}, \ldots, 2^{j_d})$, $2^{\boldsymbol{j}} = 2^{j_1} \cdot \ldots \cdot 2^{j_d}$. If X and Y are two (quasi-)normed spaces, the (quasi-)norm of an element x in X will be denoted by $\|x|X\|$. If $T : X \to Y$ is a continuous operator we write $T \in \mathcal{L}(X, Y)$. The symbol $X \hookrightarrow Y$ indicates that the identity operator from X to Y is continuous. For two sequences $(a_n)_{n=1}^{\infty}, (b_n)_{n=1}^{\infty} \subset \mathbb{R}$ we write $a_n \lesssim b_n$ if there exists a constant $c > 0$ such that $a_n \leq c\, b_n$ for all n. We will write $a_n \asymp b_n$ if $a_n \lesssim b_n$ and $b_n \lesssim a_n$ and use the Landau symbol $(a_n)_n = o((b_n)_n) :\Longleftrightarrow \lim_{n\to\infty} a_n/b_n = 0$. We use In addition, we use the following notation $[d] := \{1, \ldots, d\}$, $\mathbb{Z}^d(e) := \{\boldsymbol{k} \in \mathbb{Z}^d : k_i = 0 : \quad i \notin e\}$, $\mathbb{N}_0^d(e) := \{\boldsymbol{k} \in \mathbb{N}_0^d : k_i = 0 : \quad i \notin e\}$ where $e \subset [d]$, $\sigma_p := \max\{0, \frac{1}{p} - 1\}$, $\sigma_{p,\theta} := \max\{0, \frac{1}{p} - 1, \frac{1}{\theta} - 1\}$, where $0 < p, \theta \leq \infty$. For $a \in \mathbb{Z}$ and $\boldsymbol{\ell}, \boldsymbol{j} \in \mathbb{Z}^d$ we use the notation $\boldsymbol{\ell} > a :\Longleftrightarrow \ell_i > a$ for all $i \in [d]$ and $\boldsymbol{\ell} > \boldsymbol{j} :\Longleftrightarrow \ell_i > j_i$ for all $i \in [d]$.

2.2 Distributions

Let $\Omega \subset \mathbb{R}^d$ be a domain (meaning open connected set). We introduce the space of test functions as the set of all compactly supported infinitely many times differentiable functions $f : \Omega \to \mathbb{C}$. We define a topology in $D(\Omega)$ by the convergence of sequences. We say $(f_j)_{j\in\mathbb{N}} \subset D(\Omega)$ converges to $f \in D(\Omega)$ if there is a compact set $K \subset \Omega$ such that

(i) $\operatorname{supp} f_j \subset K, j \in \mathbb{N}$

(ii) $D^{\boldsymbol{\alpha}} f_j \to D^{\boldsymbol{\alpha}} f$ uniformly for all multiindices $\boldsymbol{\alpha} \in \mathbb{N}_0^d$.

As the topological dual we define the space $D'(\Omega)$ as the set of all linear functionals $f : D(\Omega) \to \mathbb{C}$ for that $\varphi_j \xrightarrow{D(\Omega)} \varphi$ implies $f(\varphi_j) \xrightarrow{\mathbb{C}} f(\varphi)$. We use the weak topology for $D'(\Omega)$. That means $(f_j)_{j\in\mathbb{N}_0^d} \subset D'(\Omega)$ converges to f in $D'(\Omega)$ if and only if

$$f_j(\varphi) \to f(\varphi)$$

in $\mathbb{C}$ for all $\varphi \in D(\Omega)$. We introduce the locally convex Schwartz space of infinitely times differentiable fast decreasing functions by

$$S(\mathbb{R}^d) := \{f \in C^\infty(\mathbb{R}^d) : \|f\|_{\boldsymbol{\alpha},\boldsymbol{\beta}} < \infty, \forall \boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{N}_0^d\}$$

with

$$\|f\|_{\boldsymbol{\alpha},\boldsymbol{\beta}} := \sup_{\boldsymbol{x}\in\mathbb{R}^d} \prod_{i=1}^{d} (1 + |x_i|)^{\beta_i} |D^{\boldsymbol{\alpha}} f(\boldsymbol{x})|.$$

A trivial extension by zero yields $D(\Omega) \hookrightarrow S(\mathbb{R}^d)$ with both a set theoretical and topological interpretation. The topological dual of $S(\mathbb{R}^d)$ is denoted by $S'(\mathbb{R}^d)$ and is called the space of tempered distributions. It consists of all continuous linear mappings $f : S(\mathbb{R}^d) \to \mathbb{C}$. Such a mapping is continuous if and only if there exists $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{N}_0^d$ and $C > 0$ such that

$$|f(\varphi)| \le C\|\varphi|S(\mathbb{R}^d)\|_{\boldsymbol{\alpha},\boldsymbol{\beta}}$$

for all $\varphi \in S(\mathbb{R}^d)$. The space $S'(\mathbb{R}^d)$ is equipped with the weak topology. That means a sequence $(f_j)_{j\in\mathbb{N}} \subset S'(\mathbb{R}^d)$ converges to $f \in S'(\mathbb{R}^d)$ if and only if $\lim_{j\to\infty} f_j(\varphi) = f(\varphi)$ in $\mathbb{C}$ for all $\varphi \in S(\mathbb{R}^d)$. A locally integrable function f is interpreted as a distribution by

$$f(\varphi) = \int_{\mathbb{R}^d} f(\boldsymbol{x})\varphi(\boldsymbol{x})d\boldsymbol{x}. \tag{2.2.1}$$

A distribution f is called regular if there is a locally integrable function $\tilde{f}$ such that (2.2.1) holds with $\tilde{f}$ on the right hand side for all test functions φ. If $f \in S'(\mathbb{R}^d)$ is a tempered distribution then the restriction $f|_\Omega$ denotes the restricted mapping f to $D(\Omega)$.

Now we turn to the periodic situation on $\mathbb{T}^d = [-\pi, \pi]^d$ and introduce the space of test functions $D(\mathbb{T}^d)$. It consists of all infinitely times differentiable functions f on $\mathbb{R}^d$ where opposite points are identified, i.e. $f(\boldsymbol{x}) = f(\boldsymbol{x} + 2\pi\boldsymbol{k})$ for all $\boldsymbol{x} \in \mathbb{T}^d$ and $\boldsymbol{k} \in \mathbb{N}_0^d$. It's topology is generated by the family of norms

$$\|f|D(\mathbb{T}^d)\|_N = \sum_{|\boldsymbol{\alpha}|_1\le N} \|D^\alpha f|L_\infty(\mathbb{T}^d)\|, \quad N \in \mathbb{N}_0.$$

A distribution $f : D(\mathbb{T}^d) \to \mathbb{C}$ belongs to the class $D'(\mathbb{T}^d)$ if and only if there exists $C_N > 0$ such that

$$|f(\varphi)| \le C_N\|\varphi|D(\mathbb{T}^d)\|_N, \quad \text{for all } f \in D(\mathbb{T}^d).$$

Again we equip $D'(\mathbb{T}^d)$ with the weak topology, meaning $f_n \xrightarrow{n\to\infty} f$ in $D'(\mathbb{T}^d)$ if and only if $f_n(\varphi) \xrightarrow{n\to\infty} f(\varphi)$ in $\mathbb{C}$ for all $f \in D(\mathbb{T}^d)$.

2.3 Elementary function spaces, Fourier transform and vector-valued spaces

For $\Omega \subset \mathbb{R}^d$ the set of all bounded and continuous functions $f : \Omega \to \mathbb{C}$ is denoted by $C(\Omega)$ equipped with the sup-norm $\|f|L_\infty(\Omega)\| = \sup_{\boldsymbol{x}\in\Omega} |f(\boldsymbol{x})|$. We denote by $L_p(\Omega)$, $0 < p \leq \infty$, the space of all measurable functions $f : \Omega \to \mathbb{C}$ where $\|f|L_p(\Omega)\| := (\int_\Omega |f(\boldsymbol{x})|^p d\boldsymbol{x})^{1/p}$ is finite (with the usual modification if $p = \infty$). For $f, g \in S(\mathbb{R}^d)$ the convolution is always defined as $f * g(\boldsymbol{x}) := \int_{\mathbb{R}^d} f(\boldsymbol{y})g(\boldsymbol{x} - \boldsymbol{y})\, d\boldsymbol{y} \in S(\mathbb{R}^d)$. For $f \in S'(\mathbb{R}^d)$ and $\varphi \in S(\mathbb{R}^d)$ we define the convolution by $\varphi * f(x) = f(\varphi(x - \cdot)) \in S'(\mathbb{R}^d)$, which makes sense also pointwise. For $f \in L_1(\mathbb{R}^d)$ and $\boldsymbol{x}, \boldsymbol{\xi} \in \mathbb{R}^d$ we define the Fourier transform and its inverse by

$$\mathcal{F}f(\boldsymbol{\xi}) := \frac{1}{(2\pi)^{\frac{d}{2}}} \int_{\mathbb{R}^d} f(\boldsymbol{x}) e^{-i\boldsymbol{\xi}\cdot\boldsymbol{x}} d\boldsymbol{x} \quad \text{and} \quad \mathcal{F}^{-1}f(\boldsymbol{x}) := \frac{1}{(2\pi)^{\frac{d}{2}}} \int_{\mathbb{R}^d} f(\boldsymbol{\xi}) e^{i\boldsymbol{\xi}\cdot\boldsymbol{x}} d\boldsymbol{\xi}. \quad (2.3.1)$$

For $f \in L_1(\mathbb{T}^d)$ the $\boldsymbol{k}$-th Fourier coefficient is defined by $\hat{f}(\boldsymbol{k}) := 1/(2\pi)^d \int_{\mathbb{T}^d} f(\boldsymbol{x}) e^{-i\boldsymbol{k}\cdot\boldsymbol{x}} d\boldsymbol{x}$. More generally for $f \in S'(\mathbb{R}^d)$ we define the Fourier transform by

$$\mathcal{F}f(\cdot) := f(\mathcal{F}\cdot).$$

This makes sense for all $f \in S'(\mathbb{R}^d)$. For periodic distributions $f \in D'(\mathbb{T}^d)$ we define the $\boldsymbol{k}$-th Fourier coefficient by

$$\hat{f}_{\boldsymbol{k}} := f(e^{-i\boldsymbol{k}\cdot}).$$

Definition 2.1. *Let $\omega = (\omega_{\boldsymbol{j}})_{\boldsymbol{j}\in A} \subset \mathbb{R}$ be a sequence of weights, where $A \subset \mathbb{N}_0^d$ and let $\Omega \subset \mathbb{R}^d$ be a (Lebesgue) measurable set. We define for $0 < p, \theta \leq \infty$ the spaces $L_p(\ell_\theta(\omega, A), \Omega)$ and $\ell_\theta(\omega, L_p(\Omega), A)$ as the collection of all sequences of functions $(f_{\boldsymbol{j}})_{\boldsymbol{j}\in A} \subset L_p(\Omega)$ with finite (quasi)-norm*

$$\|f_{\boldsymbol{j}}|L_p(\ell_\theta(\omega, A), \Omega)\| := \begin{cases} \left\| \left(\sum_{\boldsymbol{j}\in A} |\omega_{\boldsymbol{j}} f_{\boldsymbol{j}}|^\theta \right)^{\frac{1}{\theta}} \middle| L_p(\Omega) \right\| & : \quad 0 < \theta < \infty, \\ \left\| \sup_{\boldsymbol{j}\in A} |\omega_{\boldsymbol{j}} f_{\boldsymbol{j}}| \middle| L_p(\Omega) \right\| & : \quad \theta = \infty, \end{cases}$$

and

$$\|f_{\boldsymbol{j}}|\ell_\theta(\omega, L_p(\Omega), A)\| := \begin{cases} \left(\sum_{\boldsymbol{j}\in A} |\omega_{\boldsymbol{j}}|^\theta \|f_{\boldsymbol{j}}|L_p(A)\|^\theta \right)^{\frac{1}{\theta}} & : \quad 0 < \theta < \infty, \\ \sup_{\boldsymbol{j}\in A} |\omega_{\boldsymbol{j}}| \|f_{\boldsymbol{j}}|L_p(\Omega)\| & : \quad \theta = \infty, \end{cases}$$

respectively.

For $0 < p, \theta \leq \infty$ the (quasi-)norms $\|\cdot|L_p(\ell_\theta(\omega, \mathbb{N}_0^d), \Omega)\|$ and $\|\cdot|\ell_\theta(\omega, L_p(\Omega), A)\|$ fulfill a μ-triangle inequality with $\mu = \min\{p, \theta, 1\}$. If the domain and summation index set is clear from the context we drop it out of the notation and use the shorter denotations

$$L_p(\ell_\theta(\omega)) = L_p(\ell_\theta(\omega, A), \Omega) \quad \text{and} \quad \ell_\theta(\omega, L_p) = \ell_\theta(\omega, L_p(\Omega), A).$$

In case $\omega = (1)_{\boldsymbol{j}\in A}$ we drop it in the notation and use

$$L_p(\ell_\theta) = L_p(\ell_\theta(\omega, A), \Omega) \quad \text{and} \quad \ell_\theta(L_p) = \ell_\theta(\omega, L_p(\Omega), A).$$

Chapter 3

Besov-Triebel-Lizorkin spaces of mixed smoothness

In this section we start with the classical Fourier analytical definition for Triebel-Lizorkin spaces $S^r_{p,\theta}F(\mathbb{R}^d)$ and Besov spaces $S^r_{p,\theta}B(\mathbb{R}^d)$ with dominating mixed smoothness defined on $\mathbb{R}^d$. We state embedding results and describe equivalent norms characterizing this function spaces. Furthermore we consider function spaces on domains and periodic boundary conditions.

3.1 Basic definitions

We start introducing the following concept decomposing the Fourier image called resolution of unity.

Definition 3.1 (univariate resolution of unity). *A system* $\varphi = (\varphi_j)_{j=0}^\infty \subset C_0^\infty(\mathbb{R})$ *belongs to the class* $\Phi(\mathbb{R})$ *if and only if*

(i) It exists $A > 0$ *such that* $\operatorname{supp}\varphi_0 \subset [-A, A]$.

(ii) There are constants $0 < B < C$, *such that* $\operatorname{supp}\varphi_j \subset \{\xi \in \mathbb{R} : \ B2^j \leq |\xi| \leq C2^j\}$.

(iii) For all $\alpha \in \mathbb{N}_0$ *there are constants* $C_\alpha > 0$ *such that*

$$\sup_{\xi \in \mathbb{R}, j \in \mathbb{N}_0} 2^\alpha |D^\alpha \varphi_j(\xi)| \leq C_\alpha < \infty.$$

(iv) For all $\xi \in \mathbb{R}$

$$\sum_{j=0}^\infty \varphi_j(\xi) = 1.$$

Applying (iv) in Definition 3.1 we obtain the following decomposition of $f \in S'(\mathbb{R}^d)$

$$f = \sum_{\boldsymbol{j} \in \mathbb{N}_0^d} \delta_{\boldsymbol{j}}[f]$$

with convergence in $S'(\mathbb{R}^d)$ where

$$\delta_{\boldsymbol{j}}[f](\boldsymbol{x}) := \mathcal{F}^{-1}(\varphi_{\boldsymbol{j}}\mathcal{F}f)(\boldsymbol{x}), \quad \boldsymbol{j} \in \mathbb{N}_0^d \tag{3.1.1}$$

with

$$\varphi_{\boldsymbol{j}}(\boldsymbol{\xi}) = \prod_{i=1}^{d} \varphi_{j_i}(\xi_i), \quad \boldsymbol{\xi} \in \mathbb{R}^d, \quad \boldsymbol{j} \in \mathbb{N}_0^d. \tag{3.1.2}$$

Later we use the convention $\delta_{\boldsymbol{j}}[f] = 0$ if there exists an $i \in [d]$ with $j_i < 0$. We introduce the function spaces $S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{R}^d)$ and $S^{\boldsymbol{r}}_{p,\theta}B(\mathbb{R}^d)$ using this Fourier-analytic building blocks.

Definition 3.2. *Let $\varphi = \{\varphi_j(x)\}_{j=0}^{\infty} \in \Phi(\mathbb{R})$, and $\boldsymbol{r} \in \mathbb{R}^d$. Let further*

(i) $0 < p < \infty$ and $0 < \theta \leq \infty$. Then

$$S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{R}^d) := \Big\{ f \in S'(\mathbb{R}^d) : \|f|S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{R}^d)\| < \infty \Big\},$$

where

$$\|f|S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{R}^d)\| := \|\delta_{\boldsymbol{j}}[f]|L_p(\ell_\theta(2^{\boldsymbol{r}\cdot\boldsymbol{j}}))\|.$$

(ii) $0 < p, \theta \leq \infty$. Then

$$S^{\boldsymbol{r}}_{p,\theta}B(\mathbb{R}^d) := \Big\{ f \in S'(\mathbb{R}^d) : \|f|S^{\boldsymbol{r}}_{p,\theta}B(\mathbb{R}^d)\| < \infty \Big\},$$

where

$$\|f|S^{\boldsymbol{r}}_{p,\theta}B(\mathbb{R}^d)\| := \|\delta_{\boldsymbol{j}}[f]|\ell_\theta(2^{\boldsymbol{r}\cdot\boldsymbol{j}}, L_p)\|.$$

Remark 3.3. *(i) Different resolutions of unity $\varphi, \psi \in \Phi(\mathbb{R})$ employed in (3.1.1) generate equivalent norms in $S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{R}^d)$ and $S^{\boldsymbol{r}}_{p,\theta}B(\mathbb{R}^d)$, respectively, cf. [99, 2.2.3, Proposition 1]*

(ii) In case $d = 1$ the concepts of dominating mixed smoothness and isotropic smoothness coincide. We use the notation

$$F^r_{p,\theta}(\mathbb{R}) := S^r_{p,\theta}F(\mathbb{R}) \text{ and } B^r_{p,\theta}(\mathbb{R}) := S^r_{p,\theta}B(\mathbb{R}).$$

(iii) In case $\theta = 2$ and $1 < p < \infty$ the space $S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{R}^d)$ coincides with the Sobolev space of dominating mixed smoothness $S^{\boldsymbol{r}}_pW(\mathbb{R}^d)$ including $L_p(\mathbb{R}^d)$ if $\boldsymbol{r} = 0$. $S^{\boldsymbol{r}}_pW(\mathbb{R}^d)$ is classically normed by

$$\|f|S^{\boldsymbol{r}}_pW(\mathbb{R}^d)\|' := \Big\|\mathcal{F}^{-1}\Big(\prod_{i=1}^{d}(1+|\xi_i|)^{r_i}\mathcal{F}f(\boldsymbol{\xi})\Big)(\boldsymbol{x})\Big|L_p(\mathbb{R}^d)\Big\|. \tag{3.1.3}$$

(iv) In case $\theta = 2$, $1 < p < \infty$ and $r \in \mathbb{N}_0$ we have the equivalence

$$\|f|S^r_pW(\mathbb{R}^d)\| \asymp \|f|L_p(\mathbb{R}^d)\| + \sum_{0<|\boldsymbol{\alpha}|_\infty \leq r} \|D^{\boldsymbol{\alpha}}f|L_p(\mathbb{R}^d)\|, \tag{3.1.4}$$

cf. [99, p. 104].

3.2 Embeddings

We state the following embedding results without proof. For a reference see [99, 127] and [57]. For a complete history of the non-trivial embedding in Lemma 3.5 we refer to [33, Remark 3.8].

Lemma 3.4. *(i) Let $0 < p \le \infty$ (F-case: $p < \infty$), $0 < \theta \le \infty$, $\boldsymbol{r} > \sigma_p$. Then*

$$S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{R}^d) \hookrightarrow L_{\max\{p,1\}}(\mathbb{R}^d) \quad \text{and} \quad S^{\boldsymbol{r}}_{p,\theta}B(\mathbb{R}^d) \hookrightarrow L_{\max\{p,1\}}(\mathbb{R}^d)\,,$$

which means $S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{R}^d)$ and $S^{\boldsymbol{r}}_{p,\theta}B(\mathbb{R}^d)$ consist of regular distributions that allow an interpretation as functions.

(ii) Let $0 < p \le \infty$ (F-case: $p < \infty$), $0 < \theta \le \infty$, $\boldsymbol{r} > \frac{1}{p}$. Then

$$S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{R}^d) \hookrightarrow C(\mathbb{R}^d) \quad \text{and} \quad S^{\boldsymbol{r}}_{p,\theta}B(\mathbb{R}^d) \hookrightarrow C(\mathbb{R}^d)\,,$$

which means that we find in every equivalence class of $S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{R}^d)$ and $S^{\boldsymbol{r}}_{p,\theta}B(\mathbb{R}^d)$ a unique continuous representative making discrete point evaluations possible.

(iii) Let $0 < p \le \infty$ (F-case: $p < \infty$), $0 < \theta_1 < \theta_2 \le \infty$ and $\boldsymbol{r} \in \mathbb{R}^d$. Then

$$S^{\boldsymbol{r}}_{p,\theta_1}F(\mathbb{R}^d) \hookrightarrow S^{\boldsymbol{r}}_{p,\theta_2}F(\mathbb{R}^d) \quad \text{and} \quad S^{\boldsymbol{r}}_{p,\theta_1}B(\mathbb{R}^d) \hookrightarrow S^{\boldsymbol{r}}_{p,\theta_2}B(\mathbb{R}^d).$$

(iv) Let $0 < p < \infty$, $0 < \theta \le \infty$ and $\boldsymbol{r} \in \mathbb{R}^d$. Then

$$S^{\boldsymbol{r}}_{p,\min\{p,\theta\}}B(\mathbb{R}^d) \hookrightarrow S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{R}^d) \hookrightarrow S^{\boldsymbol{r}}_{p,\max\{p,\theta\}}B(\mathbb{R}^d).$$

(v) Let $0 < p \le \infty$ (F-case: $p < \infty$), $0 < \theta, \nu \le \infty$ and $\boldsymbol{r}_1, \boldsymbol{r}_2 \in \mathbb{R}^d$ with $\boldsymbol{r}_1 > \boldsymbol{r}_2$ Then

$$S^{\boldsymbol{r}_1}_{p,\theta}F(\mathbb{R}^d) \hookrightarrow S^{\boldsymbol{r}_2}_{p,\nu}F(\mathbb{R}^d) \quad \text{and} \quad S^{\boldsymbol{r}_1}_{p,\theta}B(\mathbb{R}^d) \hookrightarrow S^{\boldsymbol{r}_2}_{p,\nu}B(\mathbb{R}^d).$$

(vi) Let $0 < p < q < \infty$, $0 < \theta, \nu \le \infty$ and $\boldsymbol{r}_1, \boldsymbol{r}_2 \in \mathbb{R}^d$ with $\boldsymbol{r}_1 > \boldsymbol{r}_2$ fulfilling

$$\boldsymbol{r}_1 - \frac{1}{p} = \boldsymbol{r}_2 - \frac{1}{q}.$$

Then

$$S^{\boldsymbol{r}_1}_{p,\theta}F(\mathbb{R}^d) \hookrightarrow S^{\boldsymbol{r}_2}_{q,\nu}F(\mathbb{R}^d) \quad \text{and} \quad S^{\boldsymbol{r}_1}_{p,\theta}B(\mathbb{R}^d) \hookrightarrow S^{\boldsymbol{r}_2}_{q,\theta}B(\mathbb{R}^d).$$

Observe that, in contrast to the diagonal Besov embedding in Lemma 3.4, (iv), the fine index θ and ν play no role for the F-case.

Lemma 3.5 (Jawerth-Franke embedding)**.** *Let $0 < p < q \le \infty$, $0 < \theta \le \infty$, $\boldsymbol{r}_1, \boldsymbol{r}_2 \in \mathbb{R}^d$ such that*

$$\boldsymbol{r}_1 - \frac{1}{p} = \boldsymbol{r}_2 - \frac{1}{q}$$

is fulfilled.

(i) Then

$$S^{\boldsymbol{r}_1}_{p,\theta}F(\mathbb{R}^d) \hookrightarrow S^{\boldsymbol{r}_2}_{q,p}B(\mathbb{R}^d).$$

(ii) If additionally $q < \infty$ then

$$S^{\boldsymbol{r}_1}_{p,q}B(\mathbb{R}^d) \hookrightarrow S^{\boldsymbol{r}_2}_{q,\theta}F(\mathbb{R}^d).$$

3.3 Further characterizations

In this section we describe equivalent ways of characterizing Besov-Triebel-Lizorkin spaces. As a first approach we consider convolutions with so called local mean kernels. Classically the Fourier analytical building blocks of a function f are bandlimited functions that are generated as convolutions of the function f with bandlimited kernels whose Fourier image is sufficiently smooth. Local mean characterizations allow to replace this bandlimited kernels by for instance compactly supported ones. Let $\Psi_0, \Psi_1 \in S(\mathbb{R}^d)$ such that

(i) $|\mathcal{F}\Psi_0(\xi)| > 0$ for $|\xi| < \varepsilon$

(ii) $|\mathcal{F}\Psi_1(\xi)| > 0$ for $\frac{\varepsilon}{2} < |\xi| < 2\varepsilon$ and

(iii) $D^\alpha \mathcal{F}\Psi_1(0) = 0$ for all $0 \leq \alpha < L$

(3.3.1)

hold for some $\varepsilon > 0$. As usual, the j-th dilation of Ψ_1 is given by

$$\Psi_j := 2^{j-1}\Psi_1(2^{j-1}x), \quad j \geq 2.$$

For $\boldsymbol{j} \in \mathbb{N}_0^d$ we define by tensorization

$$\Psi_{\boldsymbol{j}}(\boldsymbol{x}) = \prod_{i=1}^{d} \Psi_{j_i}(x_i), \quad \boldsymbol{x} \in \mathbb{R}^d.$$

Remark 3.6. *(i) Inserting the definitions, (iii) in* (3.3.1) *means*

$$\int_{\mathbb{R}} x^\alpha \Psi_1(x)dx = 0$$

for all $0 \leq \alpha < L$*. This condition is called* L*-th order moment condition.*

(ii) There are local mean kernels fulfilling arbitrary (but finite) moment conditions and have compact supports. Let us consider the function

$$g(t) = \chi_{[-\frac{1}{2},\frac{1}{2}]}(t)$$

and define

$$g_k(t) = 2^k g(2^k t).$$

For the infinite convolution

$$\varphi = g * g_1 * g_2 * \dots$$

one can show $\varphi \in C_0^\infty(\mathbb{R})$*,* $0 \leq \varphi \leq 1$*,* $\|\varphi\|_1 = \|g\|_1 = 1$ *with* $\operatorname{supp}\varphi = [-1,1]$*. The Fourier transform is given as the infinite product*

$$\mathcal{F}\varphi(\xi) = \prod_{k=0}^{\infty} \operatorname{sinc}(2^{-k}\xi).$$

We define

$$\Psi_0(x) := 2\varphi(2x),$$

$$\Psi_1(x) := \frac{d^L}{dx^L}(2\Psi_0(2\cdot) - \Psi_0(\cdot))(x).$$

Observing the identity

$$\mathcal{F}\Psi_1(\xi) = (2\pi i\xi)^L(\mathcal{F}\Psi_0(\xi/2) - \mathcal{F}\Psi_0(\xi)),$$

it is easy to check that Ψ_0, Ψ_1 *fulfill the conditions in* (3.3.1). *For further information we refer to [97], [61, Section 6.1] and [121, p. 10].*

Theorem 3.7. *Let* $0 < p, \theta \leq \infty$ *(F-case:* $\theta < \infty$*),* $(\Psi_{\boldsymbol{j}})_{\boldsymbol{j}\in\mathbb{N}_0^d}$ *as above with* $L+1 > \boldsymbol{r}$*. Then*

$$\|f|S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{R}^d)\|^* := \|\Psi_{\boldsymbol{j}} * f|L_p(\ell_\theta(2^{\boldsymbol{r}\cdot\boldsymbol{j}}))\|$$

describes an equivalent norm in $S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{R}^d)$ *and*

$$\|f|S^{\boldsymbol{r}}_{p,\theta}B(\mathbb{R}^d)\|^* := \|\Psi_{\boldsymbol{j}} * f|\ell_\theta(2^{\boldsymbol{r}\cdot\boldsymbol{j}}, L_p)\|$$

in $S^{\boldsymbol{r}}_{p,\theta}B(\mathbb{R}^d)$.

Proof. Such characterizations for function spaces of dominating mixed smoothness were studied first in [127]. For local mean representations with the assumptions from above we refer to [123]. For more details on the interesting history of this characterization we refer to [121, Remark 4.5]. □

Mixed B-spaces are classically defined as a space of functions with L_p-bounded mixed differences. A related characterization by differences is also available for F-spaces. Before we start we introduce some notation concerning iterated differences. For a multivariate function f on $\mathbb{R}^d$ we denote the first order differences with stepwidth $h \in \mathbb{R}$ acting in direction $i \in [d]$ by

$$\Delta_h^{1,i} f(\boldsymbol{x}) := f(\boldsymbol{x} + h\boldsymbol{e}_i) - f(\boldsymbol{x}),$$

where $\boldsymbol{e}_i = (0, \ldots, 0, 1, 0, \ldots, 0)$. m-th order differences can be defined iteratively by

$$\Delta_h^{m,i} f(\boldsymbol{x}) := \Delta_h^{1,i}\Delta_h^{m-1,i} f(\boldsymbol{x}).$$

This allows us to define for $e \subset [d]$ and $\boldsymbol{h} \in \mathbb{R}^d$ the $\boldsymbol{m}$-th order difference operator acting in the directions contained in e by

$$\Delta_{\boldsymbol{h}}^{\boldsymbol{m},e} f(\boldsymbol{x}) := \Big(\prod_{i\in e} \Delta_{h_i}^{m_i,i}\Big) f(\boldsymbol{x}). \tag{3.3.2}$$

This allows us to state the characterization by rectangle means:

Theorem 3.8. *Let* $0 < p < \infty$, $0 < \theta \le \infty$ *and* $\boldsymbol{m} \in \mathbb{N}_0^d$ *such that* $\sigma_{p,\theta} < \boldsymbol{r} < \boldsymbol{m}$ *is fulfilled. Then*

$$\|f|S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{R}^d)\| \asymp \sum_{e\subset[d]} \|f|S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{R}^d)\|_{e,\boldsymbol{m}}$$

holds with

$$\|f|S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{R}^d)\|_{e,\boldsymbol{m}} := \left\|\Big[\sum_{\boldsymbol{j}\in\mathbb{N}_0^d(e)} 2^{\theta\boldsymbol{r}\cdot\boldsymbol{j}}\Big(\Big(\prod_{i\in e}2^{j_i}\Big)\int_{\substack{|h_i|\le 2^{-j_i}\\ i\in[d]}} |\Delta^{\boldsymbol{m},e}_{\boldsymbol{h}} f(\cdot)|d\boldsymbol{h}\Big)^{\theta}\Big]^{\frac{1}{\theta}}\Bigg|L_p(\mathbb{R}^d)\right\|$$

and the usual modification in case $\theta = \infty$.

Proof. We refer to [66, Theorem 3.7]. There the case for constant smoothness vector $\boldsymbol{r} = (r, \dots, r)$ has been considered. The necessary modifications are straight forward. □

Theorem 3.9. *Let* $0 < p, \theta \le \infty$ *and* $\boldsymbol{m} \in \mathbb{N}$ *such that* $\sigma_p < \boldsymbol{r} < \boldsymbol{m}$ *is fulfilled. Then*

$$\|f|S^{\boldsymbol{r}}_{p,\theta}B(\mathbb{R}^d)\| \asymp \sum_{e\subset[d]} \|f|S^{\boldsymbol{r}}_{p,\theta}B(\mathbb{R}^d)\|_{e,\boldsymbol{m}}$$

holds with

$$\|f|S^{\boldsymbol{r}}_{p,\theta}B(\mathbb{R}^d)\|_{e,\boldsymbol{m}} := \left[\sum_{\boldsymbol{j}\in\mathbb{N}_0^d(e)} 2^{\theta\boldsymbol{r}\cdot\boldsymbol{j}}\left\|\Big(\prod_{i\in e}2^{j_i}\Big)\int_{\substack{|h_i|\le 2^{-j_i}\\ i\in[d]}} |\Delta^{\boldsymbol{m},e}_{\boldsymbol{h}} f(\cdot)|d\boldsymbol{h}\Big|L_p(\mathbb{R}^d)\right\|^{\theta}\right]^{\frac{1}{\theta}}$$

and the usual modification in case $\theta = \infty$.

Proof. We refer to [122, Theorem 3.7.1 and Remark 3.7.1]. There the outer sum is an integral. By decomposing this into dyadic blocks one obtains the form stated above. □

3.4 Spaces on domains

In this section we deal with function spaces on domains. From a general point of view we mean with a domain $\Omega \subset \mathbb{R}^d$ a open connected set. Later dealing with continuous functions trivial extensions allow us to deal with the compact set $[0,1]^d$.

Definition 3.10. *Let* Ω *be a domain and* $\boldsymbol{r} \in \mathbb{R}^d$.

(i) Let additionally $0 < p < \infty$ *and* $0 < \theta \le \infty$. *Then we define*

$$S^{\boldsymbol{r}}_{p,\theta}F(\Omega) := \{f \in D'(\Omega) : \exists g \in S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{R}^d) \text{ with } g|_\Omega = f\}$$

where

$$\|f|S^{\boldsymbol{r}}_{p,\theta}F(\Omega)\| := \inf\{\|g|S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{R}^d)\| : g \in S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{R}^d), g|_\Omega = f\}.$$

(ii) Let additionally $0 < p, \theta \leq \infty$. *Then we define*

$$S^{\boldsymbol{r}}_{p,\theta}B(\Omega) := \{f \in D'(\Omega) : \exists g \in S^{\boldsymbol{r}}_{p,\theta}B(\mathbb{R}^d) \text{ with } g|_\Omega = f\}$$

where

$$\|f|S^{\boldsymbol{r}}_{p,\theta}B(\Omega)\| := \inf\{\|g|S^{\boldsymbol{r}}_{p,\theta}B(\mathbb{R}^d)\| : g \in S^{\boldsymbol{r}}_{p,\theta}B(\mathbb{R}^d), g|_\Omega = f\}.$$

On bounded domains Ω we have additionally the following embedding.

Lemma 3.11. *Let* $0 < q < p \leq \infty$ *(F-case:* $p < \infty$*),* $\boldsymbol{r} \in \mathbb{R}^d$, $0 < \theta \leq \infty$ *and* $|\Omega| < \infty$. *Then*

$$S^{\boldsymbol{r}}_{p,\theta}F(\Omega) \hookrightarrow S^{\boldsymbol{r}}_{q,\theta}F(\Omega)$$

and

$$S^{\boldsymbol{r}}_{p,\theta}B(\Omega) \hookrightarrow S^{\boldsymbol{r}}_{q,\theta}B(\Omega).$$

Proof. The proof follows trivially by definition using the embedding

$$L_p(\Omega) \hookrightarrow L_q(\Omega).$$

□

3.5 Hyperbolic representation of isotropic Sobolev spaces

In this section we introduce isotropic Sobolev spaces and discuss their representation in terms of Section 3.1. We start extending Definition 3.1 to a multivariate isotropic version.

Definition 3.12 (Resolution of unity - isotropic)**.** *A system* $\psi = (\psi_j)_{j=0}^{\infty} \subset C_0^{\infty}(\mathbb{R}^d)$ *belongs to the class* $\Phi(\mathbb{R}^d)$ *if and only if*

(i) It exists $A > 0$ *such that* $\operatorname{supp}\psi_0 \subset \{\boldsymbol{\xi} \in \mathbb{R}^d : |\boldsymbol{\xi}|_2 < A\}$.

(ii) There are constants $0 < B < C$*, such that* $\operatorname{supp}\psi_j \subset \{\boldsymbol{\xi} \in \mathbb{R}^d : B2^j \leq |\boldsymbol{\xi}|_2 \leq C2^j\}$.

(iii) For all $\boldsymbol{\alpha} \in \mathbb{N}_0^d$ *holds*

$$\sup_{\boldsymbol{\xi} \in \mathbb{R}^d, j \in \mathbb{N}_0} 2^{j|\boldsymbol{\alpha}|_1}|D^{\boldsymbol{\alpha}}\psi_j(\boldsymbol{\xi})| \leq c_{\boldsymbol{\alpha}} < \infty \text{ and}$$

(iv) For all $\boldsymbol{\xi} \in \mathbb{R}^d$

$$\sum_{j=0}^{\infty} \psi_j(\boldsymbol{\xi}) = 1.$$

Again, applying (iv) in Definition 3.12 we obtain the following decomposition of $f \in S'(\mathbb{R}^d)$. For $\psi = \{\psi_j\}_{j=0}^{\infty} \in \Phi(\mathbb{R}^d)$ let

$$\eta_j[f](\boldsymbol{x}) := \mathcal{F}^{-1}(\psi_j \mathcal{F} f)(\boldsymbol{x}). \tag{3.5.1}$$

Then it holds

$$f = \sum_{j\in\mathbb{N}_0} \eta_j[f]$$

with convergence in $S'(\mathbb{R}^d)$.

Definition 3.13. *Let $1 < p < \infty$ and $r \in \mathbb{R}$. Then*

$$W_p^r(\mathbb{R}^d) := \Big\{ f \in S'(\mathbb{R}^d) : \|f|W_p^r(\mathbb{R}^d)\| < \infty \Big\},$$

where

$$\|f|W_p^r(\mathbb{R}^d)\| := \|\eta_j[f]|L_p(\ell_2(2^{rj}, \mathbb{N}_0), \mathbb{R}^d)\|.$$

Lemma 3.14. *Let $1 < p < \infty$ and $r \in \mathbb{N}_0$. Then we have*

$$\|f|W_p^r(\mathbb{R}^d)\| \asymp \|f|L_p(\mathbb{R}^d)\| + \sum_{|\boldsymbol{\alpha}|_1 \le r} \|D^{\boldsymbol{\alpha}} f|L_p(\mathbb{R}^d)\|.$$

Proof. For more details on the proof we refer to [119, Theorem 2.5.6]. □

Remark 3.15. *The next figure shows the different Fourier supports of the hyperbolic and the isotropic resolutions of unity.*

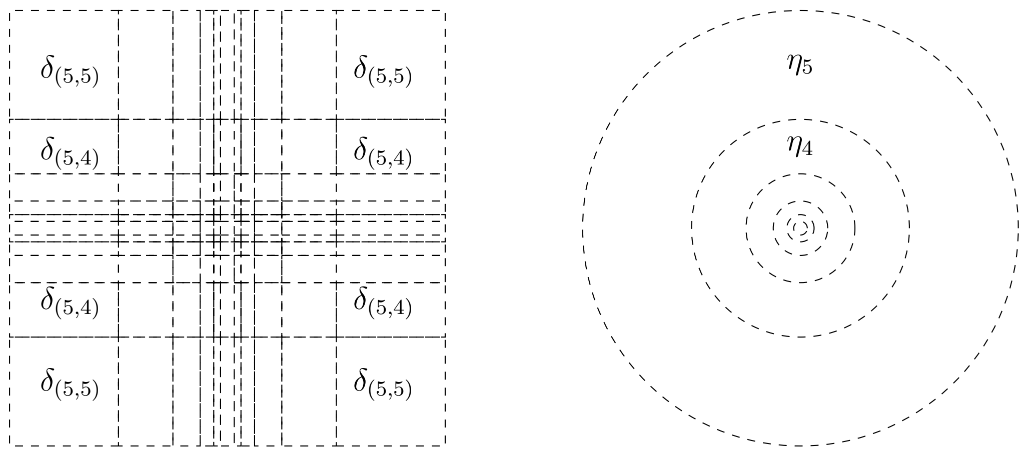


Figure 3.1: Fourier support of hyperbolic and isotropic resolution of unity

Theorem 3.16. *Let $1 < p < \infty$ and $r \in \mathbb{R}$. Then the space $W_p^r(\mathbb{R}^d)$ can be equivalently normed by*

$$\|f|W_p^r(\mathbb{R}^d)\| \asymp \|\delta_{\boldsymbol{j}}[f]|L_p(\ell_2(2^{r|\boldsymbol{j}|_\infty}, \mathbb{N}_0^d), \mathbb{R}^d)\|,$$

where $\delta_{\boldsymbol{j}}[f]$ is as in (3.1.1)*. That means we use a hyperbolic resolution of unity to give an equivalent norm for an isotropic space.*

Proof. We refer to [126]. □

Remark 3.17. *According to [126] a similar result for B-spaces can hold only in case $p = \theta = 2$.*

3.6 Periodic spaces

A function defined on $\mathbb{R}^d$ is 2π-periodic, if and only if for all $\boldsymbol{x} \in \mathbb{T}^d$ we have

$$f(\boldsymbol{x}) = f(\boldsymbol{x} + 2\pi \boldsymbol{k})$$

for all $\boldsymbol{k} \in \mathbb{Z}^d$. The function spaces defined in Section 3.1 are based on L_p functions or even more general distributions, where in general no point evaluations are available. We use periodic distributions $f \in D'(\mathbb{T}^d)$. Based on this periodic distributions we can define periodic Besov and Triebel-Lizorkin-Sobolev spaces. We need the following building blocks. Let $\varphi \in \Phi(\mathbb{R})$ with $\varphi_{\boldsymbol{j}}$ as in (3.1.2) then we define

$$\delta_{\boldsymbol{j}}^{\pi}[f] := \sum_{\boldsymbol{k} \in \mathbb{Z}^d} \hat{f}_{\boldsymbol{k}} \varphi_{\boldsymbol{j}}(\boldsymbol{k}) e^{i\boldsymbol{k}\cdot}, . \tag{3.6.1}$$

This allows us to decompose $f \in D'(\mathbb{T}^d)$ by

$$f = \sum_{\boldsymbol{j} \in \mathbb{N}_0^d} \delta_{\boldsymbol{j}}^{\pi}[f] \tag{3.6.2}$$

with convergence in $D'(\mathbb{T}^d)$.

Definition 3.18. *Let $\varphi = \{\varphi_j\}_{j=0}^{\infty} \in \Phi(\mathbb{R})$, and $\boldsymbol{r} \in \mathbb{R}^d$.*

(i) Let $0 < p < \infty$ and $0 < \theta \le \infty$. Then

$$S_{p,\theta}^{\boldsymbol{r}} F(\mathbb{T}^d) := \Big\{ f \in D'(\mathbb{T}^d) : \|f | S_{p,\theta}^{\boldsymbol{r}} F(\mathbb{T}^d)\| < \infty \Big\},$$

where

$$\|f | S_{p,\theta}^{\boldsymbol{r}} F(\mathbb{T}^d)\| := \|\delta_{\boldsymbol{j}}^{\pi}[f] | L_p(\ell_q(2^{\boldsymbol{r}\cdot\boldsymbol{j}}), \mathbb{T}^d)\|.$$

(ii) Let $0 < p, \theta \le \infty$. Then

$$S_{p,\theta}^{\boldsymbol{r}} B(\mathbb{T}^d) := \Big\{ f \in D'(\mathbb{T}^d) : \|f | S_{p,\theta}^{\boldsymbol{r}} B(\mathbb{T}^d)\| < \infty \Big\},$$

where

$$\|f | S_{p,\theta}^{\boldsymbol{r}} B(\mathbb{T}^d)\| := \|\delta_{\boldsymbol{j}}^{\pi}[f] | \ell_\theta(2^{\boldsymbol{r}\cdot\boldsymbol{j}}, L_p(\mathbb{T}^d))\|.$$

Compared to Section 3.1 we integrate here over $\mathbb{T}^d$ instead of $\mathbb{R}^d$.

Remark 3.19. *All aspects of this chapter have more or less obvious periodic counterparts. The embeddings of Lemma 3.4 hold in the periodic case including Lemma 3.11. We refer to [99, Chap. 3]. For our purpose interestingly to mention, characterization by differences in Theorems 3.8 and 3.9 work by replacing the $L_p(\mathbb{R}^d)$ integration by $L_p(\mathbb{T}^d)$, cf. [122].*

Chapter 4

The Faber-Schauder basis in multivariate smoothness spaces

In this section we introduce the Faber-Schauder system as a basis in $C([0,1]^d)$. Later we proof equivalent characterizations of $S^r_{p,\theta}F([0,1]^d)$ and $S^r_{p,\theta}B([0,1]^d)$ in terms of decreasing properties for sequences of Faber-Schauder coefficients. They allow us to deal with sampling approximation in terms of sequence spaces. Transferring complicate approximation problems from the level of function spaces to the easier to handle level of sequence spaces is a well known technique for several estimates of (pseudo) s-numbers in approximation theory, see for instance [127, 77, 79, 81]. Quite new is the approach to handle sampling with a similar method. This originally goes back to Dinh Dung [26, 29]. We extend this technique to construct and analyze (energy) sparse-grid sampling operators for functions in $S^r_pW(\mathbb{R}^d)$.

4.1 The (tensorized) Faber-Schauder system

In this section we introduce the Faber-Schauder system. Faber proved in [38] that every continuous function f in $[0,1]$ can be expanded into a basis of hat functions. Introducing this system we refer to the notation of iterated differences $\Delta^{\boldsymbol{m},e}_{\boldsymbol{h}}f(\boldsymbol{x})$ given in (3.3.2).

Definition 4.1. *We define the univariate L_∞-normalized hat function*

$$v(x) = v_{0,0}(x) := \begin{cases} 2x & : \quad 0 < x \leq \frac{1}{2}, \\ 2(1-x) & : \quad \frac{1}{2} < x < 1, \\ 0 & : \quad \textit{otherwise}. \end{cases}$$

The hat function of level $j \in \mathbb{N}_0$ and translation $k \in \mathbb{Z}$ is given by

$$v_{j,k}(x) := v(2^j x - k).$$

Additionally for $k \in \mathbb{Z}$ we use the notation

$$v_{-1,k}(x) := v\Big(\frac{1}{2}(x-k+1)\Big).$$

Faber [38] originally considered the interval $[0,1]$. Based on his arguments it is a trivial exercise to show that every continuous function on $\mathbb{R}$ can be represented by the series

$$f = \sum_{j\in\mathbb{N}_{-1}} \sum_{k\in\mathbb{Z}} d_{j,k}(f) v_{j,k} \tag{4.1.1}$$

with conditional convergence in $C(K)$, where K is an arbitrary compact subset of $\mathbb{R}$. The coefficients $d_{j,k}(f)$ are given by

$$d_{j,k}(f) := \begin{cases} f(x_{j,k}) & : \quad j = -1, \\ -\frac{1}{2}\Delta^2_{2^{-(j+1)}} f(x_{j,k}) & : \quad j \geq 0, \end{cases}$$

with

$$x_{j,k} := \begin{cases} k & : \quad j = -1 \\ 2^{-j}k & : \quad j \geq 0. \end{cases}$$

Definition 4.2. *We introduce for $j \in \mathbb{N}_0$, $k \in \mathbb{Z}$ the overlapping hat functions*

$$v^*_{j,k} := v\Big(2^j x - \frac{k-1}{2}\Big).$$

This allows us to give to following obvious refinement equation.

Lemma 4.3. *For $j \in \mathbb{N}_0$, $k \in \mathbb{Z}$ we have*

$$v_{j,k} = v^*_{j,2k+1} = \frac{1}{2}(v^*_{j+1,4k+1} + v^*_{j+1,4k+3}) + v^*_{j+1,4k+2}.$$

Tensorization of the univariate hat functions yields a d-variate version of the Faber-Schauder system. For $\boldsymbol{j} \in \mathbb{N}^d_{-1}$ and $\boldsymbol{k} \in \mathbb{Z}^d$ we define the d-variate tensor hat function by

$$v_{\boldsymbol{j},\boldsymbol{k}}(\boldsymbol{x}) := v_{j_1,k_1}(x_1) \cdot \ldots \cdot v_{j_d,k_d}(x_d)$$

with coefficients $d_{\boldsymbol{j},\boldsymbol{k}}(f)$ given by

$$d_{\boldsymbol{j},\boldsymbol{k}}(f) := \Big(-\frac{1}{2}\Big)^{|e(\boldsymbol{j})|} \Delta^{2,e(\boldsymbol{j})}_{2^{-(\boldsymbol{j}+1)}} f(x_{\boldsymbol{j},\boldsymbol{k}}) \quad \text{with} \quad x_{\boldsymbol{j},\boldsymbol{k}} := (x_{j_1,k_1}, \ldots, x_{j_d,k_d}) \tag{4.1.2}$$

and

$$e(\boldsymbol{j}) := \{i \in [d] : \ j_i \geq 0\}.$$

Since the convergence in (4.1.1) is conditional we have to say some words about the order of summation. For $\mu \in \mathbb{N}$ we define the open intervals

$$E^d_\mu = (-\mu, \mu)^d. \tag{4.1.3}$$

Definition 4.4. *Let*

$$B^d_n := \{(\boldsymbol{j},\boldsymbol{k}) \in \mathbb{N}^d_{-1} \times \mathbb{Z}^d : \boldsymbol{j} : |\boldsymbol{j}|_\infty \leq n, \boldsymbol{k} : \operatorname{supp} v_{\boldsymbol{j},\boldsymbol{k}} \cap E^d_n \neq \emptyset\}.$$

Then we define for $f \in C(\mathbb{R}^d)$ the operator

$$F^d_n f(\boldsymbol{x}) := \sum_{(\boldsymbol{j},\boldsymbol{k})\in B^d_n} d_{\boldsymbol{j},\boldsymbol{k}} v_{\boldsymbol{j},\boldsymbol{k}}(\boldsymbol{x}). \tag{4.1.4}$$

Lemma 4.5. *Applying the operator F_n^1 to $f \in \mathbb{C}(\mathbb{R})$ gives a continuous function that is piecewise linear in the intervals $I_{n,k} := [2^{-(n+1)}k, 2^{-(n+1)}(k+1)]$, $k \in \mathbb{Z}$.*

Proof. The proof is a simple consequence of Lemma 4.3. Applying it iteratively there exists a sequence $(\lambda_{\boldsymbol{j},\boldsymbol{k}})_{\boldsymbol{j},\boldsymbol{k}} \subset \mathbb{C}$ such that we obtain a representation

$$F_n^1 f(x) := \sum_{(j,k)\in B_n^1} d_{j,k} v_{j,k} = \sum_{k\in\mathbb{Z}} \lambda_{n,k} v_{n,k}^*.$$

We have

$$\operatorname{supp} v_{n,k} = [2^{-n}k, 2^{-n}(k+1)]$$

and piecewise linearity in

$$L_{n,k} := \Big[2^{-(n+1)}2k, 2^{-(n+1)}(2k+1)\Big] \quad \text{and} \quad R_{n,k} := \Big[2^{-(n+1)}(2k+1), 2^{-(n+1)}(2k+2)\Big].$$

For each interval $L_{n,k}$ (or $R_{n,k}$) there are only two translated hat functions $v_{n,u}^*$ and $v_{n,u+1}^*$, $u \in \mathbb{Z}$ with

$$|\operatorname{supp} v_{n,u}^* \cap \operatorname{supp} v_{n,u+1}^* \cap L_{n,k}| > 0,$$

(or $|\operatorname{supp} v_{n,u}^* \cap \operatorname{supp} v_{n,u+1}^* \cap R_{n,k}| > 0$). Both are piecewise linear in $L_{n,k}$ (or $R_{n,k}$). For that reason their sum $\lambda_1 v_{n,u}^* + \lambda_2 v_{n,u+1}^*$ is also piecewise linear in $L_{n,k}$ (or $R_{n,k}$). Applying this argument iteratively for each interval $I_{n,k}, k \in \mathbb{Z}$ proves the claim. □

Lemma 4.6. *$F_n^d f$ interpolates $f \in C(\mathbb{R}^d)$ in the nodes*

$$\begin{aligned}
\mathcal{G}_n^{int,d} &:= \{(2^{-(n+1)}k_1, \ldots, 2^{-(n+1)}k_d) : |\boldsymbol{k}|_\infty \leq n2^n\} \\
&= \bigtimes_{i=1}^{d} \Big(\bigcup_{j=0}^{n} \{2^{-j}(k+\frac{1}{2}) : -n2^{n-j} \leq k < n2^{n-j}\} \cup \{-n, \ldots, n\} \Big).
\end{aligned}$$

Proof. To avoid technical issues concerning the order of summation we proof this interpolation property only in the interval $[-1,1]$. Additionally we restrict in the beginning to the case $d = 1$. We use induction. The case $n = 1$ can be easily checked inserting the definitions. Assuming the result holds for F_{n-1}^1 we prove that it holds also for F_n^1. It suffices to prove

$$F_n^1 f\Big(2^{-n}\Big(u+\frac{1}{2}\Big)\Big) = f\Big(2^{-n}\Big(u+\frac{1}{2}\Big)\Big).$$

Since

$$F_n^1 f\Big(2^{-j}\Big(u+\frac{1}{2}\Big)\Big) = F_{n-1}^1 f\Big(2^{-j}\Big(u+\frac{1}{2}\Big)\Big) = f\Big(2^{-j}\Big(u+\frac{1}{2}\Big)\Big)$$

for $0 < j < n$. We obtain

$$\begin{aligned}
F_n^1 f\Big(2^{-n}\Big(u+\frac{1}{2}\Big)\Big)) &= F_{n-1}^1 f\Big(2^{-n}\Big(u+\frac{1}{2}\Big)\Big) + \sum_{-2^j \leq k < 2^j} d_{n,k} \underbrace{v_{n,k}\Big(2^{-n}\Big(u+\frac{1}{2}\Big)\Big)}_{=1} \\
&= F_{n-1}^1 f\Big(2^{-n}\Big(u+\frac{1}{2}\Big)\Big) - \frac{1}{2}[f(2^{-n}u + 2^{-n}) - 2f(2^{-n}u + 2^{-(n+1)}) \\
&\quad + f(2^{-n}u)]. \qquad (4.1.5)
\end{aligned}$$

From Lemma 4.5 we know that F^1_{n-1} is piecewise linear in $[2^{-(n-1)}\frac{u}{2}, 2^{-(n-1)}\frac{u+1}{2}]$. That gives

$$F^1_{n-1}f\Big(2^{-n}\Big(u+\frac{1}{2}\Big)\Big) = \frac{f(2^{-(n-1)}\frac{u}{2}) + f(2^{-(n-1)}\frac{u+1}{2})}{2}.$$

Inserting this into (4.1.5) yields the desired result. The result for F^d_n can be obtained by interpreting F^d_n as an iterated application of F^1_n to each direction of $f \in C(\mathbb{R}^d)$ □

Remark 4.7. *The condition $|\boldsymbol{k}|_\infty \le n2^n$ in the definition of $\mathcal{G}^{int,d}_n$ is due to the fact that the order of summation in Definition 4.4 has the property that with increasing n it covers not only refined dilations of hat functions it covers also new translations on further intervals. This kind of property allows us to prove uniform convergence in the next theorem.*

Theorem 4.8. *Every $f \in C(\mathbb{R}^d)$ can be represented by the series*

$$f = \sum_{\boldsymbol{j}\in\mathbb{N}^d_{-1}} \sum_{\boldsymbol{k}\in\mathbb{Z}^d} d_{\boldsymbol{j},\boldsymbol{k}}(f) v_{\boldsymbol{j},\boldsymbol{k}}$$

with (conditional) convergence in every $C(K)$, where K is a compact subset of $\mathbb{R}^d$. The order of summation should be understood in the sense of $F^d_n f$, $n \to \infty$.

Proof. Basically we extend the arguments in [120] (which are for $[0,1]^2$) to an arbitrary compact set $K \subset \mathbb{R}^d$. Without loss of generality we can assume $K = [A,B]^d, A, B \in \mathbb{Z}$ (K is a cube, since we can always embed a compact set K in such a cube). Let $\varepsilon > 0$. Due to Lemma 4.6 F^d_n interpolates a continuous function in the points

$$\mathcal{G}^{int,d}_n := \{2^{-(n+1)}k_1, \ldots, 2^{-(n+1)}k_d)\boldsymbol{k} : |\boldsymbol{k}|_\infty \le n2^n\}.$$

For $n > n_0$ (sufficient large) the set IF_n is a δ-net of K, that means

$$K \subset \bigcap_{\boldsymbol{x}\in\mathcal{G}^{int,d}_n} \underbrace{\{\boldsymbol{y} \in \mathbb{R}^d : |\boldsymbol{x}-\boldsymbol{y}|_\infty \le \delta\}}_{:=B^\infty_\delta(\boldsymbol{x})}.$$

Consequently, we find for every $\boldsymbol{x} \in K$

$$\boldsymbol{x}^* := \operatorname{argmin}_{\boldsymbol{y}\in\mathcal{G}^{int,d}_n\cap K} |\boldsymbol{x}-\boldsymbol{y}|_\infty$$

with

$$|\boldsymbol{x}-\boldsymbol{x}^*|_\infty < \delta.$$

Continuity of f in $\mathbb{R}^d$ implies uniform continuity on K. That means we find $\delta > 0$ such that

$$|f(\boldsymbol{x}) - f(\boldsymbol{y})| \le \varepsilon \tag{4.1.6}$$

for all $\boldsymbol{x}, \boldsymbol{y} \in K$ with $|\boldsymbol{x}-\boldsymbol{y}| \le \delta$. This allows us to estimate

$$\begin{aligned} |f(\boldsymbol{x}) - F_n f(\boldsymbol{x})| &= |f(\boldsymbol{x}) - f(\boldsymbol{x}^*) + F_n f(\boldsymbol{x}^*) - F_n f(\boldsymbol{x})| \\ &\le |f(\boldsymbol{x}) - f(\boldsymbol{x}^*)| + |F_n f(\boldsymbol{x}^*) - F_n f(\boldsymbol{x})|. \end{aligned}$$

The first summand is bounded by (4.1.6). We have to estimate the second term. Lemma 4.5 gives piecewise linearity of

$$F_n^d f(u_1, \ldots, u_{i-1}, x_i, u_{i+1}, \ldots, u_d) = C F_n^1 f(u_1, \ldots, u_{i-1}, x_i, u_{i+1}, \ldots, u_d)$$

for fixed $\boldsymbol{u} \in \mathbb{R}^{d-1}$, $i = 1, \ldots, d$ and

$$x_i \in [2^{-(n+1)}k, 2^{-(n+1)}(k+1)],$$

where F_n^1 is applied to the i-th direction. Linear functions are monotone. Using this monotonicity iteratively in every single direction we find $\boldsymbol{x}^{**} \in \mathcal{G}_n^{int,d} \cap \overline{B_\delta^\infty(\boldsymbol{x}^*)}$ such that

$$|F_n f(\boldsymbol{x}^*) - F_n f(\boldsymbol{x})| \leq |F_n f(\boldsymbol{x}^*) - F_n f(\boldsymbol{x}^{**})|.$$

Hence,

$$\begin{aligned} |f(\boldsymbol{x}) - F_n f(\boldsymbol{x})| &\leq |f(\boldsymbol{x}) - f(\boldsymbol{x}^*| + |F_n f(\boldsymbol{x}^*) - F_n f(\boldsymbol{x}^{**})| \\ &= |f(\boldsymbol{x}) - f(\boldsymbol{x}^*)| + |f(\boldsymbol{x}^*) - f(\boldsymbol{x}^{**})| \\ &\leq 2\varepsilon. \end{aligned}$$

Since we can proceed in that way for every $\boldsymbol{x} \in K$ and the choice of δ does not depend on $\boldsymbol{x}$ we obtain uniform convergence in K. □

4.2 Sequence spaces

In this section we define discrete function spaces of f and b-type. For the first moment the denotation discrete function space seems unusual, since they consist of sequences of coefficients instead of functions. In the upcoming sections we use the Faber-Schauder system to connect $f \in S^r_{p,\theta}F(\mathbb{R}^d)$ or $f \in S^r_{p,\theta}B(\mathbb{R}^d)$ with a corresponding sequence $\lambda \in s^r_{p,\theta}f$ or $\lambda \in s^r_{p,\theta}b$. For a simplified notation we introduce for $j \in \mathbb{N}_{-1}$, $k \in \mathbb{Z}$ the intervals

$$I_{j,k} := \begin{cases} [2^{-j}k, 2^{-j}(k+1)) & : \quad j \geq 0, \\ [k - \frac{1}{2}, k + \frac{1}{2}) & : \quad j = -1. \end{cases} \tag{4.2.1}$$

For $\boldsymbol{j} \in \mathbb{N}^d_{-1}$ and $\boldsymbol{k} \in \mathbb{Z}^d$ we use the cross product

$$I_{\boldsymbol{j},\boldsymbol{k}} = \bigtimes_{i=1}^{d} I_{j_i,k_i}.$$

This notation allows us to define the characteristic function

$$\chi_{\boldsymbol{j},\boldsymbol{k}}(\boldsymbol{x}) := \begin{cases} 1 & : \quad \boldsymbol{x} \in I_{\boldsymbol{j},\boldsymbol{k}}, \\ 0 & : \quad \text{otherwise.} \end{cases}$$

Definition 4.9. *We define for $0 < p, \theta \leq \infty$ (f-case: $p < \infty$), $\boldsymbol{r} \in \mathbb{R}^d$ the spaces $s^{\boldsymbol{r}}_{p,\theta}f$ and $s^{\boldsymbol{r}}_{p,\theta}b$ as the space of all sequences of coefficients $(\lambda_{\boldsymbol{j},\boldsymbol{k}})_{\boldsymbol{j}\in\mathbb{N}_0^d,\boldsymbol{k}\in\mathbb{Z}^d} \subset \mathbb{C}$ with finite (quasi)-norms*

$$\|\lambda_{\boldsymbol{j},\boldsymbol{k}}|s^{\boldsymbol{r}}_{p,\theta}f\| := \begin{cases} \left\|\left(\sum_{\boldsymbol{j}\in\mathbb{N}^d_{-1}} 2^{\theta \boldsymbol{j}\boldsymbol{r}}|\sum_{\boldsymbol{k}\in\mathbb{Z}^d} \lambda_{\boldsymbol{j},\boldsymbol{k}}\chi_{\boldsymbol{j},\boldsymbol{k}}|^\theta\right)^{\frac{1}{\theta}}\middle|L_p(\mathbb{R}^d)\right\| & : \quad 0 < \theta < \infty, \\ \left\|\sup_{\boldsymbol{j}\in\mathbb{N}^d_{-1}} 2^{\boldsymbol{j}\boldsymbol{r}}|\sum_{\boldsymbol{k}\in\mathbb{Z}^d} \lambda_{\boldsymbol{j},\boldsymbol{k}}\chi_{\boldsymbol{j},\boldsymbol{k}}|\middle|L_p(\mathbb{R}^d)\right\| & : \quad \theta = \infty, \end{cases}$$

and

$$\|\lambda_{\boldsymbol{j},\boldsymbol{k}}|s^{\boldsymbol{r}}_{p,\theta}b\| := \begin{cases} \left(\sum_{\boldsymbol{j}\in\mathbb{N}^d_{-1}} 2^{\theta \boldsymbol{j}\boldsymbol{r}}\|\sum_{\boldsymbol{k}\in\mathbb{Z}^d} \lambda_{\boldsymbol{j},\boldsymbol{k}}\chi_{\boldsymbol{j},\boldsymbol{k}}|L_p(\mathbb{R}^d)\|^\theta\right)^{\frac{1}{\theta}} & : \quad 0 < \theta < \infty, \\ \sup_{\boldsymbol{j}\in\mathbb{N}^d_{-1}} 2^{\boldsymbol{j}\boldsymbol{r}}\|\sum_{\boldsymbol{k}\in\mathbb{Z}^d} \lambda_{\boldsymbol{j},\boldsymbol{k}}\chi_{\boldsymbol{j},\boldsymbol{k}}|L_p(\mathbb{R}^d)\| & : \quad \theta = \infty, \end{cases}$$

respectively.

Analogously to Lemma 3.4 the following embedding results hold for discrete function spaces.

Lemma 4.10. *(i) Let $0 < p \leq \infty$ (f-case: $p < \infty$), $0 < \theta_1 < \theta_2 \leq \infty$ and $\boldsymbol{r} \in \mathbb{R}^d$. Then*

$$s^{\boldsymbol{r}}_{p,\theta_1}f \hookrightarrow s^{\boldsymbol{r}}_{p,\theta_2}f \quad \text{and} \quad s^{\boldsymbol{r}}_{p,\theta_1}b \hookrightarrow s^{\boldsymbol{r}}_{p,\theta_2}b.$$

(ii) Let $0 < p < \infty$, $0 < \theta \leq \infty$ and $\boldsymbol{r} \in \mathbb{R}^d$. Then

$$s^{\boldsymbol{r}}_{p,\min\{p,\theta\}}b \hookrightarrow s^{\boldsymbol{r}}_{p,\theta}f \hookrightarrow s^{\boldsymbol{r}}_{p,\max\{p,\theta\}}b.$$

(iii) Let $0 < p \leq \infty$ (f: $p < \infty$), $0 < \theta, \nu \leq \infty$ and $\boldsymbol{r}_1, \boldsymbol{r}_2 \in \mathbb{R}^d$ with $\boldsymbol{r}_1 > \boldsymbol{r}_2$ Then

$$s^{\boldsymbol{r}_1}_{p,\theta}f \hookrightarrow s^{\boldsymbol{r}_2}_{p,\nu}f \quad \text{and} \quad s^{\boldsymbol{r}_1}_{p,\theta}b \hookrightarrow s^{\boldsymbol{r}_2}_{p,\nu}b.$$

(iv) Let $0 < p < q < \infty$, $0 < \theta, \nu \leq \infty$ and $\boldsymbol{r}_1, \boldsymbol{r}_2 \in \mathbb{R}^d$ fulfilling

$$\boldsymbol{r}_1 - \frac{1}{p} = \boldsymbol{r}_2 - \frac{1}{q}.$$

Then

$$s^{\boldsymbol{r}_1}_{p,\theta}f \hookrightarrow s^{\boldsymbol{r}_2}_{q,\nu}f \quad \text{and} \quad s^{\boldsymbol{r}_1}_{p,\theta}b \hookrightarrow s^{\boldsymbol{r}_2}_{q,\theta}b.$$

Proof. The proofs are similar to that in the references of Lemma 3.4. For the proof of (iv) we refer to [53, Prop. 5.3.3, Prop. 5.3.1]. The proof there adapts the proof in [99] for the diagonal embedding with respect to the Fourier analytical definition. □

4.3 Equivalent characterizations on $\mathbb{R}^d$

In this section we prove equivalent norm characterizations for $S^r_{p,\theta}F(\mathbb{R}^d)$ and $S^r_{p,\theta}B(\mathbb{R}^d)$ by (decreasing) properties of Faber-Schauder coefficients $d_{\boldsymbol{j},\boldsymbol{k}}(f)$. For $S^r_{p,\theta}B([0,1]^d)$ such a characterization was considered in [120]. First we give a characterization for spaces based on $\mathbb{R}^d$. We start with the following technical lemma.

Lemma 4.11. *Let $j \in \mathbb{N}_0$, $k \in \mathbb{Z}$, $\ell \in \mathbb{Z}$ with $j+\ell \geq -1$ and $R>0$ then for local means with* $\operatorname{supp}\Psi_0 \subset [-\frac{1}{2},\frac{1}{2}]$ *and* $\operatorname{supp}\Psi_j \subset [-2^{-j},2^{-j}]$ *(as in Remark 3.6) with $L \geq 2$ the following estimate holds.*

(i) There is a $C_R > 0$ such that

$$|\Psi_j * v_{j+\ell,k}(x)| \leq C_R 2^{-|\ell|}(1+2^{\min\{j,j+\ell\}}|x-x_{j+\ell,k}|)^{-R}.$$

(ii) A refined version of the inequality above is provided by

$$|\Psi_j * v_{j+\ell,k}(x)| \leq C 2^{-|\ell|}\chi_{A_{j+\ell,k}}(x),$$

where $A_{j+\ell,k}$ with $|A_{j+\ell,k}| \asymp 2^{-j}$ is a set that fulfills

$$A_{j+\ell,k} \subset \bigcup_{|u-k|\lesssim 2^{\ell_+}} I_{j+\ell_+,u}. \tag{4.3.1}$$

Proof. First we prove the case $j>0$. The compact supports of Ψ_j and $v_{j,k}$ yield for a non-vanishing integrand of

$$|\Psi_j * v_{j+\ell,k}(x)| = \Big|\int_{\mathbb{R}} 2^{j-1}\Psi_1(2^{j-1}(x-y))v(2^{j+\ell}y-k)dy\Big|.$$

the necessary conditions

$$|x-y| \leq 2^{-j}$$

and additionally for fixed $k \in \mathbb{Z}$

$$|y-x_{j+\ell,k}| \leq 2^{-(j+\ell)}.$$

Triangle inequality implies for a non vanishing integrand

$$|x-x_{j+\ell,k}| \leq |x-y| + |y-x_{j+\ell,k}| \lesssim 2\max\{2^{-j},2^{-(j+\ell)}\}. \tag{4.3.2}$$

Defining $A_{j+\ell,k} := \{x \in \mathbb{R} : |x-x_{j+\ell,k}| \leq 2^{-\min\{j+\ell,j\}}2\}$ we obtain the identity

$$|\Psi_j * v_{j+\ell,k}(x)| = |\Psi_j * v_{j+\ell,k}(x)|\chi_{A_{j+\ell,k}}(x).$$

We proceed considering the case $\ell<0$. Here the support of $v_{j+\ell,k}$ is larger than the support of Ψ_j. The assumption that Ψ_j fulfills moment conditions of order 2 and due to the fact that $v_{j+\ell,k}$ is piecewise linear allows us to shrink the set $A_{j+\ell,k}$ to a set

$A^*_{j+\ell,k} \subset A_{j+\ell,k}$ fulfilling $|A^*_{j+\ell,k}| \leq D2^{-j}$. To be more precise $A^*_{j+\ell,k}$ is the union of 3 intervals of size $\asymp 2^{-j}$ centered in the non-smooth locations of $v_{j+\ell,k}$. A simple change of variable yields

$$\begin{aligned} |\Psi_j * v_{j+\ell,k}(x)| &= \Big| \int_{\mathbb{R}} 2^{j-1}\Psi_1(2^{j-1}(x-y))v(2^{j+\ell}y-k)dy \Big| \\ &= |\Psi_{-\ell} * v_{0,k}(2^{j+\ell}x)|\chi_{A^*_{j+\ell,k}}(x). \end{aligned}$$

The characterization of $B^1_{\infty,\infty}(\mathbb{R})$ by differences easily yields that $v_{0,0} \in B^1_{\infty,\infty}(\mathbb{R})$, cf. [60, Proposition 3.5]. Since Ψ_j is a local mean with $L \geq 2$ we can interpret the convolution as a part of the $B^1_{\infty,\infty}(\mathbb{R})$ norm of $v_{0,k}$. We obtain

$$\begin{aligned} |\Psi_j * v_{j+\ell,k}(x)| &\leq \|v_{0,k}|B^1_{\infty,\infty}(\mathbb{R}^d)\|2^{\ell+1}\chi_{A^*_{j+\ell,k}}(x) \\ &\leq C2^{\ell}\chi_{A^*_{j+\ell,k}}(x). \end{aligned}$$

We continue with the case $\ell \geq 0$ and obtain

$$\begin{aligned} |\Psi_j * v_{j+\ell,k}|(x) &\leq C2^j\|\Psi_1\|_\infty\|v_{j+\ell,k}\|_\infty \int_{2^{-(j+\ell)}k}^{2^{-(j+\ell)}(k+1)} 1dy\chi_{A_{j+\ell,k}}(x) \\ &= D2^{-\ell}\chi_{A_{j+\ell,k}}(x). \end{aligned} \tag{4.3.3}$$

Recognizing the (piecewise) interval structure of $A^*_{j+\ell,k}$ and $A_{j+\ell,k}$ then the inclusion in (4.3.1) follows by simple volume arguments. Finally, concerning the weaker estimate in (i), $|A_{j+\ell,k}| < 2^{-\min\{j+\ell,j\}}$ yields that we find for every $R > 0$ a constant C_R such that

$$\chi_{A^*_{j+\ell,k}}(x) \leq \chi_{A_{j+\ell,k}}(x) \leq C_R(1+2^{\min\{j,j+\ell\}}|x-x_{j,k})^{-R}$$

holds. The case $j = 0$ (where no moment conditions are available) can be estimated with the arguments used to estimate (4.3.3). Formally this computations are for the case $j+\ell \geq 0$. In case $j+\ell = -1$ the slightly shifted translation of the hat function has to be considered. That finishes the proof. □

Lemma 4.12. *Let $\boldsymbol{j} \in \mathbb{N}_0^d$, $\boldsymbol{k} \in \mathbb{Z}^d$, $\boldsymbol{\ell} \in \mathbb{Z}^d$ with $\boldsymbol{j}+\boldsymbol{\ell} \geq -1$ and $R > 0$ then*

(i)

$$|\Psi_{\boldsymbol{j}} * v_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}}(\boldsymbol{x})| \leq C_R 2^{-|\boldsymbol{\ell}|_1}\prod_{i=1}^d(1+2^{\min\{j_i,j_i+\ell_i\}}|x_i - x_{j_i+\ell_i,k_i})^{-R}.$$

(ii) A sharper version of the inequality above is provided by

$$|\Psi_{\boldsymbol{j}} * v_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}}(\boldsymbol{x})| \leq C2^{-|\boldsymbol{\ell}|_1}\chi_{A_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}}}(\boldsymbol{x}),$$

where $A_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}}$ is the cross product of the sets in Lemma 4.11, (ii) with $|A_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}}| \asymp 2^{-|\boldsymbol{j}|_1}$.

Proof. Since $\Psi_{\boldsymbol{j}}$ and $v_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}}$ are tensor products of univariate functions Fubini's theorem allows to write $\Psi_{\boldsymbol{j}} * v_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}}$ as a product of d univariate convolutions. Applying the arguments in Lemma 4.11 to every single factor yields the Lemma stated above. □

For the rest of the paper we use the convention

$$v_{\boldsymbol{j},\boldsymbol{k}} := 0$$

if there exists $i \in [d]$ with $j_i < -1$.

Definition 4.13. *Let $\boldsymbol{v} \in \{0,1\}^d$. For $(\lambda_{\boldsymbol{j},\boldsymbol{k}})_{\boldsymbol{j},\boldsymbol{k}} \in s^{\boldsymbol{r}}_{p,\theta}f$ we define the linear operator*

$$T_{\boldsymbol{v}} : s^{\boldsymbol{r}}_{p,\theta}f \to L_p(\ell_\theta, 2^{\boldsymbol{j}\cdot\boldsymbol{r}})$$

given by

$$(\lambda_{\boldsymbol{j},\boldsymbol{k}})_{\boldsymbol{j}\in\mathbb{N}^d_{-1},\boldsymbol{k}\in\mathbb{Z}^d} \mapsto \Big(\sum_{\boldsymbol{\ell}\in B(\boldsymbol{v})}\sum_{\boldsymbol{k}\in\mathbb{Z}^d} \lambda_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}} \Psi_{\boldsymbol{j}} * v_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}}\Big)_{\boldsymbol{j}\in\mathbb{N}^d_0}$$

where

$$B(\boldsymbol{v}) := \{\boldsymbol{\ell} \in \mathbb{Z}^d : \quad \ell_i \geq 0 \Longleftrightarrow v_i = 1\}.$$

For a sign vector $\boldsymbol{v} \in \{0,1\}^d$ and an integer vector $\boldsymbol{\ell}$ we define

$$\ell_{\boldsymbol{v}} := (\ell^*_1, \ldots, \ell^*_d)$$

where

$$\ell^*_i := \begin{cases} \ell_i & : \quad v_i = 1, \\ 0 & : \quad v_i = 0. \end{cases}$$

Additionally, we define the complement of $\boldsymbol{v}$ by

$$\boldsymbol{v}^c := \boldsymbol{1} - \boldsymbol{v}.$$

Lemma 4.14. *Let $0 < p, \theta < \infty$ ($\theta = \infty$), $\boldsymbol{v} \in \{0,1\}^d$ and $\boldsymbol{r} \in \mathbb{R}^d$ fulfilling*

$$r_i > \sigma_{p,\theta} \text{ if } v_i = 1 \tag{4.3.4}$$

and

$$r_i < 1 \text{ if } v_i = 0, \tag{4.3.5}$$

respectively. Then there is a $C > 0$ such that

$$\|T_{\boldsymbol{v}}\lambda | L_p(\ell_\theta(2^{\boldsymbol{j}\boldsymbol{r}}))\| \leq C \|\lambda | s^r_{p,\theta}f\|.$$

Proof. First we choose a parameter $a < \min\{p,\theta,1\}$ such that $\boldsymbol{r}_i > \frac{1}{a} - 1$ holds for all $i \in [d]$ with $v_i = 1$. We start applying u-triangle inequality in $L_p(\ell_\theta)$ with $u := \min\{p,\theta,1\}$.

$$\|T_{\boldsymbol{v}}\lambda | L_p(\ell_\theta(2^{\boldsymbol{j}\boldsymbol{r}}))\| \;\leq\; \Big(\sum_{\boldsymbol{\ell}\in B(\boldsymbol{v})} \Big\|\Big(\sum_{\boldsymbol{j}\in\mathbb{N}^d_0} 2^{\theta\boldsymbol{j}\cdot\boldsymbol{r}}\Big[\sum_{\boldsymbol{k}\in\mathbb{Z}^d} |\lambda_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}}||\Psi_{\boldsymbol{j}} * v_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}}|\Big]^\theta\Big)^{\frac{1}{\theta}}\Big| L_p(\mathbb{R}^d)\Big\|^u\Big)^{\frac{1}{u}}.$$

Lemma 4.12, (i) yields

$$\|T_{\boldsymbol{v}}\lambda|L_p(\ell_\theta(2^{\boldsymbol{j}\boldsymbol{r}}))\| \lesssim \Big(\sum_{\boldsymbol{\ell}\in B(\boldsymbol{v})} 2^{-u|\boldsymbol{\ell}|_1}$$
$$\times\Big\|\Big(\sum_{\boldsymbol{j}\in\mathbb{N}_0^d} 2^{\theta\boldsymbol{j}\cdot\boldsymbol{r}}\Big[\sum_{\boldsymbol{k}\in\mathbb{Z}^d}|\lambda_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}}|\prod_{i=1}^d(1+2^{\min\{j_i,j_i+\ell_i\}}|x_i-2^{-j_i}k_i|)^{-R}\Big]^\theta\Big)^{\frac{1}{\theta}}\Big|L_p(\mathbb{R}^d)\Big\|^u\Big)^{\frac{1}{u}}.$$

with $R>\frac{1}{a}$. Proceeding by applying Lemma B.13 gives

$$\|T_{\boldsymbol{v}}\lambda|L_p(\ell_\theta(2^{\boldsymbol{j}\boldsymbol{r}}))\| \lesssim \Big(\sum_{\boldsymbol{\ell}\in B(\boldsymbol{v})} 2^{-u|\boldsymbol{\ell}|_1}2^{u|\boldsymbol{\ell}_+|_1/a}$$
$$\times\Big\|\Big(\sum_{\boldsymbol{j}\in\mathbb{N}_0^d} 2^{\theta\boldsymbol{j}\cdot\boldsymbol{r}}\Big[\Big[M\Big|\sum_{\boldsymbol{k}\in\mathbb{Z}^d}\lambda_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}}\chi_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}}\Big|^a(\boldsymbol{x})\Big]^{\frac{1}{a}}\Big]^\theta\Big)^{\frac{1}{\theta}}\Big|L_p(\mathbb{R}^d)\Big\|^u\Big)^{\frac{1}{u}}.$$

We observe the trivial identity $\|(M|f_j|^a)^{\frac{1}{a}}|L_p(\ell_\theta)\| = \|(M|f_j|^a)|L_{\frac{p}{a}}(\ell_{\frac{\theta}{a}})\|^{\frac{1}{a}}$. The fact $\min\{\frac{p}{a},\frac{\theta}{a}\}>1$ allows us to apply Theorem B.6 which yields

$$\|T_{\boldsymbol{v}}\lambda|L_p(\ell_\theta(2^{\boldsymbol{j}\boldsymbol{r}}))\| \lesssim \Big(\sum_{\boldsymbol{\ell}\in B(\boldsymbol{v})} 2^{-u|\boldsymbol{\ell}|_1}2^{u|\boldsymbol{\ell}_+|_1/a}$$
$$\times\Big\|\Big(\sum_{\boldsymbol{j}\in\mathbb{N}_0^d} 2^{\theta\boldsymbol{j}\cdot\boldsymbol{r}}\Big|\sum_{\boldsymbol{k}\in\mathbb{Z}^d}\lambda_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}}\chi_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}}(\boldsymbol{x})\Big|^\theta\Big)^{\frac{1}{\theta}}\Big|L_p(\mathbb{R}^d)\Big\|^u\Big)^{\frac{1}{u}}.$$

Extending the summation index implies

$$\|T_{\boldsymbol{v}}\lambda|L_p(\ell_\theta(2^{\boldsymbol{j}\boldsymbol{r}}))\| \lesssim \|\lambda|s^{\boldsymbol{r}}_{p,\theta}f\|\Big(\sum_{\boldsymbol{\ell}\in B(\boldsymbol{v})} 2^{-u\boldsymbol{\ell}_{\boldsymbol{v}}\cdot(\boldsymbol{r}-(\frac{1}{a}-\boldsymbol{1}))}2^{u\boldsymbol{\ell}_{v^c}(\boldsymbol{1}-\boldsymbol{r})}\Big)^{\frac{1}{u}}.$$

Due to the choice of $\boldsymbol{r}$ in (4.3.4) and (4.3.5) the sum converges to a constant if a is chosen sufficient close to $\min\{p,\theta,1\}$. That finishes the proof. □

Lemma 4.15. *Let $0<p,\theta\le\infty$, $\boldsymbol{v}\in\{0,1\}^d$ and $\boldsymbol{r}\in\mathbb{R}^d$ fulfilling*

$$r_i>\sigma_p,\quad v_i=1 \tag{4.3.6}$$

and

$$r_i<1+\frac{1}{p},\quad v_i=0. \tag{4.3.7}$$

Then

$$\|T_{\boldsymbol{v}}\lambda|\ell_\theta(2^{\boldsymbol{j}\cdot\boldsymbol{r}},L_p)\|\lesssim\|\lambda|s^{\boldsymbol{r}}_{p,\theta}b\|.$$

Proof. We restrict our proof to the case $p,\theta<\infty$, the modifications in case $p,\theta=\infty$ are obvious. For a shorter notation we define

$$G_{\boldsymbol{\ell}}(\boldsymbol{k}) := \{\boldsymbol{u}\in\mathbb{Z}^d : |u_i-k_i|\le C2^{(\ell_i)_+}, i\in[d]\},$$

where $C > 0$ is the constant (4.3.1). We start applying u-triangle inequality in $\ell_\theta(2^{\boldsymbol{j}\cdot\boldsymbol{r}}, L_p(\mathbb{R}^d))$ with $u := \min\{p, 1\}$.

$$\|T_{\boldsymbol{v}}\lambda|\ell_\theta(2^{\boldsymbol{j}\cdot\boldsymbol{r}}, L_p)\| \leq \Big(\sum_{\boldsymbol{\ell}\in B(\boldsymbol{v})}\Big(\sum_{\boldsymbol{j}\in\mathbb{N}_0^d} 2^{\theta\boldsymbol{j}\cdot\boldsymbol{r}}\Big\|\sum_{\boldsymbol{k}\in\mathbb{Z}^d}|\lambda_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}}||\Psi_{\boldsymbol{j}} * v_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}}|\Big|L_p(\mathbb{R}^d)\Big\|^\theta\Big)^{\frac{u}{\theta}}\Big)^{\frac{1}{u}}.$$

Applying Lemma 4.12 yields

$$\begin{aligned}&\|T_{\boldsymbol{v}}\lambda|\ell_\theta(2^{\boldsymbol{j}\cdot\boldsymbol{r}}, L_p(\mathbb{R}^d))\| \\ \leq\ &\Big(\sum_{\boldsymbol{\ell}\in B(\boldsymbol{v})} 2^{-u|\boldsymbol{\ell}|_1}\Big(\sum_{\boldsymbol{j}\in\mathbb{N}_0^d} 2^{\theta\boldsymbol{j}\cdot\boldsymbol{r}}\Big\|\sum_{\boldsymbol{k}\in\mathbb{Z}^d}|\lambda_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}}|\chi_{A_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}}}(\boldsymbol{x})\Big|L_p(\mathbb{R}^d)\Big\|^\theta\Big)^{\frac{u}{\theta}}\Big)^{\frac{1}{u}}.\end{aligned} \tag{4.3.8}$$

The fact $A_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}} \subset \bigcup_{|u_i-k_i|\leq 2^{\ell_{i,+}}} I_{\boldsymbol{j}+\boldsymbol{\ell_v},\boldsymbol{u}}$ allows to decompose

$$\Big[\sum_{\boldsymbol{k}\in\mathbb{Z}^d}|\lambda_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}}|\chi_{A_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}}}(\boldsymbol{x})\Big]^p \leq \Big[\sum_{\boldsymbol{k}\in\mathbb{Z}^d}|\lambda_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}}|\sum_{\boldsymbol{u}\in G_{\boldsymbol{\ell}}(\boldsymbol{k})}\chi_{I_{\boldsymbol{j}+\boldsymbol{\ell_v},\boldsymbol{u}}}(\boldsymbol{x})\Big]^p.$$

Interchanging summation yields

$$\Big[\sum_{\boldsymbol{k}\in\mathbb{Z}^d}|\lambda_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}}|\sum_{\boldsymbol{u}\in G_{\boldsymbol{\ell}}(\boldsymbol{k})}\chi_{I_{\boldsymbol{j}+\boldsymbol{\ell_v},\boldsymbol{u}}}(\boldsymbol{x})\Big]^p = \Big[\sum_{\boldsymbol{u}\in\mathbb{Z}^d}\chi_{I_{\boldsymbol{j}+\boldsymbol{\ell_v},\boldsymbol{u}}}(\boldsymbol{x})\sum_{\boldsymbol{k}:\ \boldsymbol{u}\in G_{\boldsymbol{\ell}}(\boldsymbol{k})}|\lambda_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}}|\Big]^p.$$

Disjoint supports of $I_{\boldsymbol{j}+\boldsymbol{\ell_v},\boldsymbol{u}}$ for different $\boldsymbol{u} \in \mathbb{Z}^d$ yield

$$\Big[\sum_{\boldsymbol{u}\in\mathbb{Z}^d}\chi_{I_{\boldsymbol{j}+\boldsymbol{\ell_v},\boldsymbol{u}}}(\boldsymbol{x})\sum_{\boldsymbol{k}:\ \boldsymbol{u}\in G_{\boldsymbol{\ell}}(\boldsymbol{k})}|\lambda_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}}|\Big]^p = \sum_{\boldsymbol{u}\in\mathbb{Z}^d}\chi_{I_{\boldsymbol{j}+\boldsymbol{\ell_v},\boldsymbol{u}}}(\boldsymbol{x})\Big[\sum_{\boldsymbol{k}:\ \boldsymbol{u}\in G_{\boldsymbol{\ell}}(\boldsymbol{k})}|\lambda_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}}|\Big]^p.$$

Taking the structure of $G_{\boldsymbol{\ell}}(\boldsymbol{k})$ into account with $|G_{\boldsymbol{\ell}}(\boldsymbol{k})| \asymp 2^{|\boldsymbol{\ell}_+|_1}$ then Hölder's inequality in case $p > 1$ or simply the embedding $\ell_p \hookrightarrow \ell_1$ in case $p < 1$ respectively, implies

$$\Big[\sum_{\boldsymbol{k}\in\mathbb{Z}^d}|\lambda_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}}|\chi_{A_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}}}(\boldsymbol{x})\Big]^p \lesssim 2^{p|\boldsymbol{\ell_v}|_1(1-\frac{1}{p})_+}\sum_{\boldsymbol{u}\in\mathbb{Z}^d}\chi_{I_{\boldsymbol{j}+\boldsymbol{\ell_v},\boldsymbol{u}}}(\boldsymbol{x})\sum_{\boldsymbol{k}:\ \boldsymbol{u}\in G_{\boldsymbol{\ell}}(\boldsymbol{k})}|\lambda_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}}|^p.$$

Furthermore it yields

$$\sum_{\boldsymbol{u}\in\mathbb{Z}^d}\sum_{\boldsymbol{k}:\ \boldsymbol{u}\in G_{\boldsymbol{\ell}}(\boldsymbol{k})}|\lambda_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}}|^p \leq 2^{|\boldsymbol{\ell}_v|_1}\sum_{\boldsymbol{k}\in\mathbb{Z}^d}|\lambda_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}}|^p.$$

Considering the $L_p(\mathbb{R}^d)$ norm gives

$$\Big\|\sum_{\boldsymbol{k}\in\mathbb{Z}^d}|\lambda_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}}|\chi_{A_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}}}(\boldsymbol{x})\Big|L_p(\mathbb{R}^d)\Big\|^p \lesssim 2^{-|\boldsymbol{j}+\boldsymbol{\ell}|_1}2^{-|\boldsymbol{\ell}_{\boldsymbol{v}^c}|_1}2^{|\boldsymbol{\ell_v}|_1[(p-1)_++1]}\sum_{\boldsymbol{k}\in\mathbb{Z}^d}|\lambda_{\boldsymbol{j}+\boldsymbol{\ell},k}|^p.$$

Inserting this into 4.3.8 yields

$$\|T_{\boldsymbol{v}}\lambda|\ell_\theta(2^{\boldsymbol{j}\cdot\boldsymbol{r}},L_p(\mathbb{R}^d))\|$$
$$\lesssim \Big(\sum_{\boldsymbol{\ell}\in B(\boldsymbol{v})} 2^{-u\boldsymbol{\ell}_{\boldsymbol{v}}\cdot\left[\boldsymbol{r}_{\boldsymbol{v}}-(\frac{1}{p}-1)_+\right]}2^{u\boldsymbol{\ell}_{\boldsymbol{v}^c}\cdot\left[1+\frac{1}{p}-\boldsymbol{r}_{\boldsymbol{v}^c}\right]}\Big[\sum_{\boldsymbol{j}\in\mathbb{N}_0^d} 2^{\theta(\boldsymbol{j}+\boldsymbol{\ell})\cdot(\boldsymbol{r}-\frac{1}{p})}\Big(\sum_{\boldsymbol{k}\in\mathbb{Z}^d}|\lambda_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}}|^p\Big)^{\frac{\theta}{p}}\Big]^{\frac{u}{\theta}}\Big)^{\frac{1}{u}}.$$

Finally extending the summation index gives

$$\|T_{\boldsymbol{v}}\lambda|\ell_\theta(2^{\boldsymbol{j}\cdot\boldsymbol{r}},L_p(\mathbb{R}^d))\| \lesssim \Big(\sum_{\boldsymbol{\ell}\in B(\boldsymbol{v})} 2^{-u\boldsymbol{\ell}_{\boldsymbol{v}}\cdot\left[\boldsymbol{r}-(\frac{1}{p}-1)_+\right]}2^{u\boldsymbol{\ell}_{\boldsymbol{v}^c}\cdot\left[1+\frac{1}{p}-\boldsymbol{r}\right]}\Big)^{\frac{1}{u}}\|\lambda|s^r_{p,\theta}b\|.$$

Due to the choice of the parameters in (4.3.6) and (4.3.7) the sum converges to an absolute constant. That proves the claim. □

Theorem 4.16. *Let $0<p,\theta<\infty$ ($\theta=\infty$) and for $r\geq 1$*

$$1+\min\Big\{\frac{1}{p},\frac{1}{\theta}\Big\}>r>\begin{cases}\sigma_{p,\theta} & : \quad \min\{p,\theta\}>1 \text{ or } \max\{p,\theta\}<1,\\ \left|\frac{1}{p}-\frac{1}{\theta}\right| & : \quad \min\{p,\theta\}\leq 1\leq\max\{p,\theta\},\end{cases} \tag{4.3.9}$$

or in case $r<1$ simply

$$1>r>\sigma_{p,\theta}$$

being fulfilled. Further let $\lambda\in s^r_{p,\theta}f$. Then

(i)

$$f:=\sum_{\boldsymbol{j}\in\mathbb{N}^d_{-1}}\sum_{\boldsymbol{k}\in\mathbb{Z}^d}\lambda_{\boldsymbol{j},\boldsymbol{k}}v_{\boldsymbol{j},\boldsymbol{k}}(\cdot)$$

converges unconditionally in every $S^{r-\varepsilon}_{p,\nu}F(\mathbb{R}^d)$ with $0<\nu\leq\infty$ and $\varepsilon>0$. In case $\theta<\infty$ there is unconditional convergence in $S^r_{p,\theta}F(\mathbb{R}^d)$, itself.

(ii) Additionally, there is a constant $C>0$ such that

$$\|f|S^r_{p,\theta}F(\mathbb{R}^d)\|\leq C\|\lambda|s^r_{p,\theta}f\||. \tag{4.3.10}$$

holds.

Proof. Step 1. We assume the unconditional convergence of f in at least $L_1(\mathbb{R}^d)$ and prove (ii). We start representing the norm in terms of local means, cf. Theorem 3.7

$$\begin{aligned}\|f|S^r_{p,\theta}F(\mathbb{R}^d)\| &= \Big\|\Big(\sum_{\boldsymbol{j}\in\mathbb{N}_0^d}2^{\theta|\boldsymbol{j}|_1 r}|\Psi_{\boldsymbol{j}}*f|^\theta\Big)^{\frac{1}{\theta}}\Big|L_p(\mathbb{R}^d)\Big\|\\ &\leq \Big(\sum_{\boldsymbol{v}\in\{0,1\}^d}\Big\|\Big(\sum_{\boldsymbol{j}\in\mathbb{N}_0^d}2^{\theta|\boldsymbol{j}|_1 r}\Big|\sum_{\boldsymbol{\ell}\in B(\boldsymbol{v})}\sum_{\boldsymbol{k}\in\mathbb{Z}^d}\lambda_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}}\Psi_{\boldsymbol{j}}*v_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}}\Big|^\theta\Big)^{\frac{1}{\theta}}\Big|L_p(\mathbb{R}^d)\Big\|^u\Big)^{\frac{1}{u}}\\ &= \Big(\sum_{\boldsymbol{v}\in\{0,1\}^d}\|T_{\boldsymbol{v}}\lambda|L_p(\ell_\theta(2^{|\boldsymbol{j}|_1 r}))|L_p(\mathbb{R}^d)\|^u\Big)^{\frac{1}{u}}\end{aligned}$$

with $u=\min\{p,\theta,1\}$. In case $\sigma_{p,\theta}<r<1$ applying Lemma 4.14 finishes the proof. In case $\max\{1,\sigma_{p,\theta}\}\le r<1+\min\{\frac{1}{p},\frac{1}{\theta}\}$ we use complex interpolation of quasi Banach spaces, cf. [127] and the references therein, to prove the boundedness of

$$\|T_{\boldsymbol{v}}\lambda|s^r_{p,\theta}f\to L_p(\ell_\theta(2^{|\boldsymbol{j}|_1 r}))\|.$$

The basic idea is borrowed from [120, Proposition]. We distinguish two cases. First we consider the case $\frac{1}{\theta}\le\frac{1}{p}$, where we use the interpolation identities (Banach case: cf. [118, Sec. 1.18.1 and 1.18.4], quasi-Banach case: [127, Chap. 4])

$$s^r_{p,\theta}f=[s^{\boldsymbol{r}_0}_{p_0,\theta}f,s^{\boldsymbol{r}_1}_{\theta,\theta}f]_\nu,\quad L_p(\ell_\theta(2^{|\boldsymbol{j}|_1 r}))=[L_{p_0}(\ell_\theta(2^{\boldsymbol{j}\cdot\boldsymbol{r_0}})),L_\theta(\ell_\theta(2^{\boldsymbol{j}\cdot\boldsymbol{r_1}}))]_\nu$$

for a $\nu\in(0,1)$, $0<p_0<\infty$ such that

$$\frac{1}{p}=\frac{1-\nu}{p_0}+\frac{\nu}{\theta} \tag{4.3.11}$$

and

$$r\mathbf{1}=(1-\nu)\boldsymbol{r_0}+\nu\boldsymbol{r_1},$$

where $\boldsymbol{r}_0,\boldsymbol{r}_1\in\mathbb{R}^d$ such that

$$\begin{cases} r_0^i>\sigma_{p_0,\theta} & : \quad i\in[d] \text{ with } v_i=1\\ r_0^i<1 & : \quad i\in[d] \text{ with } v_i=0\end{cases} \quad\text{and}\quad \begin{cases} r_1^i>\sigma_\theta & : \quad i\in[d] \text{ with } v_i=1\\ r_1^i<1+\frac{1}{\theta} & : \quad i\in[d] \text{ with } v_i=0\end{cases}$$

are fulfilled. For convenience of the reader we explain how to choose these interpolation parameters. We set

$$r_1^i=1+\frac{1}{\theta}-\varepsilon>r,$$

where ε is sufficient small, for all $i\in[d]$ with $v_i=0$. Additionally, we fix

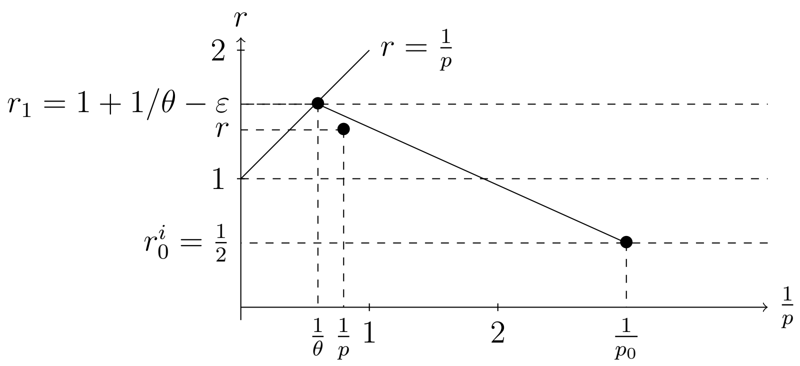
$$r_0^i=\frac{1}{2}$$

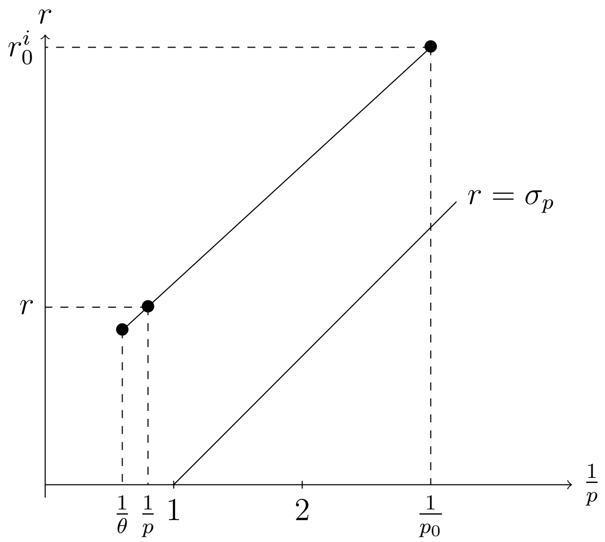
for all $i\in[d]$ with $v_i=0$. From this we determine $\nu\in(0,1)$. Clearly, since $\frac{1}{\theta}<\frac{1}{p}$ we will find $p_0\in(0,p)$ such that

$$\frac{1}{p}=\frac{1-\nu}{p_0}+\frac{\nu}{\theta}.$$

It remains to choose r_0^i and r_1^i for all $i\in[d]$ with $v_i=1$. p_0 can be become very small. For that reason we choose $r>r_1^i=\sigma_\theta+\varepsilon$ with ε sufficient small such that

$$\frac{r-r_1^i}{\frac{1}{p}-\frac{1}{\theta}}\ge 1 \tag{4.3.12}$$





is fulfilled. This implies the condition

$$r > \frac{1}{p} - \frac{1}{\theta} - \sigma_\theta = \begin{cases} \frac{1}{p} - \frac{1}{\theta} & , \quad \theta \geq 1, \\ \frac{1}{p} - 1 & , \quad \theta < 1. \end{cases}$$

In case $\frac{1}{\theta} < \frac{1}{p} \leq 1$ (4.3.12) is always fulfilled since we are in case $r \geq 1$. (4.3.12) guarantees to find $r_0^i > \sigma_{p_0,\theta} = \sigma_{p_0}$ fulfilling

$$r = (1-\nu) r_0^i + \nu r_1^i$$

for all $i \in [d]$ with $v_i = 1$, since the derivation of σ_{p_0} is smaller or equal to 1 in p_0. This finishes the case $\frac{1}{\theta} < \frac{1}{p}$. The case $\frac{1}{\theta} \geq \frac{1}{p}$ works similar. Here we interpolate,

$$s^r_{p,\theta} f = [s^{\boldsymbol{r}_0}_{p,\theta_0} f, s^{\boldsymbol{r}_1}_{p,p} b]_\nu, \quad L_p(\ell_\theta(2^{|\boldsymbol{j}|_1 r})) = [L_p(\ell_{\theta_0}(2^{\boldsymbol{j}\cdot\boldsymbol{r}_0})), L_p(\ell_p(2^{\boldsymbol{j}\cdot\boldsymbol{r}_1}))]_\nu.$$

The parameters are chosen analogously, where the role of p_0 is replaced by θ_0. *Step 2.* We show the unconditional convergence of f in $S^r_{p,\theta}F(\mathbb{R}^d)$. We prove (i) in case $\theta < \infty$. To begin with, we denote the set of Faber indices by $\nabla = \{(\boldsymbol{j}, \boldsymbol{k}) : \boldsymbol{j} \in \mathbb{N}^d_{-1}, \boldsymbol{k} \in \mathbb{Z}^d\}$.

Based on this we define the set of sequences with finite index sets given by

$$\mathfrak{E} := \Big\{ \mathcal{E} = (\mathcal{E}_n)_{n\in\mathbb{N}} : \ \mathcal{E}_n \subset \nabla, \ |\mathcal{E}_n| = n, \ \mathcal{E}_n \subset \mathcal{E}_{n+1} \text{ for all } n \in \mathbb{N}, \text{ and } \bigcup_{n=1}^{\infty} \mathcal{E}_n = \nabla \Big\}.$$

Every sequence in $\mathfrak{E}$ defines an order of summation. Furthermore for $\mathcal{E} \in \mathfrak{E}$ we define $F_{\mathcal{E}_n} := \sum_{(\boldsymbol{j},\boldsymbol{k})\in\mathcal{E}_n} \lambda_{\boldsymbol{j},\boldsymbol{k}} v_{\boldsymbol{j},\boldsymbol{k}}$. We take a second sequence $A \in \mathfrak{E}$ and consider $F_{\mathcal{E}_n} - F_{\mathcal{A}_m}$. This difference can be written as a sum with finitely many $\lambda_{\boldsymbol{j},\boldsymbol{k}}$. This fulfills the assumptions necessary in Step 1 and yields

$$\|F_{\mathcal{E}_n} - F_{\mathcal{A}_m} | S^r_{p,\theta}F(\mathbb{R}^d)\| \lesssim \Big\| \Big(\sum_{(\boldsymbol{j},\boldsymbol{k})\in(\mathcal{E}_n\cup\mathcal{A}_m)\setminus(\mathcal{E}_n\cap\mathcal{A}_m)} 2^{r|\boldsymbol{j}|_1\theta} |\lambda_{\boldsymbol{j},\boldsymbol{k}}|^\theta \chi_{\boldsymbol{j},\boldsymbol{k}} \Big)^{\frac{1}{\theta}} \Big| L_p(\mathbb{R}^d) \Big\|.$$

Due to the disjoint support of $\chi_{\boldsymbol{j},\boldsymbol{k}_1}$ and $\chi_{\boldsymbol{j},\boldsymbol{k}_2}$ for $\boldsymbol{k}_1 \neq \boldsymbol{k}_2$ we obtain that

$$\begin{aligned} &\Big(\sum_{(\boldsymbol{j},\boldsymbol{k})\in(\mathcal{E}_n\cup\mathcal{A}_m)\setminus(\mathcal{E}_n\cap\mathcal{A}_m)} 2^{r|\boldsymbol{j}|_1\theta} |\lambda_{\boldsymbol{j},\boldsymbol{k}}|^\theta \chi_{\boldsymbol{j},\boldsymbol{k}} \Big)^{\frac{1}{\theta}} \\ \leq\ &\Big(\sum_{\boldsymbol{j}\in\mathbb{N}^d_{-1}} 2^{r|\boldsymbol{j}|_1\theta} \Big| \sum_{\boldsymbol{k}\in\mathbb{Z}^d} \lambda_{\boldsymbol{j},\boldsymbol{k}} v_{\boldsymbol{j},\boldsymbol{k}} \Big|^\theta \Big)^{\frac{1}{\theta}} \in L_p(\mathbb{R}^d) \end{aligned}$$

holds almost everywhere. Therefore Lebesgue's dominated convergence theorem yields that we find for every $\varepsilon > 0$ a $n_0 \in \mathbb{N}$ such that

$$\|F_{\mathcal{E}_n} - F_{\mathcal{A}_m} | S^r_{p,\theta}F(\mathbb{R}^d)\| \leq \varepsilon$$

for all $m, n > n_0$. Finally this implies unconditional convergence in $S^r_{p,\theta}F(\mathbb{R}^d)$. In case $\theta = \infty$ we stress on the embeddings

$$S^s_{p,1}F(\mathbb{R}^d) \hookrightarrow S^{\tilde{r}}_{p,\nu}F(\mathbb{R}^d)$$

and

$$\|\lambda | s^s_{p,1}f\| \lesssim \|\lambda | s^r_{p,\infty}f\|,$$

where $r > s > \sigma_{p,\nu}$, $s > \tilde{r}$ and $0 < \nu \leq \infty$. Applying the arguments from above to $S^s_{p,1}F(\mathbb{R}^d)$ yields the result for $S^{\tilde{r}}_{p,\nu}F(\mathbb{R}^d)$. □

Remark 4.17. *The conditions on r in Theorem 4.16 look partly unnatural and are probably not sharp. This seems to be a technical issue of the interpolation technique. One would expect that this result holds for all $\sigma_{p,\theta} < r < 1+\min\{\frac{1}{p}, \frac{1}{\theta}\}$ with $0 < p, \theta < \infty$ ($\theta = \infty$). Nevertheless our technique works for all $0 < p, \theta < \infty$ ($\theta = \infty$) with r such that $\max\{\frac{1}{p}, \frac{1}{\theta}\} < r < 1 + \min\{\frac{1}{p}, \frac{1}{\theta}\}$, which is important for an equivalent characterization we will give later.*

The next Theorem is the B-case analog of Theorem 4.16.

Theorem 4.18. *Let $0 < p, \theta \leq \infty$ and $\sigma_p < r < 1 + \frac{1}{p}$. Further let $\lambda \in s^r_{p,\theta}b$. Then*

(i)

$$f := \sum_{\boldsymbol{j}\in\mathbb{N}_{-1}^d} \sum_{\boldsymbol{k}\in\mathbb{Z}^d} \lambda_{\boldsymbol{j},\boldsymbol{k}} v_{\boldsymbol{j},\boldsymbol{k}}(\cdot)$$

converges unconditionally in every $S^{r-\varepsilon}_{p,\nu}F(\mathbb{R}^d)$ *with* $0 < \nu \le \infty$ *and* $\varepsilon > 0$. *In case* $\max\{p,\theta\} < \infty$ *there is unconditional convergence in* $S^r_{p,\theta}B(\mathbb{R}^d)$, *itself.*

(ii) Additionally, there is a constant $C > 0$ *such that*

$$\|f|S^r_{p,\theta}B(\mathbb{R}^d)\| \le C\|\lambda|s^r_{p,\theta}b|\|. \tag{4.3.13}$$

holds.

Proof. The proof is a trivial B-case modification of Theorem 4.16. The inequality in (ii) can be obtained in the following way

$$\begin{aligned}
\|f|S^r_{p,\theta}B(\mathbb{R}^d)\| &= \Big(\sum_{\boldsymbol{j}\in\mathbb{N}_0^d} 2^{\theta|\boldsymbol{j}|_1 r}\|\Psi_{\boldsymbol{j}} * f\|_p^\theta\Big)^{\frac{1}{\theta}} \\
&\le \Bigg(\sum_{\boldsymbol{v}\in\{0,1\}^d}\Big(\sum_{\boldsymbol{j}\in\mathbb{N}_0^d} 2^{\theta|\boldsymbol{j}|_1 r}\Big\|\sum_{\boldsymbol{\ell}\in B(\boldsymbol{v})}\sum_{\boldsymbol{k}\in\mathbb{Z}^d}\lambda_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}}\Psi_{\boldsymbol{j}} * v_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}}\Big|L_p(\mathbb{R}^d)\Big\|^\theta\Big)^{\frac{u}{\theta}}\Bigg)^{\frac{1}{u}} \\
&= \Big(\sum_{\boldsymbol{v}\in\{0,1\}^d}\|T_{\boldsymbol{v}}\lambda|\ell_\theta(2^{|\boldsymbol{j}|_1 r}, L_p)|L_p(\mathbb{R}^d)\|^u\Big)^{\frac{1}{u}}
\end{aligned}$$

Inserting the estimate from Lemma 4.15 finishes the proof. □

Theorem 4.19. *(i) Let* $\frac{1}{2} < p,\theta \le \infty$ *(*$p < \infty$*) and* $\max\{\frac{1}{p},\frac{1}{\theta}\} < r < 2$. *Then for* $f \in S^r_{p,\theta}F(\mathbb{R}^d)$ *the inequality*

$$\|d_{\boldsymbol{j},\boldsymbol{k}}(f)|s^r_{p,\theta}f\| \lesssim \|f|S^r_{p,\theta}F(\mathbb{R}^d)\|$$

holds.

(ii) Let $0 < p,\theta \le \infty$ *(*$p > \frac{1}{2}$*) and* $\frac{1}{p} < r < 2$. *Then for* $f \in S^r_{p,\theta}B(\mathbb{R}^d)$ *the inequality*

$$\|d_{\boldsymbol{j},\boldsymbol{k}}(f)|s^r_{p,\theta}b\| \lesssim \|f|S^r_{p,\theta}B(\mathbb{R}^d)\|$$

holds.

Proof. The proof provided here is a trivial modification of [60, Proposition 3.4], where the B-case was considered in the periodic setting. We prove the F-case. We use for fixed $\boldsymbol{j} \in \mathbb{N}^d_{-1}$ the pointwise decomposition

$$f = \sum_{\boldsymbol{\ell}\in\mathbb{Z}^d} \delta_{\boldsymbol{j}+\boldsymbol{\ell}}[f]$$

with $\delta_{\boldsymbol{j}+\boldsymbol{\ell}}[f]$ as in (3.1.1) and restrict first to the F-case. This allows us to estimate

$$\begin{aligned}\|d_{\boldsymbol{j},\boldsymbol{k}}(f)|s^r_{p,\theta}f\| &= \Big\|\Big(\sum_{\boldsymbol{j}\in\mathbb{N}^d_{-1}} 2^{r|\boldsymbol{j}|_1\theta}\Big|\sum_{\boldsymbol{k}\in\mathbb{Z}^d} d_{\boldsymbol{j},\boldsymbol{k}}(f)\chi_{\boldsymbol{j},\boldsymbol{k}}(\cdot)\Big|^\theta\Big)^{\frac{1}{\theta}}\Big|L_p(\mathbb{R}^d)\Big\| \\ &\le \Big(\sum_{\boldsymbol{\ell}\in\mathbb{Z}^d}\Big\|\Big(\sum_{\boldsymbol{j}\in\mathbb{N}^d_{-1}} 2^{r|\boldsymbol{j}|_1\theta}\Big|\sum_{\boldsymbol{k}\in\mathbb{Z}^d} d_{\boldsymbol{j},\boldsymbol{k}}(\delta_{\boldsymbol{j}+\boldsymbol{\ell}}[f])\chi_{\boldsymbol{j},\boldsymbol{k}}(\cdot)\Big|^\theta\Big)^{\frac{1}{\theta}}\Big|L_p(\mathbb{R}^d)\Big\|^u\Big)^{\frac{1}{u}}.\end{aligned} \tag{4.3.14}$$

where $u=\min\{p,\theta,1\}$. We consider

$$F_{\boldsymbol{j},\boldsymbol{\ell}}(\boldsymbol{x}) := \sum_{\boldsymbol{k}\in\mathbb{Z}^d} d_{\boldsymbol{j},\boldsymbol{k}}(\delta_{\boldsymbol{j}+\boldsymbol{\ell}}[f])\chi_{\boldsymbol{j},\boldsymbol{k}}(\boldsymbol{x}).$$

Clearly, whenever $\boldsymbol{x}\in I_{\boldsymbol{j},\boldsymbol{k}}$ we have

$$|F_{\boldsymbol{j},\boldsymbol{\ell}}(\boldsymbol{x})| \le \Big|d_{\boldsymbol{j},\boldsymbol{k}}(\delta_{\boldsymbol{j}+\boldsymbol{\ell}}[f])\Big| \lesssim |\Delta^{2,e(\boldsymbol{j})}_{2^{-(\boldsymbol{j}+1)}}(\delta_{\boldsymbol{j}+\boldsymbol{\ell}}[f],\boldsymbol{x}_{\boldsymbol{j},\boldsymbol{k}})|. \tag{4.3.15}$$

We estimate the iterated differences $\Delta^{2,e(\boldsymbol{j})}_{2^{-(\boldsymbol{j}+1)}}(\delta_{\boldsymbol{j}+\boldsymbol{\ell}}[f],\boldsymbol{x}_{\boldsymbol{j},\boldsymbol{k}})$ one by one now. Let $g_{j_i+\ell_i}(t)$ be an univariate bandlimited function with frequencies in $[-A2^{j_i+\ell_i}, B2^{j_i+\ell_i}]$. We start with the case $i\in e(\boldsymbol{j})$ ($j_i\ge 0$). Here, Lemma B.10 yields

$$|\Delta^{2,i}_{2^{-(j_i+1)}}(g_{j_i+\ell_i},x_{j_i,k_i})| \lesssim \min\{1,2^{2\ell_i}\}\max\{1,2^{\ell_i a}\}P_{2^{\ell_i+j_i},a|i}g_{j_i+\ell_i}(x_{j_i,k_i}).$$

Obviously, $x_i\in I_{j_i,k_i}$ implies $|x_{j_i,k_i}-x_i|\le 2^{-(j_i)_+}$. For that reason Lemma B.12 gives

$$|\Delta^{2,i}_{2^{-(j_i+1)}}(g_{j_i+\ell_i},x_{j_i,k_i})| \lesssim \min\{2^{2\ell_i,1}\}\max\{2^{\ell_i a},1\}P_{2^{\ell_i+j_i},a|i}g_{j_i+\ell_i}(x). \tag{4.3.16}$$

In case $i\notin e(\boldsymbol{j})$ we have $j_i=-1$ and

$$|g_{j_i+\ell_i}(k)| \le \sup_{|y|\le 1}|g_{j_i+\ell_i}(x+y)| \lesssim \sup_{|y|\le 1}\frac{|g_{j_i+\ell_i}(x+y)|}{(1+2^{j_i}|y|)^a} \le P_{2^{j_i},a|i}g_{j_i+\ell_i}(x). \tag{4.3.17}$$

Additionally, assuming $\ell_i\ge 0$ then Lemma B.12, (ii) yields

$$|g_{j_i+\ell_i}(k)| \le 2^{\ell_i a}P_{2^{j_i+\ell_i},a|i}g_{j_i+\ell_i}(x) \tag{4.3.18}$$

Applying iteratively pointwise estimates (4.3.18) and (4.3.16) to the right hand of (4.3.15) yields

$$|F_{\boldsymbol{j},\boldsymbol{\ell}}(x)| \lesssim P_{\boldsymbol{2}^{\boldsymbol{\ell}+\boldsymbol{j}},a}\delta_{\boldsymbol{j}+\boldsymbol{\ell}}[f](\boldsymbol{x})\prod_{i\in e(\boldsymbol{j})}\min\{2^{2\ell_i},1\}\max\{2^{\ell_i a},1\}. \tag{4.3.19}$$

Inserting this into (4.3.14) we obtain

$$\|d_{\boldsymbol{j},\boldsymbol{k}}(f)|s^r_{p,\theta}f\| \lesssim \Big(\sum_{\boldsymbol{\ell}\in\mathbb{Z}^d}\Big\|\Big(\sum_{\boldsymbol{j}\in\mathbb{N}^d_{-1}} 2^{r|\boldsymbol{j}+\boldsymbol{\ell}|_1\theta}\Big|P_{\boldsymbol{2}^{\boldsymbol{\ell}+\boldsymbol{j}},a}\delta_{\boldsymbol{j}+\boldsymbol{\ell}}[f](\boldsymbol{x})\prod_{i=1}^d A_{\ell_i}\Big|^\theta\Big)^{\frac{1}{\theta}}\Big|L_p(\mathbb{R}^d)\|^u\Big)^{\frac{1}{u}} \tag{4.3.20}$$

where

$$A_n = \begin{cases} 2^{(2-r)n} & : n < 0 \\ 2^{(a-r)n} & : n \geq 0. \end{cases}$$

Note, $\delta_{\boldsymbol{j}+\boldsymbol{\ell}}[f]$ are bandlimited functions with frequencies in $[-A2^{j_1+\ell_1}, B2^{j_1+\ell_1}] \times \ldots \times [-A2^{j_d+\ell_d}, B2^{j_d+\ell_d}]$. We fix $a > 0$ such that

$$2 > r > a > \max\left\{\frac{1}{p}, \frac{1}{\theta}\right\}. \tag{4.3.21}$$

Under these conditions we are allowed to apply Theorem B.17 that yields

$$\begin{aligned} &\Big(\sum_{\boldsymbol{\ell}\in\mathbb{Z}^d} \Big\| \Big(\sum_{\boldsymbol{j}\in\mathbb{N}^d_{-1}} 2^{r|\boldsymbol{j}+\boldsymbol{\ell}|_1\theta} \Big| P_{\boldsymbol{2}^{\boldsymbol{\ell}+\boldsymbol{j}},a}\delta_{\boldsymbol{j}+\boldsymbol{\ell}}[f] \prod_{i=1}^d A_{\ell_i}\Big|^\theta\Big)^{\frac{1}{\theta}} \Big| L_p(\mathbb{R}^d)\Big\|^u\Big)^{\frac{1}{u}} \\ \lesssim\ &\Big(\sum_{\boldsymbol{\ell}\in\mathbb{Z}^d} \Big(\prod_{i=1}^d A_{\ell_i}\Big)^u \Big\| \Big(\sum_{\boldsymbol{j}\in\mathbb{N}^d_{-1}} 2^{r|\boldsymbol{j}+\boldsymbol{\ell}|_1\theta} \Big|\delta_{\boldsymbol{j}+\boldsymbol{\ell}}[f]\Big|^\theta\Big)^{\frac{1}{\theta}} \Big| L_p(\mathbb{R}^d)\Big\|^u\Big)^{\frac{1}{u}}. \end{aligned}$$

Furthermore the choice of a in (4.3.21) together with the assumptions on r implies that there is a $\delta > 0$ such that $A_n \leq 2^{-\delta|n|}$ and hence

$$\begin{aligned} &\Big(\sum_{\boldsymbol{\ell}\in\mathbb{Z}^d} \Big(\prod_{i=1}^d A_{\ell_i}\Big)^u \Big\| \Big(\sum_{\boldsymbol{j}\in\mathbb{N}^d_{-1}} 2^{r|\boldsymbol{j}+\boldsymbol{\ell}|_1\theta} \Big|\delta_{\boldsymbol{j}+\boldsymbol{\ell}}[f]\Big|^\theta\Big)^{\frac{1}{\theta}} \Big| L_p(\mathbb{R}^d)\|^u\Big)^{\frac{1}{u}} \\ \lesssim\ &\Big(\sum_{\boldsymbol{\ell}\in\mathbb{Z}^d} 2^{-u\delta|\boldsymbol{\ell}|_1}\Big)^{\frac{1}{u}} \|f|S^r_{p,\theta}F(\mathbb{R}^d)\| \\ \lesssim\ &\|f|S^r_{p,\theta}F(\mathbb{R}^d)\| \end{aligned} \tag{4.3.22}$$

holds. In B-case inserting the estimate from (4.3.19) gives

$$\begin{aligned} \|d_{\boldsymbol{j},\boldsymbol{k}}(f)|s^r_{p,\theta}b\| &\leq \Big(\sum_{\boldsymbol{\ell}\in\mathbb{Z}^d} \Big(\sum_{\boldsymbol{j}\in\mathbb{N}^d_{-1}} 2^{r|\boldsymbol{j}|_1\theta} \Big\| \sum_{\boldsymbol{k}\in\mathbb{Z}^d} d_{\boldsymbol{j},\boldsymbol{k}}(\delta_{\boldsymbol{j}+\boldsymbol{\ell}}[f])\chi_{\boldsymbol{j},\boldsymbol{k}} \Big| L_p(\mathbb{R}^d)\Big\|^\theta\Big)^{\frac{u}{\theta}}\Big)^{\frac{1}{u}} \\ &\lesssim \Big(\sum_{\boldsymbol{\ell}\in\mathbb{Z}^d} \Big(\sum_{\boldsymbol{j}\in\mathbb{N}^d_{-1}} 2^{r|\boldsymbol{j}+\boldsymbol{\ell}|_1\theta} \Big\| P_{\boldsymbol{2}^{\boldsymbol{\ell}+\boldsymbol{j}},a}\delta_{\boldsymbol{j}+\boldsymbol{\ell}}[f] \prod_{i=1}^d A_{\ell_i} \Big| L_p(\mathbb{R}^d)\Big\|^\theta\Big)_p^{\frac{u}{\theta}}\Big)^{\frac{1}{u}}. \end{aligned}$$

Applying Theorem B.11 yields

$$\|d_{\boldsymbol{j},\boldsymbol{k}}(f)|s^r_{p,\theta}b\| \lesssim \Big(\sum_{\boldsymbol{\ell}\in\mathbb{Z}^d} \Big(\prod_{i=1}^d A_{\ell_i}\Big)^u \Big(\sum_{\boldsymbol{j}\in\mathbb{N}^d_{-1}} 2^{r|\boldsymbol{j}+\boldsymbol{\ell}|_1\theta} \|\delta_{\boldsymbol{j}+\boldsymbol{\ell}}[f]\|_p^\theta\Big)_p^{\frac{u}{\theta}}\Big)^{\frac{1}{u}}.$$

Finally the calculations in (4.3.22) show

$$\|d_{\boldsymbol{j},\boldsymbol{k}}(f)|s^r_{p,\theta}b\| \lesssim \|f|S^r_{p,\theta}B(\mathbb{R}^d)\|.$$

That proves the claim. □

Theorem 4.20. *(i) Let* $\frac{1}{2} < p, \theta < \infty$ *(*$\theta = \infty$*) and* $\max\{\frac{1}{p}, \frac{1}{\theta}\} < r < 1 + \min\{\frac{1}{p}, \frac{1}{\theta}\}$*. Then* $f \in S^r_{p,\theta}F(\mathbb{R}^d)$ *can be represented by*

$$f = \sum_{\boldsymbol{j} \in \mathbb{N}^d_{-1}} \sum_{\boldsymbol{k} \in \mathbb{Z}^d} d_{\boldsymbol{j},\boldsymbol{k}}(f) v_{\boldsymbol{j},\boldsymbol{k}}$$

with unconditional convergence in every $S^{r-\varepsilon}_{p,\theta}F(\mathbb{R}^d)$ *with* $\varepsilon > 0$*. Additionally, if* $\theta < \infty$ *the unconditional convergence holds in the space* $S^r_{p,\theta}F(\mathbb{R}^d)$*, itself. The following norms are equivalent*

$$\|d_{\boldsymbol{j},\boldsymbol{k}}(f)|s^r_{p,\theta}f\| \asymp \|f|S^r_{p,\theta}F(\mathbb{R}^d)\|. \tag{4.3.23}$$

(ii) Let $0 < p, \theta \leq \infty$ *(*$p > \frac{1}{2}$*) and* $\frac{1}{p} < r < 1 + \frac{1}{p}$*. Then* $f \in S^r_{p,\theta}B(\mathbb{R}^d)$ *can be represented by*

$$f = \sum_{\boldsymbol{j} \in \mathbb{N}^d_{-1}} \sum_{\boldsymbol{k} \in \mathbb{Z}^d} d_{\boldsymbol{j},\boldsymbol{k}}(f) v_{\boldsymbol{j},\boldsymbol{k}}$$

with unconditional convergence in every $S^{r-\varepsilon}_{p,\theta}B(\mathbb{R}^d)$ *with* $\varepsilon > 0$*. Additionally, if* $\theta < \infty$ *the unconditional convergence holds in the space* $S^r_{p,\theta}B(\mathbb{R}^d)$*, itself. The following norms are equivalent*

$$\|d_{\boldsymbol{j},\boldsymbol{k}}(f)|s^r_{p,\theta}b\| \asymp \|f|S^r_{p,\theta}B(\mathbb{R}^d)\|. \tag{4.3.24}$$

Proof. We prove the F-case here. The B-case can be obtained by replacing Theorem 4.16 by Theorem 4.18 and the usual modifications. We restrict to the case $\theta < \infty$, the modifications in case $\theta = \infty$ are analogous to the proof of Theorem 4.16. Theorem 4.19 implies that for $f \in S^r_{p,\theta}F(\mathbb{R}^d)$ the sequence $(d_{\boldsymbol{j},\boldsymbol{k}})_{\boldsymbol{j},\boldsymbol{k}}$ is in $s^r_{p,\theta}f$. Theorem 4.16 yields that

$$\sum_{\boldsymbol{j} \in \mathbb{N}^d_{-1}} \sum_{\boldsymbol{k} \in \mathbb{Z}^d} d_{\boldsymbol{j},\boldsymbol{k}} v_{\boldsymbol{j},\boldsymbol{k}} \tag{4.3.25}$$

converges unconditionally in $S^r_{p,\theta}F(\mathbb{R}^d)$ to some element $g \in S^r_{p,\theta}F(\mathbb{R}^d)$. It remains to show $f(x) = g(x)$ for all $x \in \mathbb{R}^d$. Let $K \subset \mathbb{R}^d$ (K compact). We use the order of summation provided by F^d_n (cf. Definition 4.1.4). Let $\varepsilon > 0$. Then

$$\begin{aligned} \|f - g|C(K)\| &\leq \left\|f - F^d_n f\middle|C(K)\right\| + \left\|F^d_n f - g\middle|C(K)\right\| \\ &\leq \left\|f - F^d_n f\middle|C(K)\right\| + \left\|F^d_n f - g\middle|S^r_{p,\theta}F(\mathbb{R}^d)\right\| \\ &< 2\varepsilon. \end{aligned}$$

This follows immediately by Theorem 4.8 and the unconditional convergence of the series (4.3.25) in $S^r_{p,\theta}F(\mathbb{R}^d)$. The equivalence of the norms follows by Theorem 4.16 and Theorem 4.19. □

Remark 4.21. *We should remark some facts about the sharpness of the smoothness restrictions in the theorem above. Theoretically, for* $r > \frac{1}{p}$ *one deals with continuous functions, this is required to give a sense to function evaluations in* $S^r_{p,\theta}F(\mathbb{R}^d)$ *and* $S^r_{p,\theta}B(\mathbb{R}^d)$*. The upper bound* $1 + \frac{1}{p}$ *in B-case seems also to be sharp, since* $v_{\boldsymbol{j},\boldsymbol{k}} \notin S^r_{p,\theta}B(\mathbb{R}^d)$ *for* $r \geq 1 + \frac{1}{p}$, $\theta < \infty$*. The F-case becomes more exciting in case* $0 < \theta < p$*. Recently, Seeger and Ullrich proved in [100, Rem. 7.3], [101] that the close related Haar system is not an unconditional basis in* $W^r_p(\mathbb{R}) = F^r_{p,2}(\mathbb{R})$ *in case* $\frac{1}{p} - 1 < r < \frac{1}{\theta} - 1$*. The Faber-Schauder hat function* $v(x)$ *can be written as the integral of the Haar function* $h(x)$*. We have the identity*

$$v(x) = \int_0^x h(t)dt$$

where

$$h(t) = \begin{cases} 1 & , \quad 0 \leq t \leq \frac{1}{2} \\ -1 & , \quad \frac{1}{2} < t \leq 1. \end{cases}$$

Hence, the properties of the Faber-Schauder representation can be interpreted in some sense as a 1*-lifted Haar basis representation. This indicates that the condition* $r > \max\{\frac{1}{p}, \frac{1}{\theta}\}$ *is also sharp.*

4.4 Equivalent characterizations on the unit cube

In this section we study the Faber-Schauder system as an unconditional basis for $S^r_{p,\theta}F([0,1]^d)$ and $S^r_{p,\theta}B([0,1]^d)$. In section 3.4 we define domains as open connected sets. This is required for instance since our spaces $S^r_{p,\theta}B(\Omega)$ and $S^r_{p,\theta}F(\Omega)$ are based on distributions $f \in D'(\Omega)$ which require an open set to be well defined. For that reason we formally deal with the open unit cube $\Omega := (0,1)^d$ and consider the index set

$$\nabla := \{(\boldsymbol{j}, \boldsymbol{k}) \in \mathbb{N}^d_{-1} \times \mathbb{Z}^d : \operatorname{supp} v_{\boldsymbol{j},\boldsymbol{k}} \cap (0,1)^d \neq \emptyset\}. \tag{4.4.1}$$

For fixed level $\boldsymbol{j}$ all translates with the property $\operatorname{supp} v_{\boldsymbol{j},\boldsymbol{k}} \cap (0,1)^d$ are contained in

$$D_{\boldsymbol{j}} := \bigtimes_{i=1}^{d} D_{j_i}, \tag{4.4.2}$$

which is defined as the tensor product of the sets

$$D_j := \begin{cases} \{k \in \mathbb{N}_0 : \ 0 \leq k < 2^j\} & , \quad j \geq 0 \\ \{0, 1\} & , \quad j = -1. \end{cases}$$

This allows us to define the following sequence spaces.

Definition 4.22. *We define for* $0 < p, \theta \leq \infty$ *(*$s^{r,\Omega}_{p,\theta}f$*:* $p < \infty$*) the spaces* $s^{r,\Omega}_{p,\theta}f$ *and* $s^{r,\Omega}_{p,\theta}b$ *as the space of all sequences of coefficients* $(\lambda_{\boldsymbol{j},\boldsymbol{k}})_{\boldsymbol{j}\in\mathbb{N}^d_{-1}, \boldsymbol{k}\in D_{\boldsymbol{j}}} \subset \mathbb{C}$ *with finite* *(quasi)-norm*

$$\|\lambda_{\boldsymbol{j},\boldsymbol{k}}|s^{r,\Omega}_{p,\theta}f\| := \begin{cases} \left\| \left(\sum_{\boldsymbol{j}\in\mathbb{N}_{-1}{}^d} 2^{\theta|\boldsymbol{j}|_1 r} |\sum_{\boldsymbol{k}\in D_{\boldsymbol{j}}} \lambda_{\boldsymbol{j},\boldsymbol{k}}\chi_{\boldsymbol{j},\boldsymbol{k}}|^\theta \right)^{\frac{1}{\theta}} \middle| L_p(\Omega) \right\| & : \quad 0<\theta<\infty, \\ \left\| \sup_{\boldsymbol{j}\in\mathbb{N}^d_{-1}} 2^{|\boldsymbol{j}|_1 r} |\sum_{\boldsymbol{k}\in D_{\boldsymbol{j}}} \lambda_{\boldsymbol{j},\boldsymbol{k}}\chi_{\boldsymbol{j},\boldsymbol{k}}| \middle| L_p(\Omega) \right\| & : \quad \theta=\infty, \end{cases}$$

and

$$\|\lambda_{\boldsymbol{j},\boldsymbol{k}}|s^{r,\Omega}_{p,\theta}b\| := \begin{cases} \left(\sum_{\boldsymbol{j}\in\mathbb{N}^d_{-1}} 2^{\theta|\boldsymbol{j}|_1 r} \| \sum_{\boldsymbol{k}\in D_{\boldsymbol{j}}} \lambda_{\boldsymbol{j},\boldsymbol{k}}\chi_{\boldsymbol{j},\boldsymbol{k}}|L_p(\Omega)\|^\theta \right)^{\frac{1}{\theta}} & : \quad 0<\theta<\infty, \\ \sup_{\boldsymbol{j}\in\mathbb{N}^d_{-1}} 2^{|\boldsymbol{j}|_1 r} \| \sum_{\boldsymbol{k}\in D_{\boldsymbol{j}}} \lambda_{\boldsymbol{j},\boldsymbol{k}}\chi_{\boldsymbol{j},\boldsymbol{k}}|L_p(\Omega)\| & : \quad \theta=\infty, \end{cases}$$

respectively.

Lemma 4.23. *Let $0 < q < p \le \infty$ (f-case: $p < \infty$) and $0 < \theta \le \infty$. Then*

$$s^{r,\Omega}_{p,\theta}f \hookrightarrow s^{r,\Omega}_{q,\theta}f$$

and

$$s^{r,\Omega}_{p,\theta}b \hookrightarrow s^{r,\Omega}_{q,\theta}b$$

Proof. The proof follows trivially by definition using the embedding

$$L_p((0,1)^d) \hookrightarrow L_q((0,1)^d).$$

□

Remark 4.24. *According to the definition of the spaces in Section 3.4 the functions f (distributions) in $S^r_{p,\theta}X(\Omega)$, $X \in \{B,F\}$ can be extended to functions f^* belonging to $S^r_{p,\theta}X(\mathbb{R}^d)$. For $r > \frac{1}{p}$ this space is continuously embedded into $C(\mathbb{R}^d)$ (cf. Lemma 3.4,(ii)). In fact, this implies that there is an unique extension from f on $(0,1)^d$ to $[0,1]^d$ giving us an continuous function on $[0,1]^d$. Since the norms are based on L_p expressions that do not care about boundary values we denote (identify) the space $S^r_{p,\theta}X(\Omega)$ with $S^r_{p,\theta}X([0,1]^d)$ in case $r > \frac{1}{p}$.*

Theorem 4.25. *(i) Let $\frac{1}{2} < p,\theta \le \infty$ ($p < \infty$) and $\max\{\frac{1}{p},\frac{1}{\theta}\} < r < 1+\min\{\frac{1}{p},\frac{1}{\theta}\}$. Then $f \in S^r_{p,\theta}F([0,1]^d)$ can be represented by*

$$f = \sum_{\boldsymbol{j}\in\mathbb{N}^d_{-1}} \sum_{\boldsymbol{k}\in D_{\boldsymbol{j}}} d_{\boldsymbol{j},\boldsymbol{k}}(f) v_{\boldsymbol{j},\boldsymbol{k}} \tag{4.4.3}$$

with unconditional convergence in every $S^{r-\varepsilon}_{p,\theta}F([0,1]^d)$ with $\varepsilon > 0$. Additionally, if $\theta < \infty$ the unconditional convergence holds in the space $S^r_{p,\theta}F([0,1]^d)$, itself. The following norms are equivalent

$$\|d_{\boldsymbol{j},\boldsymbol{k}}(f)|s^{r,\Omega}_{p,\theta}f\| \asymp \|f|S^r_{p,\theta}F([0,1]^d)\|. \tag{4.4.4}$$

(ii) Let $0 < p,\theta \le \infty$ *(*$p > \frac{1}{2}$*) and* $\frac{1}{p} < r < 1 + \frac{1}{p}$*. Then* $f \in S^r_{p,\theta}B([0,1]^d)$ *can be represented by*

$$f = \sum_{\boldsymbol{j}\in\mathbb{N}^d_{-1}} \sum_{\boldsymbol{k}\in D_{\boldsymbol{j}}} d_{\boldsymbol{j},\boldsymbol{k}}(f) v_{\boldsymbol{j},\boldsymbol{k}}$$

with unconditional convergence in every $S^{r-\varepsilon}_{p,\theta}B([0,1]^d)$ *with* $\varepsilon > 0$*. Additionally, if* $\max\{p,\theta\} < \infty$ *the unconditional convergence holds in the space* $S^r_{p,\theta}B([0,1]^d)$*, itself. The following norms are equivalent*

$$\|d_{\boldsymbol{j},\boldsymbol{k}}(f)|s^{r,\Omega}_{p,\theta}b\| \asymp \|f|S^r_{p,\theta}B([0,1]^d)\|. \tag{4.4.5}$$

Proof. As usual, we prove only the F-case. The B-case works with obvious modifications and was considered in [120] and [125]. Let $f \in S^r_{p,\theta}F([0,1]^d)$. By definition we find a $g^* \in S^r_{p,\theta}F(\mathbb{R}^d)$ with $g^*|_\Omega = f$ and

$$\|f|S^r_{p,\theta}F([0,1]^d)\| \le \|g^*|S^r_{p,\theta}F(\mathbb{R}^d)\| \le 2\|f|S^r_{p,\theta}F([0,1]^d)\|.$$

Expanding g^* into the Faber-Schauder system restricted to Ω, then

$$g := \sum_{\boldsymbol{j}\in\mathbb{N}^d_{-1}} \sum_{\boldsymbol{k}\in D_{\boldsymbol{j}}} d_{\boldsymbol{j},\boldsymbol{k}}(g^*) v_{\boldsymbol{j},\boldsymbol{k}} \in S^r_{p,\theta}F(\mathbb{R}^d)$$

is also a function with the property $g|_\Omega = f$. Theorem 4.20 yields

$$\begin{aligned}\Big\|\Big(\sum_{\boldsymbol{j}\in\mathbb{N}^d_{-1}} 2^{\theta r|\boldsymbol{j}|_1}\Big|\sum_{\boldsymbol{k}\in D_{\boldsymbol{j}}} d_{\boldsymbol{j},\boldsymbol{k}}(g^*)\chi_{\boldsymbol{j},\boldsymbol{k}}\Big|^\theta\Big)^{\frac{1}{\theta}}\Big|L_p(\mathbb{R}^d)\Big\| &\asymp \|g|S^r_{p,\theta}F(\mathbb{R}^d)\|\\ \asymp \|g^*|S^r_{p,\theta}F(\mathbb{R}^d)\| &\asymp \|f|S^r_{p,\theta}F([0,1]^d)\|.\end{aligned}$$

It remains to show

$$\Big\|\Big(\sum_{\boldsymbol{j}\in\mathbb{N}^d_{-1}} 2^{\theta r|\boldsymbol{j}|_1}\Big|\sum_{\boldsymbol{k}\in D_{\boldsymbol{j}}} d_{\boldsymbol{j},\boldsymbol{k}}(g^*)\chi_{\boldsymbol{j},\boldsymbol{k}}\Big|^\theta\Big)^{\frac{1}{\theta}}\Big|L_p(\mathbb{R}^d)\Big\| \asymp \|d_{\boldsymbol{j},\boldsymbol{k}}(f)|s^{r,\Omega}_{p,\theta}f\|. \tag{4.4.6}$$

Due to the definition of $d_{\boldsymbol{j},\boldsymbol{k}}(g^*)$ we obtain that for $(\boldsymbol{j},\boldsymbol{k}) \in \nabla$ only function values of g^* in Ω are considered. Since

$$f(x) = g^*(x),$$

for all $\boldsymbol{x} \in \Omega$ we have the identity

$$d_{\boldsymbol{j},\boldsymbol{k}}(g^*) = d_{\boldsymbol{j},\boldsymbol{k}}(f).$$

Finally we show that it suffices for an equivalent norm to integrate over Ω instead of $\mathbb{R}^d$. The direction "$\gtrsim$" in (4.4.6) is obvious. The core of the matter for the direction "$\lesssim$" is also very easy to see. To prevent further notation we prove only the case $d = 1$. We recognize $\operatorname{supp} g \subset [-1,2]$. That yields

$$\begin{aligned}&\Big\|\Big(\sum_{j\in\mathbb{N}_{-1}} 2^{\theta r|j|}\Big|\sum_{k\in D_j} d_{j,k}(f)\chi_{j,k}\Big|^\theta\Big)^{\frac{1}{\theta}}\Big|L_p(\mathbb{R})\Big\|\\ = \;&\Big\|\Big(\sum_{j\in\mathbb{N}_{-1}} 2^{\theta r|j|}\Big|\sum_{k\in D_j} d_{j,k}(f)\chi_{j,k}\Big|^\theta\Big)^{\frac{1}{\theta}}\Big|L_p\Big(\Big[-\frac{1}{2},\frac{3}{2}\Big]\Big)\Big\|.\end{aligned}$$

Splitting the integral in three parts gives

$$\Big\|\Big(\sum_{j\in\mathbb{N}_{-1}} 2^{\theta r|j|}\Big|\sum_{k\in D_j} d_{j,k}(f)\chi_{j,k}\Big|^{\theta}\Big)^{\frac{1}{\theta}}\Big|L_p\Big(\Big[-\frac{1}{2},\frac{3}{2}\Big]\Big)\Big\| \lesssim \|d_{j,k}(f)|s^{r,\Omega}_{p,\theta}f\|$$

$$+\Big\|\Big(\sum_{j\in\mathbb{N}_{-1}} 2^{\theta r|j|}\Big|\sum_{k\in D_j} d_{j,k}(f)\chi_{j,k}\Big|^{\theta}\Big)^{\frac{1}{\theta}}\Big|L_p\Big(\Big[-\frac{1}{2},0\Big]\Big)\Big\|$$

$$+\Big\|\Big(\sum_{j\in\mathbb{N}_{-1}} 2^{\theta r|j|}\Big|\sum_{k\in D_j} d_{j,k}(f)\chi_{j,k}\Big|^{\theta}\Big)^{\frac{1}{\theta}}\Big|L_p\Big(\Big[1,\frac{3}{2}\Big]\Big)\Big\|.$$

The term integrating over $\Big[-\frac{1}{2},0\Big]$ breaks down to

$$\Big\|\Big(\sum_{j\in\mathbb{N}_{-1}} 2^{\theta r|j|}\Big|\sum_{k\in D_j} d_{j,k}(f)\chi_{j,k}\Big|^{\theta}\Big)^{\frac{1}{\theta}}\Big|L_p\Big(\Big[-\frac{1}{2},0\Big]\Big)\Big\| = C_r\Big\|d_{-1,0}(f)\chi_{-1,0}\Big|L_p\Big(\Big[-\frac{1}{2},0\Big]\Big)\Big\|.$$

Symmetry of $\chi_{-1,0}(x)$ yields

$$\Big\|d_{-1,0}(f)\chi_{-1,0}\Big|L_p\Big(\Big[-\frac{1}{2},0\Big]\Big)\Big\| = \|d_{-1,0}(f)\chi_{-1,0}|L_p([0,1])\| \lesssim \|d_{j,k}(f)|s^{r,\Omega}_{p,\theta}f\|.$$

With the same arguments the term integrating over $\Big[1,\frac{3}{2}\Big]$ can be estimated. Sure, similar arguments can be applied in case $d>1$. Here the decomposition of $\Big[-\frac{1}{2},\frac{3}{2}\Big]^d$ causes further technical consideration. That finishes the proof. □

Theorem 4.26. *(i) Let $\frac{1}{2}<p,\theta\le\infty$ ($p<\infty$) and $\max\{\frac{1}{p},\frac{1}{\theta}\}<r<2$. Then for $f\in S^r_{p,\theta}F([0,1]^d)$ the inequality*

$$\|d_{\boldsymbol{j},\boldsymbol{k}}(f)|s^{r,\Omega}_{p,\theta}f\| \lesssim \|f|S^r_{p,\theta}F([0,1]^d)\|$$

holds.

(ii) Let $0<p,\theta\le\infty$ ($p>\frac{1}{2}$) and $\frac{1}{p}<r<2$. Then for $f\in S^r_{p,\theta}B([0,1]^d)$ the inequality

$$\|d_{\boldsymbol{j},\boldsymbol{k}}(f)|s^{r,\Omega}_{p,\theta}b\| \lesssim \|f|S^r_{p,\theta}B([0,1]^d)\|$$

holds.

Proof. We prove only the F-case, the B-case works with well known modifications. Let $g\in S^r_{p,\theta}F(\mathbb{R}^d)$ be an arbitrary extension of f, i.e. $g|_\Omega=f$. Then Theorem 4.19 yields

$$\begin{aligned}\|d_{\boldsymbol{j},\boldsymbol{k}}(f)|s^{r,\Omega}_{p,\theta}f\| &= \Big\|\Big(\sum_{\boldsymbol{j}\in\mathbb{N}^d_{-1}} 2^{|\boldsymbol{j}|_1 r\theta}\Big|\sum_{\boldsymbol{k}\in D_{\boldsymbol{j}}} d_{\boldsymbol{j},\boldsymbol{k}}(f)\chi_{\boldsymbol{j},\boldsymbol{k}}\Big|^{\theta}\Big)^{\frac{1}{\theta}}\Big|L_p([0,1]^d)\Big\| \\ &\le \Big\|\Big(\sum_{\boldsymbol{j}\in\mathbb{N}^d_{-1}} 2^{|\boldsymbol{j}|_1 r\theta}\Big|\sum_{\boldsymbol{k}\in\mathbb{Z}^d} d_{\boldsymbol{j},\boldsymbol{k}}(g)\chi_{\boldsymbol{j},\boldsymbol{k}}\Big|^{\theta}\Big)^{\frac{1}{\theta}}\Big|L_p(\mathbb{R}^d)\Big\| \\ &\asymp \|g|S^r_{p,\theta}F(\mathbb{R}^d)\|.\end{aligned}$$

Finally, taking the infimum over all $\|g|S^r_{p,\theta}F(\mathbb{R}^d)\|$ with $g|_\Omega = f$ yields

$$\|d_{\boldsymbol{j},\boldsymbol{k}}(f)|s^{r,\Omega}_{p,\theta}f\| \lesssim \|f|S^r_{p,\theta}F([0,1]^d)\|.$$

That finishes the proof. □

Lemma 4.27. *Let* $0 < p < \infty$ *and* $\lambda = (\lambda_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}})_{\boldsymbol{j}\in\mathbb{N}^d_{-1},\boldsymbol{k}\in D_{\boldsymbol{j}}} \in s^{0,\Omega}_{p,1}f$. *Then there is a* $C > 0$ *such that*

$$\Big\| \sum_{\boldsymbol{j}\in\mathbb{N}^d_{-1}} \sum_{\boldsymbol{k}\in D_{\boldsymbol{j}}} \lambda_{\boldsymbol{j},\boldsymbol{k}} v_{\boldsymbol{j},\boldsymbol{k}} \Big| L_p([0,1]^d)\Big\| \le C\|\lambda|s^{0,\Omega}_{p,1}f\|$$

holds.

Proof. The non-trivial point are the levels $j_i = -1$, since the supports of $v_{-1,0}$ and $v_{-1,1}$ have some overlap. Additionally the support of the corresponding characteristic function $\chi_{j,k}$ is only half of the size of the support of $v_{j,k}$. We introduce the auxiliary characteristic function

$$\chi^*_{-1,k} := \begin{cases} \chi_{-1,0} & : \quad k = 1, \\ \chi_{-1,1} & : \quad k = 0. \end{cases}$$

and use the following simple decomposition estimate for $x \in [0,1]$

$$|v_{-1,k}| \le \chi_{[0,1]} = \chi_{-1,k} + \chi^*_{-1,k}.$$

To prevent further notation we estimate only the case $d = 2$. The modifications for $d > 2$ are obvious.

$$\begin{aligned}
&\Big\| \sum_{\boldsymbol{j}\in\mathbb{N}^2_{-1}} \sum_{\boldsymbol{k}\in D_{\boldsymbol{j}}} \lambda_{\boldsymbol{j},\boldsymbol{k}} v_{\boldsymbol{j},\boldsymbol{k}} \Big| L_p([0,1]^d)\Big\|^p \\
\le\ &\Big\| \sum_{\boldsymbol{j}\in\mathbb{N}^2_{-1}} \sum_{\boldsymbol{k}\in D_{\boldsymbol{j}}} |\lambda_{\boldsymbol{j},\boldsymbol{k}} v_{\boldsymbol{j},\boldsymbol{k}}| \Big| L_p([0,1]^2)\Big\|^p \\
\lesssim\ &\Big\| \sum_{\boldsymbol{j}\in\mathbb{N}^2_{-1}} \Big| \sum_{\boldsymbol{k}\in D_{\boldsymbol{j}}} \lambda_{\boldsymbol{j},\boldsymbol{k}} \chi_{\boldsymbol{j},\boldsymbol{k}} \Big| \Big| L_p([0,1]^2)\Big\|^p \\
&+ \Big\| \sum_{j_2\in\mathbb{N}_0} \sum_{\boldsymbol{k}\in D_{(-1,j_2)}} |\lambda_{(-1,j_2),(k_1,k_2)} \chi^*_{-1,k_1} \chi_{j_2,k_2}| \Big| L_p([0,1]^2)\Big\|^p \\
&+ \Big\| \sum_{j_1\in\mathbb{N}_0} \sum_{\boldsymbol{k}\in D_{(j_1,-1)}} |\lambda_{(j_1,-1),(k_1,k_2)} \chi_{j_1,k_1} \chi^*_{-1,k_2}| \Big| L_p([0,1]^2)\Big\|^p \\
&+ \Big\| \sum_{\boldsymbol{k}\in D_{(-1,-1)}} |\lambda_{(-1,-1),(k_1,k_2)} \chi^*_{-1,k_1} \chi^*_{-1,k_2}| \Big| L_p([0,1]^2)\Big\|^p.
\end{aligned}$$

The term

$$\sum_{j_2\in\mathbb{N}_0} \sum_{\boldsymbol{k}\in D_{(-1,j_2)}} |\lambda_{(-1,j_2),(k_1,k_2)} \chi^*_{-1,k_2}(x_1) \chi_{j_2,k_2}(x_2)|$$

is for fixed $x_2 \in [0,1]$ a step function. Interchanging the interval of the integration we obtain

$$\begin{aligned}
&\Big\| \sum_{j_2 \in \mathbb{N}_0} \sum_{\boldsymbol{k} \in D_{(-1,j_2)}} |\lambda_{(-1,j_2),(k_1,k_2)} \chi^*_{-1,k_2} \chi_{j_2,k_2}| \Big| L_p([0,1]^2) \Big\| \\
&= \Big\| \sum_{j_2 \in \mathbb{N}_0} \sum_{\boldsymbol{k} \in D_{(-1,j_2)}} |\lambda_{(-1,j_2),(k_1,k_2)} \chi_{-1,k_2} \chi_{j_2,k_2}| \Big| L_p([0,1]^2) \Big\| \\
\leq &\ \Big\| \sum_{\boldsymbol{j} \in \mathbb{N}^2_{-1}} \Big| \sum_{\boldsymbol{k} \in D_{\boldsymbol{j}}} \lambda_{\boldsymbol{j},\boldsymbol{k}} \chi_{\boldsymbol{j},\boldsymbol{k}} \Big| \Big| L_p([0,1]^2) \Big\|.
\end{aligned}$$

The remaining terms can be estimated similarly. □

With similar arguments we obtain the following b-space counterpart.

Lemma 4.28. *Let $0 < p \leq \infty$ and $\lambda = (\lambda_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}})_{\boldsymbol{j} \in \mathbb{N}^d_{-1}, \boldsymbol{k} \in D_{\boldsymbol{j}}} \in s^{0,\Omega}_{p,1}b$. Then there is a $C > 0$ such that*

$$\Big\| \sum_{\boldsymbol{j} \in \mathbb{N}^d_{-1}} \sum_{\boldsymbol{k} \in D_{\boldsymbol{j}}} \lambda_{\boldsymbol{j},\boldsymbol{k}} v_{\boldsymbol{j},\boldsymbol{k}} \Big| L_p([0,1]^d) \Big\| \leq C \|\lambda | s^{0,\Omega}_{p,1} b\|$$

holds.

Proof. The proof follows by a simplification of the arguments presented in the proof of Lemma 4.27. □

4.5 Bounded second order weak derivatives

In this section we follow the idea of Bungartz, Griebel [8] and consider functions with bounded second order mixed weak derivatives in connection with hat functions. As a preparation we need the following lemma.

Lemma 4.29. *Let $1 \leq p < \infty$. Then there exists a constant $C > 0$ such that*

$$\Big(\sum_{k=-\infty}^{\infty} |f(k)|^p \Big)^{\frac{1}{p}} \leq C \|f | B^{\frac{1}{p}}_{p,1}(\mathbb{R})\|$$

holds for all $f \in B^{\frac{1}{p}}_{p,1}(\mathbb{R})$.

Proof. We refer to [98, Proposition 2]. □

Now we are able to prove the following theorem as an analog of Theorem 4.19.

Theorem 4.30. *For every function $f \in S^2_p W(\mathbb{R}^d)$ it holds*

$$\sup_{\boldsymbol{j} \in \mathbb{N}^d_{-1}} 2^{2|\boldsymbol{j}|_1} \Big\| \sum_{\boldsymbol{k} \in \mathbb{Z}^d} d_{\boldsymbol{j},\boldsymbol{k}}(f) \chi_{\boldsymbol{j},\boldsymbol{k}} \Big| L_p(\mathbb{R}^d) \Big\| \lesssim \|f | S^2_p W(\mathbb{R}^d)\|. \tag{4.5.1}$$

Proof. For technical reasons we state the proof for $d = 2$ here. The methods we use can be easily extended to dimensions $d > 2$. *Step 1.* Assume we have proven the above inequality for functions from $D(\mathbb{R}^2)$. Let us denote the norm on the left-hand side with

$$\|f|\tilde{S}^2_{p,\infty}B(\mathbb{R}^2)\| := \sup_{\boldsymbol{j}\in\mathbb{N}^d_{-1}} 2^{2|\boldsymbol{j}|_1}\Big\|\sum_{\boldsymbol{k}\in D_{\boldsymbol{j}}} d_{\boldsymbol{j},\boldsymbol{k}}(f)\chi_{\boldsymbol{j},\boldsymbol{k}}\Big|L_p(\mathbb{R}^2)\Big\|.$$

According to (3.1.4) the norm on the right hand side in (4.5.1) is equivalent to

$$\|f|S^2_pH(\mathbb{R}^2)\| := \|f|L_p(\mathbb{R}^2)\| + \Big\|\frac{\partial^2}{\partial x_1^2}f\Big|L_p(\mathbb{R}^2)\Big\| + \Big\|\frac{\partial^2}{\partial x_2^2}f\Big|L_p(\mathbb{R}^2)\Big\| + \Big\|\frac{\partial^2}{\partial x_1\partial x_2}f\Big|L_p(\mathbb{R}^2)\Big\|. \tag{4.5.2}$$

Assume we have a sequence $(\varphi_j)_{j\in\mathbb{N}_0} \in D(\mathbb{R}^2)$ such that $\varphi_j \xrightarrow[j\to\infty]{} \in S^2_pW(\mathbb{R}^2)$ in the norm $\|\cdot|S^2_pW(\mathbb{R}^2)\|$. Then

$$\|\varphi_j - \varphi_i|\tilde{S}^2_{p,\infty}B(\mathbb{R}^2)\| \lesssim \|\varphi_j - \varphi_i|S^2_pW\| < \varepsilon, \quad \text{for all } i,j > M.$$

This implies that $(\varphi_j)_{\boldsymbol{j}}$ is a Cauchy sequence in $\tilde{S}^2_{p,\infty}B(\mathbb{R}^2)$ and hence convergent. This implies convergence in $L_p(\mathbb{R}^2)$ (and even $C(\mathbb{R}^2)$). Hence, we have

$$\varphi_j \xrightarrow[j\to\infty]{} f \in S^2_pW(\mathbb{R}^2) \text{ in } \|\cdot|S^2_pW(\mathbb{R}^2)\|$$

and

$$\varphi_j \xrightarrow[j\to\infty]{} f^* \in C(\mathbb{R}^2) \text{ in } \|\cdot|\tilde{S}^2_{p,\infty}B(\mathbb{R}^2)\|.$$

Therefore $f = f^*$. Further

$$\|\varphi_j|\tilde{S}^2_{p,\infty}B(\mathbb{R}^2)\| \lesssim \|\varphi_j|S^2_pW(\mathbb{R}^2)\|.$$

Taking the limit on both sides yields

$$\|f|\tilde{S}^2_{p,\infty}B(\mathbb{R}^2)\| \lesssim \|f|S^2_pW(\mathbb{R}^2)\|.$$

Hence, it remains to prove the theorem for functions from $D(\mathbb{R}^2)$. *Step 2.* Let $\varphi \in D(\mathbb{R}^2)$. We consider the left hand side and decompose it into the regions

$$\sup_{\boldsymbol{j}\in\mathbb{N}^2_{-1}} \ldots = \sup_{\substack{j_1\in\mathbb{N}_0\\ j_2\in\mathbb{N}_0}} \ldots + \sup_{\substack{j_1=-1\\ j_2\in\mathbb{N}_0}} \ldots + \sup_{\substack{j_1\in\mathbb{N}_0\\ j_2=-1}} \ldots + \sup_{\substack{j_1=-1\\ j_2=-1}} \ldots .$$

We show that the corresponding supremum are bounded by the $S^2_pW(\mathbb{R}^2)$-norm. We start with $j_1 = j_2 = -1$

$$\begin{aligned}
&\Big\|\sum_{k_1\in\mathbb{Z}}\sum_{k_2\in\mathbb{Z}} d_{\boldsymbol{j},\boldsymbol{k}}(\varphi)\chi_{\boldsymbol{j},\boldsymbol{k}}\Big|L_p(\mathbb{R}^2)\Big\| \\
=\ &\Big\|\sum_{k_1\in\mathbb{Z}}\sum_{k_2\in\mathbb{Z}} \varphi(k_1,k_2)\chi_{-1,k_1}(x_1)\chi_{-1,k_2}(x_2)\Big|L_p(\mathbb{R}^2)\Big\| \\
\lesssim\ &\Big(\sum_{k_1\in\mathbb{Z}}\sum_{k_2\in\mathbb{Z}} |\varphi(k_1,k_2)|^p\Big)^{\frac{1}{p}}.
\end{aligned}$$

Lemma 4.29 allows to estimate this by

$$\Big\| \sum_{k_1\in\mathbb{Z}} \sum_{k_2\in\mathbb{Z}} d_{\boldsymbol{j},\boldsymbol{k}}(\varphi)\chi_{\boldsymbol{j},\boldsymbol{k}} \Big| L_p(\mathbb{R}^2)\Big\| \ \lesssim \ \Big(\sum_{k_1\in\mathbb{Z}} \|\varphi(k_1,\cdot)|B^{\frac{1}{p}}_{p,1}(\mathbb{R})\|^p \Big)^{\frac{1}{p}}.$$

Applying Lemma 3.4 (remember: $S^r_{p,\theta}B(\mathbb{R}) = B^r_\theta(\mathbb{R})$) yields

$$\Big\| \sum_{k_1\in\mathbb{Z}} \sum_{k_2\in\mathbb{Z}} d_{\boldsymbol{j},\boldsymbol{k}}(\varphi)\chi_{\boldsymbol{j},\boldsymbol{k}} \Big| L_p(\mathbb{R}^2)\Big\| \ \lesssim \ \Big(\sum_{k_1\in\mathbb{Z}} \|\varphi(k_1,\cdot)|W^2_p(\mathbb{R})\|^p \Big)^{\frac{1}{p}}.$$

Repeating the last two steps for the first variable we obtain

$$\Big\| \sum_{k_1\in\mathbb{Z}} \sum_{k_2\in\mathbb{Z}} d_{\boldsymbol{j},\boldsymbol{k}}(\varphi)\chi_{\boldsymbol{j},\boldsymbol{k}} \Big| L_p(\mathbb{R}^2)\Big\| \ \lesssim \ \Big\| \|\varphi(x_1,x_2)|W^2_p(\mathbb{R})\| \Big| W^2_p(\mathbb{R})\Big\|.$$

Finally the cross norm property of $S^r_pW(\mathbb{R}^2)$ (cf. [107, Theorem 2.1]) or alternatively inserting the equivalent norm known from Lemma (3.14) immediately gives

$$\Big\| \sum_{k_1\in\mathbb{Z}^d} \sum_{k_2\in\mathbb{Z}^d} d_{\boldsymbol{j},\boldsymbol{k}}(\varphi)\chi_{\boldsymbol{j},\boldsymbol{k}} \Big| L_p(\mathbb{R}^2)\Big\| \lesssim \|\varphi|S^2_pW(\mathbb{R}^2)\|.$$

In case $j_1 = -1$, $j_2 \in \mathbb{N}_0$ we argue as follows. We use the property that $\chi_{\boldsymbol{j},\boldsymbol{k}}$ has disjoint supports for different $\boldsymbol{k}$ and obtain

$$\begin{aligned}
& \sup_{j_2\in\mathbb{N}_0} 2^{2j_2p} \Big\| \sum_{k_1\in\mathbb{Z}} \sum_{k_2\in\mathbb{Z}} d_{\boldsymbol{j},\boldsymbol{k}}(\varphi)\chi_{\boldsymbol{j},\boldsymbol{k}} \Big| L_p(\mathbb{R}^2)\Big\|^p \\
= \ & \sup_{j_2\in\mathbb{N}_0} 2^{2j_2p} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \Big| \sum_{k_1\in\mathbb{Z}} \chi_{-1,k_1}(x_1) \sum_{k_2\in\mathbb{Z}} \chi_{j_2,k_2}(x_2) \Delta^2_{2^{-(j_2+1)},2}\varphi(k_1,2^{-j_2})k_2\Big|^p dx_1dx_2 \\
\lesssim \ & \sup_{j_2\in\mathbb{N}_0} 2^{2j_2p} \int_{-\infty}^{\infty} \sum_{k_1\in\mathbb{Z}} \Big| \sum_{k_2\in\mathbb{Z}} \chi_{j_2,k_2}(x_2) \Delta^2_{2^{-(j_2+1)},2}\varphi(k_1,2^{-j_2}k_2)\Big|^p dx_2 \\
= \ & \sup_{j_2\in\mathbb{N}_0} 2^{2j_2p} \sum_{k_1\in\mathbb{Z}} \int_{-\infty}^{\infty} \Big| \sum_{k_2\in\mathbb{Z}} \Delta^2_{2^{-(j_2+1)},2}\varphi(k_1,2^{-j_2}k_2)\chi_{j_2,k_2}(x_2)\Big|^p dx_2. \\
\lesssim \ & \sup_{j_2\in\mathbb{N}_0} 2^{2j_2p} \sum_{k_1\in\mathbb{Z}} \sum_{k_2\in\mathbb{Z}} \Big| \Delta^2_{2^{-(j_2+1)},2}\varphi(k_1,2^{-j_2}k_2)\Big|^p dx_2.
\end{aligned} \tag{4.5.3}$$

Using partial integration we can easily check the identity

$$\Delta^2_{2^{-(j_2+1)}} g(2^{-j_2}k_2) = 2^{-j_2} \int_{I_{j_2,k_2}} v_{j_2,k_2}(t) g^{(2)}(t)dt$$

for g being a 2-times continuously differentiable function. Hölder's inequality provides

$$|\Delta^2_{2^{-(j_2+1)}} g(2^{-j_2}k_2)|^p \lesssim 2^{-2j_2} 2^{\frac{j_2}{p}} \int_{I_{j_2,k_2}} |g^{(2)}(t)|^p dt. \tag{4.5.4}$$

Returning to (4.5.3) gives

$$\begin{aligned}
(4.5.3) &\lesssim \sup_{j_2\in\mathbb{N}_0} 2^{2j_2p} \sum_{k_1\in\mathbb{Z}} \sum_{k_2\in\mathbb{Z}} |\Delta^2_{2^{-(j_2+1)},2}\varphi(k_1, 2^{-j_2}k_2)|^p \\
&\lesssim \sup_{j_2\in\mathbb{N}_0} \sum_{k_1\in\mathbb{Z}} \underbrace{\sum_{k_2\in\mathbb{Z}} \int_{I_{j_2,k_2}} |\varphi^{(0,2)}(k_1,t)|^p dt}_{\lesssim \int_{-\infty}^{\infty} |\varphi^{(0,2)}(k_1,t)|^p dt} \\
&\lesssim \sum_{k_1\in\mathbb{Z}} \int_{-\infty}^{\infty} |\varphi^{(0,2)}(k_1,t)|^p dt \\
&= \int_{-\infty}^{\infty} \sum_{k_1\in\mathbb{Z}} |\varphi^{(0,2)}(k_1,t)|^p dt.
\end{aligned}$$

Clearly, for fixed $t \in \mathbb{R}$ Lemma 4.29 yields

$$\sum_{k_1\in\mathbb{Z}} \left|\varphi^{(0,2)}\Big(k_1, t\Big)\right|^p \lesssim \|\varphi^{(0,2)}(\cdot,t)|B^{\frac{1}{p}}_{p,1}(\mathbb{R})\|^p \lesssim \|\varphi^{(0,2)}(\cdot,t)|W^2_p(\mathbb{R})\|^p.$$

Inserting the equivalent norm from (4.5.2) gives

$$\sup_{j_2\in\mathbb{N}_0} 2^{2j_2p}\Big\| \sum_{k_1\in\mathbb{Z}} \sum_{k_2\in\mathbb{Z}} d_{\boldsymbol{j},\boldsymbol{k}}(\varphi)\chi_{\boldsymbol{j},\boldsymbol{k}}\Big|L_p(\mathbb{R}^2)\Big\|^p \lesssim \|\varphi|S^2_pW(\mathbb{R}^2)\|.$$

The case $(j_1, -1)$ works analogous and for $(j_1, j_2) \in \mathbb{N}_0^2$ we do not need Lemma 4.29. Applying (4.5.4) in both directions provides

$$\sup_{\boldsymbol{j}\in\mathbb{N}_0^2} 2^{2|\boldsymbol{j}|_1}\Big\| \sum_{\boldsymbol{k}\in\mathbb{Z}^2} d_{\boldsymbol{j},\boldsymbol{k}}(\varphi)\chi_{\boldsymbol{j},\boldsymbol{k}}\Big|L_p(\mathbb{R}^2)\Big\| \lesssim \|\varphi|S^2_pW(\mathbb{R}^2)\|.$$

That finishes the proof. □

Remark 4.31. *Compared to Theorem 4.16 we have in the limiting case $r = 2$ a B-type sequence space with fine index $\theta = \infty$ on the left hand side in (4.5.1). Later in Chapter 5 we will see that this issue causes an additional logarithmic factor in some approximation rates we have to pay.*

An argumentation analogously to the proof of Theorem 4.26 provides the following estimate on the unit cube.

Corollary 4.32. *For every function $f \in S^2_pW([0,1]^d)$ it holds*

$$\sup_{\boldsymbol{j}\in\mathbb{N}^d_{-1}} 2^{2|\boldsymbol{j}|_1}\Big\| \sum_{\boldsymbol{k}\in D_{\boldsymbol{j}}} d_{\boldsymbol{j},\boldsymbol{k}}(f)\chi_{\boldsymbol{j},\boldsymbol{k}}\Big|L_p([0,1]^d)\Big\| \lesssim \|f|S^2_pW([0,1]^d)\|. \tag{4.5.5}$$

4.6 A norm estimate for atomic superpositions

The limited regularity of the Faber-Schauder system restricts the smoothness range of Theorem 4.16 and 4.18. Later, constructing locally supported fooling functions for sampling quantities we aim to overcome the restriction $r < 1 + \frac{1}{p}$ (F-case: $r < 1 + \min\{\frac{1}{p}, \frac{1}{\theta}\}$) and cover at least the smoothness range $r < 2$ (cf. Theorem 4.19) or even more. For that purpose we consider smoother functions, that allow estimates in in b and f-type sequence spaces. We introduce the concept of atoms according to [127, p. 25].

Definition 4.33. *Let $K, L + 1 \in \mathbb{N}_0$ and $\gamma > 1$. A K-times differentiable complex valued function $a(x)$ is called $[K, L]$-atom centered at $I_{\boldsymbol{j},\boldsymbol{k}}$ (defined in (4.2.1)) if*

(i)

$$\operatorname{supp} a \subset \gamma I_{\boldsymbol{j},\boldsymbol{k}}$$

(ii)

$$|D^{\boldsymbol{\alpha}} a(x)| \leq 2^{\boldsymbol{\alpha}\cdot\boldsymbol{j}} \quad \textit{for} \quad |\boldsymbol{\alpha}|_\infty \leq K$$

(iii)

$$\int_{\mathbb{R}} x_i^m a(x) dx_i = 0 \quad \textit{if} \quad i = 1, \ldots, d \quad \textit{and} \quad m = 0, \ldots, L.$$

for $\boldsymbol{j} \in \mathbb{N}^d$.

Using the notation $S^r_{p,\theta}X$ where $X \in \{B, F\}$ and $s^r_{p,\theta}x$ with $x \in \{b, f\}$ allows us to state the following theorem.

Theorem 4.34. *Let $0 < p, \theta \leq \infty$, ($p < \infty$ in the F-case) and $r \in \mathbb{R}$. Fix $K \in \mathbb{N}_0$ and $L + 1 \in \mathbb{N}_0$ with*

$$K \geq (1 + \lceil r \rceil)_+ \quad \textit{and} \quad L \geq \max(-1, \lceil \sigma_{p,\theta} - r \rceil)$$

($L \geq \max(-1, \lceil \sigma_p - r \rceil)$ in the B-case). If $\lambda \in s^r_{p,\theta}a$ and $\{a_{\boldsymbol{j},\boldsymbol{k}}\}_{\boldsymbol{j}\in\mathbb{N}_0^d, \boldsymbol{k}\in\mathbb{Z}^d}$ are [K,L]-atoms centered at $I_{\boldsymbol{j},\boldsymbol{k}}$, then the sum

$$\sum_{\boldsymbol{j}\in\mathbb{N}_0^d} \sum_{\boldsymbol{k}\in\mathbb{Z}^d} \lambda_{\boldsymbol{j},\boldsymbol{k}} a_{\boldsymbol{j},\boldsymbol{k}}(x)$$

converges in $S'(\mathbb{R}^d)$, its limit f belongs to the space $S^r_{p,\theta}X(\mathbb{R}^d)$ and

$$\|f|S^r_{p,\theta}X(\mathbb{R}^d)\| \leq C\|\lambda|s^r_{p,\theta}x\|, \tag{4.6.1}$$

where the constant C is universal for all admissible λ and $a_{\boldsymbol{j},\boldsymbol{k}}$.

Proof. We refer to [127, Theorem 2.4]. We should mention, that the proof of this theorem is based on convolution inequalities as in Lemma 4.12. Here for $[K, L]$-atoms $(a_{\boldsymbol{j},\boldsymbol{k}})_{\boldsymbol{j}\in\mathbb{N}_0^d,\boldsymbol{k}\in\mathbb{Z}^d}$ the convolution can be estimated by

$$|\Psi_{\boldsymbol{j}} * a_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}}(x)| \leq C_R 2^{-|\boldsymbol{\ell}_+|K} 2^{-|\boldsymbol{\ell}_-|(L+2)} \prod_{i=1,\ldots,d} (1 + 2^{\min\{j_i, j_i+\ell_i\}}|x_i - 2^{-(j_i+\ell_i)}k_i|)^{-R}.$$

Then the technique presented in Lemma 4.14 (F-case) and Lemma 4.15 (B-case) together with the simple estimate

$$\|f|S^r_{p,\theta}F(\mathbb{R}^d)\| \lesssim \Big(\sum_{\boldsymbol{v}\in\{0,1\}^d} \|T^*_{\boldsymbol{v}}\lambda|L_p(\ell_\theta(2^{|\boldsymbol{j}|_1 r}, \mathbb{N}_0^d), \mathbb{R}^d)|L_p(\mathbb{R}^d)\|^u\Big)^{\frac{1}{u}} \lesssim \|\lambda|s^r_{p,\theta}f\|$$

proves the claim (obvious changes in B case). Here

$$T^*_{\boldsymbol{v}} : s^{\boldsymbol{r}}_{p,\theta}f \to L_p(\ell_\theta, 2^{\boldsymbol{j}\cdot\boldsymbol{r}})$$

is given by

$$(\lambda_{\boldsymbol{j},\boldsymbol{k}})_{\boldsymbol{j}\in\mathbb{N}_{-1}^d,\boldsymbol{k}\in\mathbb{Z}^d} \mapsto \Big(\sum_{\boldsymbol{\ell}\in B(\boldsymbol{v})} \sum_{\boldsymbol{k}\in\mathbb{Z}^d} \lambda_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}} \Psi_{\boldsymbol{j}} * a_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}}\Big)_{\boldsymbol{j}\in\mathbb{N}_0^d}.$$

□

Remark 4.35. *Obviously, if additionally* $\operatorname{supp} a_{\boldsymbol{j},\boldsymbol{k}} \subset I_{\boldsymbol{j},\boldsymbol{k}}$ *for all* $\boldsymbol{j} \in \mathbb{N}_0^d, \boldsymbol{k} \in D_{\boldsymbol{j}}$ *then* (4.6.1) *can be refined to*

$$\Big\|\sum_{\boldsymbol{j}\in\mathbb{N}_0^d} \sum_{\boldsymbol{k}\in D_{\boldsymbol{j}}} \lambda_{\boldsymbol{j},\boldsymbol{k}} a_{\boldsymbol{j},\boldsymbol{k}} |S^r_{p,\theta}A([0,1]^d)\Big\| \lesssim \|\lambda|s^{r,\Omega}_{p,\theta}a\|.$$

That means we obtain a related estimate on the domain $[0,1]^d$*.*

Chapter 5

(Energy-)Sparse grid approximation

The upcoming two chapters deal with approximation aspects of the Faber-Schauder system in spaces with non-periodic boundary conditions. A sparse grid with asymptotically $M^{d-1}2^M$ points is the set

$$\mathcal{G}_M^{sparse} := \{(2^{-j_1}k_1, \ldots, 2^{-j_d}k_d) : \boldsymbol{k} \in \bigtimes_{i=1}^{d} \{0, \ldots, 2^{j_i}\}, |\boldsymbol{j}|_1 \leq M\}. \tag{5.0.1}$$

We use samples generated on $\mathcal{G}_M^{sparse}$ to approximate functions $f \in S_p^r W([0,1]^d)$ in the $L_q([0,1]^d)$-norm. The Faber-Schauder system has a restricted regularity that causes attention concerning smoothness, fine index and integrability of the function classes we consider. For a improved visibility of this effects we restrict to model spaces $S_p^r W([0,1]^d)$ in this chapter. In Chapter 9, the requirements for the trigonometric sampling representation are less critical. There we point out more general approximation results for the periodic setting. The second half of this chapter is about measuring the error in the energy norm $H^1([0,1]^d)$. It turns out that a modification of Smolyak's algorithm which generates a so called energy sparse grid yields optimal sampling rates.

5.1 Hierarchical sparse grid approximation

First we deal with the approximation of functions where the error is measured in $L_q([0,1]^d),\ 1 < q \leq \infty$.

Definition 5.1. *We define for $f \in C([0,1]^d)$ the (linear) Smolyak Faber-Schauder sampling operator*

$$I_M f = \sum_{|\boldsymbol{j}|_1 \leq M} \sum_{\boldsymbol{k} \in D_{\boldsymbol{j}}} d_{\boldsymbol{j},\boldsymbol{k}}(f) v_{\boldsymbol{j},\boldsymbol{k}}. \tag{5.1.1}$$

By construction this operator samples a continuous function f on $\mathcal{G}_M^{sparse}$ (defined in (5.0.1)). One can easily check that the cardinality of $\mathcal{G}_M^{sparse}$ can be described by

$$|\mathcal{G}_M^{sparse}| \asymp \sum_{|\boldsymbol{j}|_1 \leq M} 2^{|\boldsymbol{j}|_1} \asymp M^{d-1}2^M,$$

cf. Lemma C.21.

Lemma 5.2. *For $f \in C([0,1]^d)$ and $M > 0$ we have*

$$I_M f(\boldsymbol{x}) = f(\boldsymbol{x})$$

for all $\boldsymbol{x} \in \mathcal{G}_M^{sparse}$.

Proof. The interpolation property of the (at level M truncated) univariate Faber-Schauder series expansion on $[0,1]$ immediately gives an interpolation property of the $|\boldsymbol{j}|_\infty \le M$-truncated multivariate expansion on a "full grid" (see Lemma 4.6). Arguing similar as in [107, Lem. 4.3] we obtain the interpolation property on sparse grids $\mathcal{G}_M^{sparse}$ stated above for I_M. □

Lemma 5.3. *Let $M > 0$. Then*

$$\operatorname{rank} I_M = M^{d-1} 2^M.$$

Proof. The direction $\operatorname{rank} I_M \lesssim M^{d-1}2^M$ is obvious, since I_M samples by construction on a sparse grid $\mathcal{G}_M^{sparse}$ with $|\mathcal{G}_M^{sparse}| \asymp M^{d-1}2^M$ sampling nodes. The lower bound comes from the fact that I_M reproduces the set

$$V_M = \operatorname{span}\{v_{\boldsymbol{j},\boldsymbol{k}} : |\boldsymbol{j}|_1 \le M, \boldsymbol{k} \in D_{\boldsymbol{j}}\}$$

with $\dim V_M = M^{d-1}2^M$ (cf. Lemma 5.2). This type of arguments are well known for Smolyak type algorithms in the periodic context and transfer one-to-one to hat functions. □

To improve the length of presentation in the upcoming proofs we use the conventions

$$u_{\boldsymbol{j}} = \sum_{\boldsymbol{k}\in D_{\boldsymbol{j}}} d_{\boldsymbol{j},\boldsymbol{k}} v_{\boldsymbol{j},\boldsymbol{k}} \quad \text{and} \quad u_{\boldsymbol{j}}^* = \sum_{\boldsymbol{k}\in D_{\boldsymbol{j}}} d_{\boldsymbol{j},\boldsymbol{k}} \chi_{\boldsymbol{j},\boldsymbol{k}}$$

and start proving some approximation rates. First we consider the case where we have less (or equal) integrability in the target space than in the source space.

Theorem 5.4. *Let $1 < q \le p < \infty$ and $\max\{\frac{1}{p}, \frac{1}{2}\} < r < 2$. Then we obtain*

$$\|f - I_M f | L_q([0,1]^d\| \lesssim M^{\frac{d-1}{2}} 2^{-Mr} \|f | S_p^r W([0,1]^d)\|$$

for all $M \in \mathbb{N}$.

Proof. The expansion in (4.4.3), the embedding $L_p([0,1]^d) \hookrightarrow L_q([0,1]^d)$ and Lemma 4.27 together with Hölder's inequality yield

$$\begin{aligned} \Big\| f - \sum_{|\boldsymbol{j}|_1 \le M} u_{\boldsymbol{j}} \Big| L_q([0,1]^d) \Big\| &\le \Big\| \sum_{|\boldsymbol{j}|_1 > M} u_{\boldsymbol{j}} \Big| L_p([0,1]^d) \Big\| \\ &\lesssim \Big(\sum_{|\boldsymbol{j}|_1 > M} 2^{-2|\boldsymbol{j}|_1 r} \Big)^{\frac{1}{2}} \Big\| \Big(\sum_{|\boldsymbol{j}|_1 > M} 2^{2r|\boldsymbol{j}|_1} |u_{\boldsymbol{j}}^*|^2 \Big)^{\frac{1}{2}} \Big| L_p([0,1]^d) \Big\|. \end{aligned}$$

The estimate for the sum in Lemma C.20 together with Theorem 4.26 yields

$$\Big\| f - \sum_{|\boldsymbol{j}|_1 \le M} u_{\boldsymbol{j}} \Big| L_q([0,1]^d) \Big\| \lesssim 2^{-rM} M^{\frac{d-1}{2}} \|f | S_p^r W([0,1]^d)\|.$$

□

Next we proof a result where the integrability in the target space is greater than in the source space.

Theorem 5.5. *Let* $1 < p < q < \infty$ *and* $\frac{1}{p} < r < 2 + \frac{1}{p} - \frac{1}{q}$*. Then we obtain*

$$\|f - I_M f|L_q([0,1]^d)\| \lesssim 2^{-M(r-\frac{1}{p}+\frac{1}{q})}\|f|S^r_pW([0,1]^d)\|$$

for all $M \in \mathbb{N}$.

Proof. The expansion in (4.4.3) together with Lemma 4.27 yield

$$\Big\|f - \sum_{|\boldsymbol{j}|_1 \le M} u_{\boldsymbol{j}}\Big|L_q([0,1]^d)\Big\| \;\lesssim\; \sup_{|\boldsymbol{j}|_1 > M} 2^{-(r-(\frac{1}{p}-\frac{1}{q}))|\boldsymbol{j}|_1}\Big\|\sum_{|\boldsymbol{j}|_1 > M} 2^{(r-(\frac{1}{p}-\frac{1}{q}))|\boldsymbol{j}|_1}|u^*_{\boldsymbol{j}}|\Big|L_q([0,1]^d)\Big\|.$$

We choose q^* with $p < q^* < q$ close to q with $r - (\frac{1}{p} - \frac{1}{q^*}) < 2$. Applying the diagonal embedding stated in Lemma 4.10, (iv) gives

$$\Big\|f - \sum_{|\boldsymbol{j}|_1 \le M} u_{\boldsymbol{j}}\Big|L_q([0,1]^d)\Big\| \;\le\; 2^{-(r-(\frac{1}{p}-\frac{1}{q}))M}\Big\|\sup_{|\boldsymbol{j}|_1 > M} 2^{r-(\frac{1}{p}-\frac{1}{q^*})|\boldsymbol{j}|_1}|u^*_{\boldsymbol{j}}|\Big|L_{q^*}([0,1]^d)\Big\|.$$

Applying Theorem 4.26 yields

$$\Big\|f - \sum_{|\boldsymbol{j}|_1 \le M} u_{\boldsymbol{j}}\Big|L_q([0,1]^d)\Big\| \;\le\; 2^{-(r-(\frac{1}{p}-\frac{1}{q}))M}\|f|S^{r-(\frac{1}{p}-\frac{1}{q^*})}_{q^*,\infty}F([0,1]^d)\|.$$

Finally the diagonal embedding

$$S^r_pW([0,1]^d) \hookrightarrow S^{r-(\frac{1}{p}-\frac{1}{q^*})}_{q^*,\infty}F([0,1]^d)$$

(cf. Lemma 3.4, (vi)) finishes the proof. □

The proof of Theorem 5.5 shows that this approximation rate holds for a bigger class of functions namely the mixed Triebel-Lizorkin space with fine index $\theta = \infty$. Finally we investigate the special case $q = \infty$.

Theorem 5.6. *Let* $1 < p < \infty$ *and* $\frac{1}{p} < r < 2 + \frac{1}{p}$*. Then*

$$\|f - I_M f|L_\infty([0,1]^d)\| \;\lesssim\; M^{(d-1)(1-\frac{1}{p})}2^{-M(r-\frac{1}{p})}\|f|S^r_pW(\mathbb{R}^d)\|$$

holds for all $M \in \mathbb{N}$.

Proof. Step 1. We prove

$$\|f - I_M f|L_\infty([0,1]^d)\| \lesssim \|f|S^{r-\frac{1}{p}}_{\infty,p}B([0,1]^d)M^{(d-1)(1-\frac{1}{p})}2^{-M(r-\frac{1}{p})}. \tag{5.1.2}$$

Expanding into (4.4.3) then Lemma 4.28 yields

$$\|f - I_M f|L_\infty([0,1]^d)\| \;=\; \Big\|\sum_{|\boldsymbol{j}|_1 > M} u_{\boldsymbol{j}}\Big|L_\infty([0,1]^d)\| \lesssim \sum_{|\boldsymbol{j}|_1 > M} \|u^*_{\boldsymbol{j}}|L_\infty([0,1]^d)\|.$$

We apply Hölder's inequality with $1 = \frac{1}{p} + \frac{1}{p'}$ and obtain

$$\|f - I_M f|L_\infty([0,1]^d)\| \lesssim \Big(\sum_{|\boldsymbol{j}|_1 > M} 2^{-p'(r-\frac{1}{p})|\boldsymbol{j}|_1}\Big)^{\frac{1}{p'}} \Big(\sum_{|\boldsymbol{j}|_1 > M} 2^{p(r-\frac{1}{p})|\boldsymbol{j}|_1} \|u^*_{\boldsymbol{j}}|L_\infty([0,1]^d)\|^p\Big)^{\frac{1}{p}}.$$

Lemma C.20 yields

$$\|f - I_M f|L_\infty([0,1]^d)\| \lesssim M^{(d-1)(1-\frac{1}{p})} 2^{-M(r-\frac{1}{p})} \Big(\sum_{|\boldsymbol{j}|_1 > M} 2^{p(r-\frac{1}{p})|\boldsymbol{j}|_1} \|u^*_{\boldsymbol{j}}|L_\infty([0,1]^d)\|^p\Big)^{\frac{1}{p}}.$$

Applying Theorem 4.26 yields (5.1.2).
Step 2. The Jawerth-Franke type embedding implies

$$S^r_p W([0,1]^d) \hookrightarrow S^{r-\frac{1}{p}}_{\infty,p} B([0,1]^d)$$

(cf. Lemma 3.5). Applying this we obtain

$$\|f - I_M f|L_\infty([0,1]^d)\| \lesssim M^{(d-1)(1-\frac{1}{p})} 2^{-M(r-\frac{1}{p})} \|f|S^r_p W([0,1]^d)\|,$$

which proves the claim. □

The next Theorem was obtained in [8, Proposition 3.8] for $p = 2$. A close related version for $1 < p < \infty$ in the context of spline interpolation is stated in [107, Corollary 5.3]. We use the Faber-Schauder sampling characterizations to reproduce such a result.

Theorem 5.7. *Let $1 < q \le p < \infty$. Then we obtain*

$$\|f - I_M f|L_q([0,1]^d\| \lesssim M^{d-1} 2^{-2M} \|f|S^2_p W([0,1]^d)\|$$

for all $M \in \mathbb{N}$.

Proof. Applying the expansion in (4.4.3) then Lemma 4.28 yields

$$\begin{aligned}
\Big\| f - \sum_{|\boldsymbol{j}|_1 \le M} u_{\boldsymbol{j}} \Big| L_q([0,1]^d)\Big\| &\le \Big\| \sum_{|\boldsymbol{j}|_1 > M} u_{\boldsymbol{j}} \Big| L_p([0,1]^d)\Big\| \\
&\lesssim \sum_{|\boldsymbol{j}|_1 > M} \|u^*_{\boldsymbol{j}}|L_p([0,1]^d)\| \\
&\le \sup_{|\boldsymbol{j}|_1 > M} 2^{2|\boldsymbol{j}|_1} \|u^*_{\boldsymbol{j}}|L_p([0,1]^d)\| \\
&\quad \sum_{|\boldsymbol{j}|_1 > M} 2^{-2|\boldsymbol{j}|_1}.
\end{aligned}$$

We apply Theorem 4.30 and obtain

$$\Big\| f - \sum_{|\boldsymbol{j}|_1 \le M} u_{\boldsymbol{j}} \Big| L_q([0,1]^d)\Big\| \lesssim \|f|S^2_2 W([0,1]^d)\| \sum_{|\boldsymbol{j}|_1 > M} 2^{-2|\boldsymbol{j}|_1}.$$

The estimate for the sum in Lemma C.20 yields

$$\Big\|f - \sum_{|\boldsymbol{j}|_1 \le M} u_{\boldsymbol{j}} \Big| L_q([0,1]^d)\Big\| \lesssim 2^{-2M} M^{d-1} \|f|S_p^2 W([0,1]^d)\|.$$

That concludes the proof. □

Remark 5.8. *Comparing the estimate for the convergence rate in Theorem 5.4 with the limiting case* $r = 2$ *considered in Theorem 5.7 we observe an additional factor* $M^{\frac{d-1}{2}}$ *for the limiting case. It is unknown whether this additional factor is seriously required or only caused by a technical issue.*

Finally we consider the case $p < q$ with smoothness $r = 2 + \frac{1}{p} - \frac{1}{q}$.

Theorem 5.9. *Let* $1 < p < q < \infty$*. Then we obtain*

$$\|f - I_M f|L_q([0,1]^d)\| \lesssim 2^{-2M} M^{d-1} \|f|S_p^{2+\frac{1}{p}-\frac{1}{q}} W([0,1]^d)\|$$

where $\operatorname{rank} I_M = M^{d-1} 2^M$*.*

Proof. The expansion in (4.4.3) and Lemma 4.28 yield

$$\begin{aligned}\Big\|f - \sum_{|\boldsymbol{j}|_1 \le M} u_{\boldsymbol{j}} \Big| L_q([0,1]^d)\Big\| &\lesssim \sum_{|\boldsymbol{j}|_1 > M} \|u_{\boldsymbol{j}}^* | L_q([0,1]^d)\|. \\ &\lesssim \sup_{|\boldsymbol{j}|_1 > M} 2^{2|\boldsymbol{j}|_1} \|u_{\boldsymbol{j}}^* | L_q([0,1]^d)\| \sum_{|\boldsymbol{j}|_1 > M} 2^{-2|\boldsymbol{j}|_1}.\end{aligned}$$

The estimate for the sum in Lemma C.20 gives

$$\Big\|f - \sum_{|\boldsymbol{j}|_1 \le M} u_{\boldsymbol{j}} \Big| L_q([0,1]^d)\Big\| \lesssim \sup_{|\boldsymbol{j}|_1 > M} 2^{|2\boldsymbol{j}|_1} \|u_{\boldsymbol{j}}^* | L_q([0,1]^d)\| 2^{-2M} M^{d-1}.$$

Theorem 4.30 provides

$$\Big\|f - \sum_{|\boldsymbol{j}|_1 \le M} u_{\boldsymbol{j}} \Big| L_q([0,1]^d)\Big\| \lesssim 2^{-2M} M^{d-1} \|f|S_q^2 W([0,1]^d)\|.$$

We apply the diagonal embedding stated in Lemma 4.10 and obtain

$$\Big\|f - \sum_{|\boldsymbol{j}|_1 \le M} u_{\boldsymbol{j}} \Big| L_q([0,1]^d)\Big\| \lesssim 2^{-2M} M^{d-1} \|f|S_p^{2+\frac{1}{p}-\frac{1}{q}} W([0,1]^d)\|.$$

That concludes the proof. □

Remark 5.10. *Here we obtain an additional factor* M^{d-1} *compared to the non-limiting case.*

5.2 Optimal sparse grid approximation

Let X, Y be (quasi-)Banach spaces with $X \hookrightarrow Y \cap C([0,1]^d)$. Then we define the quantity

$$\varrho_n^{\mathrm{SG}}(X,Y) := \inf_{\substack{M\in\mathbb{N}:|\mathcal{G}_M^{sparse}|\le n\\ \varphi:\mathbb{C}^n\to Y}} \sup_{\|f|X\|\le 1} \|f - \varphi(f(\mathcal{G}_M^{sparse})|Y\|, \tag{5.2.1}$$

which we call sparse grid sampling width. It denotes the best worst-case error for the approximation of functions belonging to the unit ball of X by algorithms that can be described as a composition of a (possibly non-linear) reconstruction map $\varphi : \mathbb{C}^n \to Y$ and an information map, which are in our case simply the functions values of f on a sparse grid $\mathcal{G}_M^{sparse}$ with $|\mathcal{G}_M^{sparse}| \le n$. This quantity is a special restriction of the IBC worst case error for standard information [84, 85, 86]. They were introduced in [35], where the focus is on $X = S^r_{p,\theta}B([0,1]^d)$ and $Y = L_q([0,1]^d)$. We use this results for the case $X = S^r_pW([0,1]^d)$ and $Y = L_q([0,1]^d)$. The following Lemma describes a method to bound this quantity from below.

Lemma 5.11. *For $1 < p, q < \infty$ ($q = \infty$) and $r > \frac{1}{p}$ a lower bound is provided by*

$$\varrho_n^{SG}(S^r_pW([0,1]^d)), L_q([0,1]^d)) \gtrsim \inf_{M\in\mathbb{N}:|\mathcal{G}_M^{sparse}|\le n} \sup_{\substack{\|f|S^r_pW([0,1]^d)\|\le 1\\ f(\boldsymbol{x})=0,\forall \boldsymbol{x}\in\mathcal{G}_M^{sparse}}} \|f|L_q([0,1]^d)\|.$$

Proof. Let $\varphi : \mathbb{C}^n \to L_q([0,1]^d)$ be an arbitrary reconstruction map and $\|f|S^r_pW([0,1]^d)\| \le 1$ with $f(\boldsymbol{x}) = 0, \forall \boldsymbol{x} \in \mathcal{G}_M^{sparse}$ with $|\mathcal{G}_M^{sparse}| \le n$. Then

$$\begin{aligned} \|f|L_q([0,1]^d)\| &= \left\| \frac{1}{2}(f - \varphi(0)) - \frac{1}{2}(-f - \varphi(0)) \middle| L_q([0,1]^d \right\| \\ &\le \frac{1}{2}\|f - \varphi(0)|L_q([0,1]^d\| + \frac{1}{2}\| - f - \varphi(0)|L_q([0,1]^d\|. \end{aligned}$$

Finally either $\|f - \varphi(0)|L_q([0,1]^d\| \ge \|f|L_q([0,1]^d\|$ or $\| - f - \varphi(0)|L_q([0,1]^d\| \ge \|f|L_q([0,1]^d\|$. That proves the claim. □

Remark 5.12. *The sparse grid structure plays no essential role in the proof provided in Lemma 5.11. Later the same arguments will be applied to obtain lower bounds for the worst case error for standard information.*

Remark 5.13. *It is easy to check that nestedness properties of the points $x_{\boldsymbol{j},\boldsymbol{k}}$ (for different levels $\boldsymbol{j}$) allow us to write the sparse grid of order M as*

$$\mathcal{G}_M^{sparse} = \{(2^{-j_1}k_1, \ldots, 2^{-j_d}k_d) : \boldsymbol{k} \in \bigtimes_{i=1}^{d} \{0, \ldots, 2^{j_i}\}, |\boldsymbol{j}|_1 = M\}. \tag{5.2.2}$$

Theorem 5.14. *Let $1 < q \le p < \infty$ and $\max\{\frac{1}{p}, \frac{1}{2}\} < r < 2$. Then we can estimate as follows*

$$\varrho_n^{SG}(S^r_pW([0,1]^d), L_q[(0,1)]^d) \asymp \sup_{\|f|S^r_pW([0,1]^d)\|} \|f - I_M f|L_q([0,1]^d\| \asymp (n^{-1}\log^{d-1} n)^r \log^{\frac{d-1}{2}}$$

with $\operatorname{rank} I_M \asymp n \asymp M^{d-1}2^M$.

Proof. Inserting the relation $n := |\mathcal{G}_M^{sparse}| \asymp M^{d-1}2^M$ into Theorem 5.4 gives the upper bound. Next we prove the lower bound. For that purpose we consider the bump function

$$b(x) = e^{-\frac{1}{x(1-x)}} e^{\frac{1}{4}} \tag{5.2.3}$$

which is a L_∞-normalized C_0^∞-function. We denote by

$$b_{\boldsymbol{j},\boldsymbol{k}} = \prod_{i=1}^{d} b(2^{j_i}x_i - k_i) \tag{5.2.4}$$

its $\boldsymbol{j}$-th dilation and $\boldsymbol{k}$-th tensorized translation. Obviously

$$\operatorname{supp} b_{\boldsymbol{j},\boldsymbol{k}} = \bigtimes_{i=1}^{d} [2^{-j_i}k_i, 2^{-j_i}(k_i+1)]$$

with

$$b_{\boldsymbol{j},\boldsymbol{k}}(2^{-j_1}(k_1+\nu_1), \ldots, 2^{-j_d}(k_d+\nu_d)) = 0 \tag{5.2.5}$$

for $\boldsymbol{\nu} \in \{0,1\}^d$, $\boldsymbol{k} \in D_{\boldsymbol{j}}$. It is easy to check that,

$$\|b_{\boldsymbol{j},\boldsymbol{k}}|L_q([0,1]^d)\| \asymp 2^{-\frac{|\boldsymbol{j}|_1}{q}} \tag{5.2.6}$$

and that due to disjoint supports

$$\Big\| \sum_{\boldsymbol{k}\in D_{\boldsymbol{j}}} b_{\boldsymbol{j},\boldsymbol{k}} \Big| L_q([0,1]^d) \Big\| \asymp 1 \tag{5.2.7}$$

holds. Defining

$$\varphi_1 := C 2^{-Mr} M^{-\frac{d-1}{2}} \sum_{|\boldsymbol{j}|_1=M} \sum_{\boldsymbol{k}\in D_{\boldsymbol{j}}} b_{\boldsymbol{j},\boldsymbol{k}}$$

we can estimate using Theorem 4.34

$$\|\varphi_1|S_p^rW([0,1]^d)\| \lesssim M^{-\frac{d-1}{2}} \Big\| \Big(\sum_{|\boldsymbol{j}|_1=M} \underbrace{\Big| \sum_{\boldsymbol{k}\in D_{\boldsymbol{j}}} \chi_{\boldsymbol{j},\boldsymbol{k}} \Big|^2}_{=1} \Big)^{\frac{1}{2}} \Big| L_p([0,1]^d) \Big\| \lesssim 1$$

($b_{\boldsymbol{j},\boldsymbol{k}} : L = -1,\ K = \infty$). By construction $\varphi_1(\boldsymbol{x}) = 0$ for all $\boldsymbol{x} \in \mathcal{G}_M^{sparse}$ (cf. (5.2.5) and the definition of $\mathcal{G}_M^{sparse}$ in (5.0.1)). This allows us to estimate

$$\varrho_n^{\mathrm{SG}}(S_p^rW([0,1]^d), L_q([0,1]^d)) \geq \|\varphi_1|L_q([0,1]^d)\| \geq \|\varphi_1|L_1([0,1]^d)\|.$$

The relation in (5.2.7) yields

$$\begin{aligned} \varrho_n^{\mathrm{SG}}(S_p^rW([0,1]^d), L_q([0,1]^d)) &\gtrsim 2^{-Mr} M^{-\frac{d-1}{2}} \underbrace{\Big\| \sum_{|\boldsymbol{j}|_1=M} \sum_{\boldsymbol{k}\in D_{\boldsymbol{j}}} b_{\boldsymbol{j},\boldsymbol{k}} \Big| L_1([0,1]^d) \Big\|}_{\asymp M^{d-1}} \\ &\asymp 2^{-Mr} M^{\frac{d-1}{2}} \asymp (n^{-1}\log^{d-1} n)^r (\log^{d-1} n)^{\frac{d-1}{2}}. \end{aligned}$$

That finishes the proof. □

Theorem 5.15. *Let $1 < p < q < \infty$ and $\frac{1}{p} < r < 2 + \frac{1}{p} - \frac{1}{q}$ then*

$$\varrho_n^{SG}(S_p^r W([0,1]^d), L_q([0,1]^d)) \asymp \sup_{\|f|S_p^r W([0,1]^d)\|} \|f - I_M f|L_q([0,1]^d)\| \asymp (n^{-1}\log^{d-1} n)^{r-(\frac{1}{p}-\frac{1}{q})}$$

with $\operatorname{rank} I_M \asymp n \asymp M^{d-1}2^M$.

Proof. Inserting the relation $n \asymp |\mathcal{G}_M^{sparse}| \asymp 2^M M^{d-1}$ into Theorem 5.5 proves the upper bound. We prove the lower bound now. Let $b_{\boldsymbol{j},\boldsymbol{k}}$ as in (5.2.4). We define

$$\varphi_2 := 2^{-(r-\frac{1}{p})M} b_{(M+1,0,\dots,0),(0,\dots,0)} \tag{5.2.8}$$

Theorem 4.34 together with (5.2.6) yields

$$\|\varphi_2|S_p^r W([0,1]^d)\| \lesssim 1.$$

Again, by construction $\varphi_2(\boldsymbol{x}) = 0$ for all $\boldsymbol{x} \in SG(M)$. This allows us to estimate

$$\begin{aligned} \varrho_n^{\mathrm{SG}}(S_p^r W([0,1]^d), L_q([0,1]^d)) &\geq \|\varphi_2\|_q \\ &\asymp 2^{-(r-\frac{1}{p})M}\|b_{(M+1,0,\dots,0),(0,\dots,0)}|L_q([0,1]^d)\|. \end{aligned}$$

Finally inserting the estimate in 5.2.6 gives

$$\varrho_n^{\mathrm{SG}}(S_p^r W([0,1]^d), L_q([0,1]^d)) \gtrsim 2^{-(r-\frac{1}{p}+\frac{1}{q})M} \asymp (n^{-1}\log^{d-1} n)^{r-(\frac{1}{p}-\frac{1}{q})}.$$

That proves the claim. □

Theorem 5.16. *Let $1 < p < \infty$ and $\frac{1}{p} < r < 2 + \frac{1}{p}$. Then*

$$\begin{aligned} \varrho_n^{SG}(S_p^r W([0,1]^d), L_\infty([0,1]^d)) &\asymp \sup_{\|f|S_p^r W([0,1]^d)\|} \|f - I_M f|L_q([0,1]^d)\| \\ &\asymp (n^{-1}\log^{d-1} n)^{r-\frac{1}{p}}(\log^{d-1} n)^{1-\frac{1}{p}} \end{aligned}$$

with $\operatorname{rank} I_M \asymp n \asymp M^{d-1}2^M$.

Proof. Inserting the relation $n \asymp |\mathcal{G}_M^{sparse}| \asymp 2^M M^{d-1}$ into Theorem 5.6 proves the upper bound. We prove the lower bound now. Let $b_{\boldsymbol{j},\boldsymbol{k}}$ as in (5.2.4). We define

$$\varphi_3 := M^{-\frac{d-1}{p}} 2^{-M(r-\frac{1}{p})} \sum_{|\boldsymbol{j}|_1 = M} b_{\boldsymbol{j},(0,\dots,0)} \tag{5.2.9}$$

and distinguish the cases $1 < p \leq 2$ and $2 < p < \infty$. In case $1 < p \leq 2$ Lemma 3.4 and Theorem 4.34 yield

$$\|\varphi_3|S_p^r W([0,1]^d)\| \lesssim \|\varphi_3|S_{p,p}^r B([0,1]^d)\| \lesssim M^{-\frac{d-1}{p}} \Big(\underbrace{\sum_{|\boldsymbol{j}|_1 = M} 1}_{\lesssim M^{d-1}} \Big)^{\frac{1}{p}} \lesssim 1.$$

In case $2 < p < \infty$ the non-compact embedding in Lemma 3.5 and Theorem 4.34 yield

$$\|\varphi_3|S^r_pW([0,1]^d)\| \lesssim \|\varphi_3|S^{r+\frac{1}{2}-\frac{1}{p}}_{2,p}B([0,1]^d)\| \lesssim M^{-\frac{d-1}{p}}\Big(\underbrace{\sum_{|\boldsymbol{j}|_1=M} 1}_{\lesssim M^{d-1}}\Big)^{\frac{1}{p}} \lesssim 1.$$

Again, by construction $\varphi_3(\boldsymbol{x}) = 0$ for all $\boldsymbol{x} \in SG(M)$. This allows us to estimate

$$\begin{aligned} \varrho^{\mathrm{SG}}_n(S^r_pW([0,1]^d), L_\infty([0,1]^d)) &\geq \|\varphi_3\|_\infty \\ &\asymp M^{(d-1)(1-\frac{1}{p})}2^{-M(r-\frac{1}{p})}\underbrace{\|b_{\boldsymbol{j},(0,\dots,0)}|L_\infty([0,1]^d)\|}_{=1}. \end{aligned}$$

Finally inserting the relation $n \asymp M^{d-1}2^M$ gives

$$\begin{aligned} \varrho^{\mathrm{SG}}_n(S^r_pW([0,1]^d), L_\infty([0,1]^d)) &\gtrsim 2^{-(r-\frac{1}{p})M}M^{(d-1)(1-\frac{1}{p})} \\ &\asymp (n^{-1}\log^{d-1} n)^{r-\frac{1}{p}}(\log^{d-1} n)^{1-\frac{1}{p}}. \end{aligned}$$

That proves the claim. □

Remark 5.17. *In the limiting case with $r = 2$ and $p \geq q$ (or $r = 2 + \frac{1}{p} - \frac{1}{q}$ in case $p < q$) we are not able to prove sharp bounds for*

$$\sup_{\|f|S^r_pW([0,1]^d)\|} \|f - I_M f|L_q([0,1]^d)\|.$$

We obtain logarithmic gaps between the upper bounds and the lower bounds for sparse grid sampling widths obtained in Theorems 5.14 and 5.15 (which are valid also for $r \geq 2$).

5.3 Sampling recovery in the energy-norm

For the rest of this chapter we are interested in measuring sampling errors in the energy norm $H^1([0,1]^d) := W^1_2([0,1]^d)$. The interest in this setting is motivated by the convergence analysis of Galerkin methods. Energy sparse grids depend on the ratio of the smoothness in the model and the target space. This point sets can be defined as

$$\mathcal{G}^{energy}_{\Delta_{\alpha,\beta}(M)} := \{(2^{-j_1}k_1,\dots,2^{-j_d}k_d) : \boldsymbol{k} \in D_{\boldsymbol{j}}, \boldsymbol{j} \in \mathbb{N}^d_{-1}, \alpha|\boldsymbol{j}|_1 - \beta|\boldsymbol{j}|_\infty \leq M\}.$$

where α and β are the mentioned degrees of freedom. The first reference where we could find this approach is the PhD thesis of Knapek [67]. Sampling in combination with measuring the error in the energy norm was also considered in [8], [9], [10], [30] and [48]. We continue considering the sampling operator

$$I_{\Delta_{\alpha,\beta}(M)}f(x) := \sum_{\boldsymbol{j}\in\Delta_{\alpha,\beta}(M)}\sum_{\boldsymbol{k}\in D_{\boldsymbol{j}}} d_{\boldsymbol{j},\boldsymbol{k}}(f)v_{\boldsymbol{j},\boldsymbol{k}} \qquad (5.3.1)$$

with

$$\Delta_{\alpha,\beta}(M) := \Big\{ \boldsymbol{j} \in \mathbb{N}_0^d : \alpha|\boldsymbol{j}|_1 - \beta|\boldsymbol{j}|_\infty \le M \Big\}. \tag{5.3.2}$$

Let $\mathcal{G}^{energy}_{\Delta_{\alpha,\beta}(M)}$ denote the grid of sampling nodes used by $I_{\Delta_{\alpha,\beta}(M)} f$. Inserting the definition one can easily verify that

$$|\mathcal{G}^{energy}_{\Delta_{\alpha,\beta}(M)}| \asymp \sum_{\boldsymbol{j} \in \Delta_{\alpha,\beta}(M)} 2^{|\boldsymbol{j}|_1}$$

which gives under the conditions of Lemma C.22

$$|\mathcal{G}^{energy}_{\Delta_{\alpha,\beta}(M)}| \asymp 2^{\frac{M}{\alpha-\beta}}.$$

For this operator we can prove the following convergence theorem.

Theorem 5.18. *Let* $1 < p < \infty$ *and*

$$1 + \Big(\frac{1}{p} - \frac{1}{2}\Big)_+ < r < 2 + \Big(\frac{1}{p} - \frac{1}{2}\Big)_+.$$

Then there exists a constant $C_\varepsilon > 0$ *(independent of* f *and* M*) such that*

$$\|f - I_{\Delta_{\alpha,\beta}(M)} f | H^1([0,1]^d)\| \le C_\varepsilon 2^{-M} \|f|S^r_pW([0,1]^d)\| \tag{5.3.3}$$

with

$$\alpha = r - \Big(\frac{1}{p} - \frac{1}{2}\Big)_+ - \varepsilon \quad \textit{and} \quad \beta = 1 - \varepsilon$$

where

$$0 < \varepsilon < 1.$$

Proof. We expand f into the series (4.4.3)

$$\|f - I_{\Delta_{\alpha,\beta}(M)} f | H^1([0,1]^d)\| \lesssim \Big\| \sum_{\boldsymbol{j} \notin \Delta_{\alpha,\beta}(M)} \sum_{\boldsymbol{k} \in D_{\boldsymbol{j}}} d_{\boldsymbol{j},\boldsymbol{k}}(f) v_{\boldsymbol{j},\boldsymbol{k}} \Big| H^1([0,1]^d) \Big\|.$$

Triangle inequality yields

$$\|f - I_{\Delta_{\alpha,\beta}(M)} f | H^1([0,1]^d)\| \le \sum_{\boldsymbol{j} \notin \Delta_{\alpha,\beta}(M)} \Big\| \sum_{\boldsymbol{k} \in D_{\boldsymbol{j}}} d_{\boldsymbol{j},\boldsymbol{k}}(f) v_{\boldsymbol{j},\boldsymbol{k}} \Big| H^1([0,1]^d) \Big\|.$$

Indeed, for fixed $\boldsymbol{j} \in \mathbb{N}_{-1}^d$ we easily check that

$$\Big\| \sum_{\boldsymbol{k} \in D_{\boldsymbol{j}}} d_{\boldsymbol{j},\boldsymbol{k}}(f) v_{\boldsymbol{j},\boldsymbol{k}} \Big| H^1([0,1]^d) \Big\|^2 \lesssim \sum_{\boldsymbol{k} \in D_{\boldsymbol{j}}} |d_{\boldsymbol{j},\boldsymbol{k}}(f)|^2 \|v_{\boldsymbol{j},\boldsymbol{k}} | H^1([0,1]^d)\|^2.$$

holds (the finite overlap of directions i_0 with $j_{i_0} = -1$ causes no problems). According to Lemma 3.14, we have

$$\|v_{\boldsymbol{j},\boldsymbol{k}} | H^1([0,1]^d)\| \asymp \|v_{\boldsymbol{j},\boldsymbol{k}} | L_2([0,1]^d)\| + \sum_{i=1}^d \Big\| \frac{\partial}{\partial x_i} v_{\boldsymbol{j},\boldsymbol{k}} \Big| L_2([0,1]^d) \Big\|.$$

Obviously,

$$\|v_{\boldsymbol{j},\boldsymbol{k}}|L_2([0,1]^d)\| \asymp 2^{-\frac{|\boldsymbol{j}|_1}{2}}.$$

Similar elementary calculations as above yield

$$\Big\|\frac{\partial}{\partial x_i}v_{\boldsymbol{j},\boldsymbol{k}}\Big|L_2([0,1]^d)\Big\| \lesssim 2^{j_i-\frac{|\boldsymbol{j}|_1}{2}}.$$

Combining both estimates gives

$$\|v_{\boldsymbol{j},\boldsymbol{k}}|H^1([0,1]^d)\| \lesssim 2^{|\boldsymbol{j}|_\infty-\frac{|\boldsymbol{j}|_1}{2}}.$$

Inserting this and applying Hölder's inequality yields

$$\begin{aligned}
&\|f-I_{\Delta_{\alpha,\beta}(M)}f|H^1([0,1]^d)\| \\
\lesssim\ & \sum_{\boldsymbol{j}\notin\Delta_{\alpha,\beta}(M)} 2^{|\boldsymbol{j}|_\infty-\frac{|\boldsymbol{j}|_1}{2}}\Big(\sum_{\boldsymbol{k}\in D_{\boldsymbol{j}}}|d_{\boldsymbol{j},\boldsymbol{k}}(f)|^2\Big)^{\frac{1}{2}}. &(5.3.4)\\
\lesssim\ & \Big(\sum_{\boldsymbol{j}\notin\Delta_{\alpha,\beta}(M)} 2^{-2[(r-(\frac{1}{p}-\frac{1}{2})_+)|\boldsymbol{j}|_1-|\boldsymbol{j}|_\infty]}\Big)^{\frac{1}{2}}\Big(\sum_{\boldsymbol{j}\notin\Delta_{\alpha,\beta}(M)} 2^{2(r-\frac{1}{2}-(\frac{1}{p}-\frac{1}{2})_+)|\boldsymbol{j}|_1}\sum_{\boldsymbol{k}\in D_{\boldsymbol{j}}}|d_{\boldsymbol{j},\boldsymbol{k}}(f)|^2\Big)^{\frac{1}{2}}. &(5.3.5)
\end{aligned}$$

Inserting the estimate from Lemma C.23 gives

$$\|f-I_{\Delta_{\alpha,\beta}(M)}f|H^1([0,1]^d)\| \ \le\ 2^{-M}\Big(\sum_{\boldsymbol{j}\notin\Delta_{\alpha,\beta}(M)} 2^{2(r-\frac{1}{2}-(\frac{1}{p}-\frac{1}{2})_+)|\boldsymbol{j}|_1}\sum_{\boldsymbol{k}\in D_{\boldsymbol{j}}}|d_{\boldsymbol{j},\boldsymbol{k}}(f)|^2\Big)^{\frac{1}{2}}.$$

We apply Theorem 4.19 and obtain

$$\|f-I_{\Delta_{\alpha,\beta}(M)}f|H^1([0,1]^d)\| \ \lesssim\ 2^{-M}\|f|S_2^{r-(\frac{1}{p}-\frac{1}{2})_+}W([0,1]^d)\|.$$

In case $p=2$ we are done. In case $p>2$ we finish with the trivial embedding

$$S_p^rW([0,1]^d) \hookrightarrow S_2^rW([0,1]^d).$$

In case $p<2$ we apply Lemma 3.4, (vi) (diagonal embedding) that yields

$$\|f-I_{\Delta_{\alpha,\beta}(M)}f|H^1([0,1]^d)\| \lesssim 2^{-M}\|f|S_p^rW([0,1]^d)\|.$$

□

Remark 5.19. *The parameter ε in Theorem 5.18 can be interpreted as a degree of freedom. Its explicit choice influences the constant C_ε and in the other way around the constant for the number of sampling nodes used by $I_{\Delta_{\alpha,\beta}}(m)$ according to Lemma C.22.*

Finally we state a result dealing with $r=2+(\frac{1}{p}-\frac{1}{2})_+$. This result was originally obtained in [8, Theorem 3.8] for $p=2$. Nevertheless the arguments there seem to contain a problematic step. We provide an alternative proof using Faber-Schauder sampling representations.

Theorem 5.20. *There exists a constant $C_\varepsilon > 0$ (independent of f and M) such that*

$$\|f - I_{\Delta_{\alpha,\beta}(M)} f | H^1([0,1]^d)\| \leq C_\varepsilon 2^{-M} \|f | S_2^{2+(\frac{1}{p}-\frac{1}{2})_+} W([0,1]^d)\| \tag{5.3.6}$$

holds with

$$\alpha = 2 - \left(\frac{1}{p} - \frac{1}{2}\right)_+ - \varepsilon \quad \textit{and} \quad \beta = 1 - \varepsilon$$

where

$$0 < \varepsilon < 1.$$

Proof. We proceed similar as in the proof of Theorem 5.18 and obtain the equivalent formulation of (5.3.4)

$$\|f - I_{\Delta_{\alpha,\beta}(M)} f | H^1([0,1]^d)\| \lesssim \sum_{\boldsymbol{j} \notin \Delta_{\alpha,\beta}(M)} 2^{|\boldsymbol{j}|_\infty} \Big\| \sum_{\boldsymbol{k} \in D_{\boldsymbol{j}}} d_{\boldsymbol{j},\boldsymbol{k}}(f) \chi_{\boldsymbol{j},\boldsymbol{k}} \Big| L_2([0,1]^d) \Big\|.$$

This can be estimated by

$$\begin{aligned} &\|f - I_{\Delta_{\alpha,\beta}(M)} f | H^1([0,1]^d)\| \\ \lesssim &\sup_{\boldsymbol{j} \notin \Delta_{\alpha,\beta}(M)} 2^{2|\boldsymbol{j}|_1} \Big\| \sum_{\boldsymbol{k} \in D_{\boldsymbol{j}}} d_{\boldsymbol{j},\boldsymbol{k}}(f) \chi_{\boldsymbol{j},\boldsymbol{k}} \Big| L_2([0,1]^d) \Big\| \sum_{\boldsymbol{j} \notin \Delta_{\alpha,\beta}(M)} 2^{-(2|\boldsymbol{j}|_1 - |\boldsymbol{j}|_\infty)}. \end{aligned}$$

We apply Theorem 4.30 and obtain

$$\|f - I_{\Delta_{\alpha,\beta}(M)} f | H^1([0,1]^d)\| \lesssim \|f | S_2^2 W([0,1]^d)\| \sum_{\boldsymbol{j} \notin \Delta_{\alpha,\beta}(M)} 2^{-(2|\boldsymbol{j}|_1 - |\boldsymbol{j}|_\infty)}.$$

The estimate for the sum in Lemma (C.23) gives

$$\|f - I_{\Delta_{\alpha,\beta}(M)} f | H^1([0,1]^d)\| \lesssim 2^{-M} \|f | S_2^2 W([0,1]^d)\|.$$

In case $p = 2$ we are done. In case $p > 2$ we finish with the trivial embedding

$$S_p^2 W([0,1]^d) \hookrightarrow S_2^2 W([0,1]^d).$$

In case $p < 2$ we apply Lemma 3.4, (vi) (diagonal embedding) that yields

$$\|f - I_{\Delta_{\alpha,\beta}(M)} f | H^1([0,1]^d)\| \lesssim 2^{-M} \|f | S_p^{2+\frac{1}{p}-\frac{1}{2}} W([0,1]^d)\|.$$

That finishes the proof. □

5.4 Optimality for standard information

The dependence on the smoothness of an energy sparse grid makes it to a very specific and non-general point set. Therefore, it does not seems to be useful to consider a benchmark quantity similar to (5.2.1). We consider a more general quantity allowing arbitrary point sets. This quantity is defined as

$$\varrho_n(S^r_pW([0,1]^d), H^1([0,1]^d)) := \inf_{\substack{X_n\subset[0,1]^d, |X_n|=n \\ \varphi:\mathbb{C}^m\to H^1([0,1]^d)}} \sup_{\|f|S^r_pW([0,1]^d)\|\le 1} \|f-\varphi(f(X_n))|H^1([0,1]^d)\|,$$

which we call worst case error for standard information (sampling width). It describes the $H^1([0,1]^d)$-best worst-case error for the approximation of functions in the unit ball of $S^r_pW([0,1]^d)$ by algorithms that can be expressed as a composition of a non-linear reconstruction map φ with vector of samples, where the sampling nodes are fixed. A simpler quantity to measure the performance of linear sampling algorithms is the *linear sampling width*,

$$\varrho^{\text{lin}}_n(S^r_pW([0,1]^d), H^1([0,1]^d)) := \inf_{X_n}\inf_{\Psi_n} \sup_{\|f|S^r_pW([0,1]^d)\|\le 1} \left\| f - \sum_{k=1}^{n} f(\boldsymbol{x}_k)\psi_k(\cdot) \middle| H^1([0,1]^d)\right\|, \tag{5.4.1}$$

$n\in\mathbb{N}$, where the sampling nodes $X_n := \{\boldsymbol{x}_k\}_{k=1}^n \subset [0,1]^d$ and associated (continuous) functions $\Psi_n := \{\psi_k\}_{k=1}^n$ determine a linear sampling recovery algorithm which is fixed in advance for the class $S^r_pW([0,1]^d)$. Let us emphasize that in (5.4.1) we restrict to *linear* recovery algorithms, whereas we admit general recovery algorithms $\varphi:\mathbb{C}^n\to L_q$ in (1.4.1).

Remark 5.21. *(i) Obviously,*

$$\varrho_n(S^r_pW([0,1]^d), H^1([0,1]^d)) \le \varrho^{lin}_n(S^r_pW([0,1]^d), H^1([0,1]^d)).$$

(ii) Similar arguments as in Lemma 5.11 yield that a lower bound for ϱ_n is provided by

$$\inf_{(x_k)_{k=1}^n\subset[0,1]^d} \sup_{\substack{\|f|S^r_pW([0,1]^d)\|\le 1, \\ f(x_k)=0,\, k\in[n])}} \|f|H^1([0,1]^d)\| \lesssim \varrho_n(S^r_pW([0,1]^d), H^1([0,1]^d)).$$

Theorem 5.22. *Let $1+(\frac{1}{p}-\frac{1}{2})_+ < r \le 2+(\frac{1}{p}-\frac{1}{2})_+$ and $1<p<\infty$. Then it holds*

$$\begin{aligned}
\varrho_n(S^r_pW([0,1]^d), H^1([0,1]^d)) &\asymp \varrho^{lin}_n(S^r_pW([0,1]^d), H^1([0,1]^d)) \\
&\asymp \sup_{\|f|S^r_pW([0,1]^d)\|\le 1} \|f - I_{\Delta_{\alpha,\beta}(M)}f|H^1([0,1]^d)\| \\
&\asymp n^{-(r-1-(\frac{1}{p}-\frac{1}{2})_+)}
\end{aligned}$$

with

$$\alpha = r - \left(\frac{1}{p}-\frac{1}{2}\right)_+ - \varepsilon, \quad \beta = 1-\varepsilon \quad \textit{and} \quad 0<\varepsilon<1.$$

and

$$n \asymp \operatorname{rank} I_{\Delta_{\alpha,\beta}(M)} \asymp |\mathcal{G}^{energy}_{\Delta_{\alpha,\beta}(M)}|.$$

Proof. The upper bound follows from Theorem 5.18 and the trivial inequality $\varrho_n \leq \varrho_n^{\mathrm{lin}}$ (limiting case: Theorem 5.20). According to Remark 5.21 a lower bound can be proven by constructing for every arbitrary point set of size n a fooling function that vanishes in all this sampling nodes. For n given sampling nodes $X = (\boldsymbol{x}_k)_{k=1}^n \subset [0,1]^d$ we find $\boldsymbol{j}^* \in \mathbb{N}_{-1}^d$ with

$$2^{|\boldsymbol{j}^*|_1} = 2^{|\boldsymbol{j}^*|_\infty} \asymp 2n. \tag{5.4.2}$$

Since we have $C2n$ translations in $D_{\boldsymbol{j}^*}$ and only Cn sampling nodes we find a set of translation indices $T_{\boldsymbol{j}^*}(X)$ such that

$$\{\boldsymbol{x} \in [0,1]^d : b_{\boldsymbol{j}^*,\boldsymbol{k}}(\boldsymbol{x}) \neq 0\} \cap \{x_i\} = \emptyset$$

for all $i = 1, \ldots, n$ and $\boldsymbol{k} \in T_{\boldsymbol{j}^*}(X)$. We have to distinguish two different cases. We start considering the case $p < q$. Here we consider the fooling function

$$f_1 = b_{\boldsymbol{j}^*,\boldsymbol{k}^*}$$

where $\boldsymbol{k}^* \in T_{\boldsymbol{j}*}(X)$. Theorem 4.34 yields

$$\|f_1|S_p^rW([0,1]^d)\| \lesssim 2^{(r-\frac{1}{p})|\boldsymbol{j}^*|_1}.$$

We stress on the equivalent norm

$$\|f|H^1([0,1]^d)\| \asymp \Big(\sum_{|\boldsymbol{\alpha}|_1 \leq 1} \|D^{\boldsymbol{\alpha}} f|L_2([0,1]^d)\|^2\Big)^{\frac{1}{2}},$$

and observe

$$\Big\|\frac{d}{dx} b_{j,k}\Big|L_2([0,1])\Big\| = 2^j \Big\|b'(2^jx - k)\Big|L_2([0,1])\Big\| = C2^{\frac{j}{2}}. \tag{5.4.3}$$

Then finally Fubini's Theorem yields

$$\|D^{\boldsymbol{\alpha}} b_{\boldsymbol{j},\boldsymbol{k}}|L_2([0,1]^d)\| = \prod_{i=1}^d \Big\|\frac{\partial^{\alpha_i}}{\partial x_i^{\alpha_i}} b_{j_i,k_i}\Big|L_2([0,1])\Big\|. \tag{5.4.4}$$

Using the identity in (5.4.2) we obtain for the single tensor bump function

$$\|b_{\boldsymbol{j}^*,\boldsymbol{k}^*}|H^1([0,1]^d)\| \gtrsim 2^{\frac{|\boldsymbol{j}^*|_1}{2}}.$$

That yields

$$\frac{\|f_1|H^1([0,1]^d)\|}{\|f_1|S_p^rW([0,1]^d)\|} \gtrsim 2^{-(r-1-(\frac{1}{p}-\frac{1}{2}))|\boldsymbol{j}^*|_1} \asymp n^{-(r-1-(\frac{1}{p}-\frac{1}{2}))}.$$

In case $p > q$ we consider the fooling function

$$f_2 = \sum_{\boldsymbol{k} \in T_{\boldsymbol{j}^*}(X)} b_{\boldsymbol{j}^*,\boldsymbol{k}}.$$

Here, due to disjoint supports for different $\boldsymbol{k}$ Theorem 4.34 yields

$$\begin{aligned}\|f_2|S_p^rW([0,1]^d)\| &\lesssim 2^{r|\boldsymbol{j}^*|_1}\Big\|\sum_{\boldsymbol{k}\in T_{\boldsymbol{j}^*}(X)}|b_{\boldsymbol{j}^*,\boldsymbol{k}}|\Big|L_p([0,1]^d)\Big\| \\ &= 2^{r|\boldsymbol{j}^*|_1}\Big(\sum_{\boldsymbol{k}\in T_{\boldsymbol{j}^*}(X)}\|b_{\boldsymbol{j}^*,\boldsymbol{k}}|L_p([0,1]^d)\|^p\Big)^{\frac{1}{p}} \\ &\asymp 2^{r|\boldsymbol{j}*|_1}.\end{aligned}$$

Additionally, we estimate

$$\begin{aligned}\|f_2|H^1([0,1]^d)\| &\asymp \Big(\sum_{|\boldsymbol{\alpha}|_1\le 1}\|D^{\boldsymbol{\alpha}}f_2|L_2([0,1]^d)\|^2\Big)^{\frac{1}{2}} \\ &\asymp \Big(\sum_{|\boldsymbol{\alpha}|_1\le 1}\sum_{\boldsymbol{k}\in T_{\boldsymbol{j}^*}(X)}\|D^{\boldsymbol{\alpha}}b_{\boldsymbol{j}^*,\boldsymbol{k}}|L_2([0,1]^d)\|^2\Big)^{\frac{1}{2}} \\ &\gtrsim \Big(\sum_{|\boldsymbol{\alpha}|_1=1}\sum_{\boldsymbol{k}\in T_{\boldsymbol{j}^*}(X)}\|D^{\boldsymbol{\alpha}}b_{\boldsymbol{j}^*,\boldsymbol{k}}|L_2([0,1]^d)\|^2\Big)^{\frac{1}{2}}. \qquad (5.4.5)\end{aligned}$$

Inserting (5.4.3) together with (5.4.4) into (5.4.5) provides

$$\|f_2|H^1([0,1]^d)\| \gtrsim 2^{|\boldsymbol{j}^*|_1}. \qquad (5.4.6)$$

Altogether, we obtain

$$\frac{\|f_2|H^1([0,1]^d)\|}{\|f_2|S_p^rW([0,1]^d)\|} \gtrsim 2^{-(r-1)|\boldsymbol{j}^*|_1} \asymp n^{-(r-1)}.$$

This concludes the proof. □

Remark 5.23. *Theorem 5.22 shows that energy sparse grid sampling provides the optimal asymptotic rate in the sense of the worst case error for standard information (sampling width).*

Chapter 6

Best m-term approximation with respect to the Faber-Schauder dictionary

In this chapter we study a concept of nonlinear approximation, so called best m-term approximation with respect to the Faber-Schauder dictionary. Let X, Y be quasi-Banach spaces and $\mathcal{D} \subset Y$ be a countable set called dictionary. For $x \in X$ we define its best m-term approximation by

$$\sigma_m(x, \mathcal{D})_Y := \inf_{(c_i)_{i=1}^m \subset \mathbb{C}, (b_i)_{i=1}^m \subset \mathcal{D}} \Big\| x - \sum_{i=1}^m c_i b_i \Big\|_Y .$$

For the space X we define the best m-term approximation with respect to the dictionary $\mathcal{D}$ by

$$\sigma_m(X, \mathcal{D})_Y := \sup_{\|x\|_X \le 1} \sigma_m(x, \mathcal{D})_Y .$$

Let $T : X \to Y$ be a linear operator. Then we define the the best m-term approximation of T by

$$\sigma_m(T : X \to Y, \mathcal{D}) := \sup_{\|x\|_X \le 1} \sigma_m(Tx, \mathcal{D})_Y .$$

6.1 Properties of best m-term widths $\sigma_m(T, \mathcal{D})$

We use the notation

$$\Sigma_m(\mathcal{D}) := \Big\{ \sum_{i=1}^m \lambda_i a_i : \lambda_i \in \mathbb{C}, a_i \in \mathcal{D}, i = 1, \ldots, m \Big\}$$

for the set of all m-terms in $\mathcal{D}$ and start with the following lemma proving some elementary properties of $\sigma_m(T, \mathcal{D})$ that we call pseudo s-number properties, cf. [90, p. 74].

Lemma 6.1. *Let W, X, Y, Z be ν-Banach spaces ($0 < \nu \leq 1$) and $\mathcal{D} \subset Y$ be a dictionary.*

(i) For $T \in \mathcal{L}(X, Y)$ we have

$$\|T|X \to Y\| = \sigma_0(T : X \to Y, \mathcal{D}) \geq \sigma_1(T : X \to Y, \mathcal{D}) \geq \sigma_2(T : X \to Y, \mathcal{D}) \geq \dots .$$

(ii) For $T_1, T_2, \dots, T_n \in \mathcal{L}(X, Y)$ and $m_1, m_2, \dots, m_n \in \mathbb{N}$ such that $m = \sum_{i=1}^n m_i$ we have

$$\sigma_m\Big(\sum_{i=1}^n T_i, \mathcal{D}\Big)^\nu \leq \sum_{i=1}^n \sigma_{m_i}(T_i, \mathcal{D})^\nu.$$

(iii) For $T \in \mathcal{L}(Z, Y)$, $A \in \mathcal{L}(X, Z)$, $B \in \mathcal{L}(Y, W)$ we have

$$\sigma_m(BTA, B(\mathcal{D})) \leq \|B\| \sigma_m(T, \mathcal{D}) \|A\|$$

Proof. *(i)* is obvious by definition. *(ii)* We prove the case $n = 2$ in detail. $n > 2$ follows by iterating the arguments. Let $x \in X$ with $\|x|X\| \leq 1$. Then for arbitrary $(c_i)_{i=1}^m \subset \mathbb{C}$ and $(b_i)_{i=1}^m \subset \mathcal{D}$ we obtain

$$\begin{aligned} \sigma_m((T_1 + T_2)x, \mathcal{D})_Y^\nu &\leq \Big\|(T_1 + T_2)x - \sum_{i=1}^m c_i b_i \Big| Y\Big\|^\nu \\ &\leq \Big\|T_1 x - \sum_{i=1}^{m_1} c_i b_i \Big| Y\Big\|^\nu + \Big\|T_2 x - \sum_{i=m_1+1}^{m_2} c_i b_i \Big| Y\Big\|^\nu . \end{aligned}$$

Taking infimum over $(c_i)_{i=1}^m \subset \mathbb{C}, (b_i)_{i=1}^m \subset \mathcal{D}$ and supremum over all $x \in X$ with $\|x|X\| \leq 1$ finishes the proof of *(ii)*. We consider *(iii)*. By definition we find $g \in \Sigma_m(\mathcal{D})$ with $g' = B(g)$ for every $g' \in \Sigma_m(B(\mathcal{D}))$ such that

$$\begin{aligned} \sigma_m(BTA, B(\mathcal{D})) &= \sup_{\|x\|_X \leq 1} \inf_{g' \in \Sigma_m(B(D))} \|BTAx - B(g)\|_W \\ &\leq \|B : Y \to W\| \sup_{\|x\|_X \leq 1} \inf_{g' \in \Sigma_m(\mathcal{D})} \|TAx - g'\|_Y \\ &= \|B\| \sigma_m(TA, \mathcal{D}). \end{aligned}$$

We continue estimating

$$\begin{aligned} \sigma_m(TA, \mathcal{D}) &= \sup_{\|x\|_X \leq 1} \inf_{g' \in \Sigma_m(\mathcal{D})} \|TAx - g'\|_Y \\ &\leq \sup_{\|z\|_Z \leq \|A\|} \inf_{g' \in \Sigma_m(\mathcal{D})} \|Tz - g'\|_Y \\ &= \sup_{\|z\| \leq 1} \inf_{g' \in \Sigma_m(\mathcal{D})} \|\|A\| T(z) - g'\|_Y \\ &= \sup_{\|z\| \leq 1} \inf_{g'' \in \Sigma_m(\mathcal{D})} \|\|A\| T(z) - \|A\| g''\|_Y \\ &= \|A : X \to Z\| \sigma_m(T, \mathcal{D}). \end{aligned}$$

□

6.2 Sparse approximation in (vector-valued) sequence spaces

Let us now discuss specific situations. The following lemma is well-known and usually referred as Stechkin's lemma. For our knowledge the first reference for the stated generality is [112, Lemma 2.1, p. 97], see also [33, Section 7.4] and the references given there.

Lemma 6.2. *Let $0 < p < q \leq \infty$ and $(a_n)_{n\in\mathbb{N}} \subset \mathbb{C}$ be a sequence of complex numbers with the property*

$$|a_1| \geq |a_2| \geq |a_3| \geq \dots$$

Then

$$\Big(\sum_{n=m+1}^{\infty} |a_n|^q\Big)^{\frac{1}{q}} \leq (m+1)^{-(\frac{1}{p}-\frac{1}{q})}\Big(\sum_{n=1}^{\infty}|a_n|^p\Big)^{\frac{1}{p}} \tag{6.2.1}$$

holds for all $m \in \mathbb{N}_0$. As usual, for $q = \infty$ the sum on the left hand side is replaced by a supremum.

Let us now turn to the vector-valued situation. Here we have X_μ, Y_μ quasi-Banach (p-Banach) spaces, $T_\mu \in \mathcal{L}(X_\mu, Y_\mu)$ linear operators and I be an index set. Let $\mathcal{D}_\mu$ denote a dictionary in Y_μ and

$$\mathcal{D} := \bigcup_{\mu\in I}\ \bigcup_{e_\mu\in\mathcal{D}_\mu} \{(0,\dots,0,e_\mu,0,\dots,0)\}.$$

Definition 6.3. *Let I be an index set and $(X_\mu)_{\mu\in I}$ be a sequence of quasi-Banach (q-Banach) spaces. We define the following sequence space*

$$\ell_p(X_\mu, I) = \Big\{x = (x_\mu)_{\mu\in I} : x_\mu \in X_\mu, \|x\|_{\ell_p(X_\mu,I)} = \Big(\sum_{\mu\in I}\|x_\mu|X_\mu\|^p\Big)^{\frac{1}{p}} < \infty\Big\}$$

with the usual modifications in case $p = \infty$.

Theorem 6.4. *Let $0 < p < q \leq \infty$ and $T = (T_\mu)_{\mu\in I} : \ell_p(X_\mu, I) \to \ell_q(Y_\mu, I)$. Then*

$$\sigma_m(T,\mathcal{D}) \leq \sup_{\mu\in I}\ \sup_{0\leq s\leq m}\Big(\frac{s+1}{m+1}\Big)^{\frac{1}{p}-\frac{1}{q}}\sigma_s(T_\mu,\mathcal{D}_\mu).$$

Proof. We have $T = (T_\mu)_{\mu\in I}$ with $T_\mu : X_\mu \to Y_\mu$. Let m be given. Define $m_\mu = \lfloor (m+1)\|x_\mu|X_\mu\|^p\rfloor$ for some $x = (x_\mu)_{\mu\in I}$ with $\|x|\ell_p(X_\mu)\| < 1$. Using m_μ-approximation in the component $T_\mu x_\mu$ we obtain the relation

$$\sigma_m(Tx,\mathcal{D})_{\ell_q(Y_\mu)} \leq \Big(\sum_{\mu\in I}\sigma_{m_\mu}(T_\mu x_\mu,\mathcal{D}_\mu)^q_{Y_\mu}\Big)^{\frac{1}{q}},$$

since

$$\sum_{\mu\in I} m_\mu \leq (m+1)\sum_{\mu\in I}\|x_\mu|X_\mu\|^p < m+1.$$

We proceed estimating as follows

$$\begin{aligned}\sigma_m(Tx,\mathcal{D})_{\ell_q(Y_\mu)} &\leq \Big(\sum_{\mu\in I}\Big(\frac{(m+1)\|x_\mu\|^p}{m+1}\Big)^{q(\frac{1}{p}-\frac{1}{q})}\Big(\frac{m+1}{(m+1)\|x_\mu|X_\mu\|^p}\Big)^{q(\frac{1}{p}-\frac{1}{q})} \\ &\quad \sigma_{m_\mu}(T_\mu,\mathcal{D}_\mu)^q\|x_\mu|X_\mu\|^q\Big)^{\frac{1}{q}} \\ &\leq \Big(\sum_{\mu\in I}\Big(\frac{m_\mu+1}{m+1}\Big)^{q(\frac{1}{p}-\frac{1}{q})}\Big(\frac{m+1}{(m+1)\|x_\mu|X_\mu\|^p}\Big)^{q(\frac{1}{p}-\frac{1}{q})} \\ &\quad \sigma_{m_\mu}(T_\mu,\mathcal{D}_\mu)^q\|x_\mu|X_\mu\|^q\Big)^{\frac{1}{q}}.\end{aligned}$$

Note, that in case $m_\mu = 0$ the respective summand will be replaced by $\|T_\mu\|^q\|x_\mu|X_\mu\|^q = \sigma_0(T_\mu,\mathcal{D}_\mu)^q\|x_\mu|X_\mu\|^q$. We obtain

$$\begin{aligned}\sigma_m(Tx,\mathcal{D})_{\ell_q(Y_\mu)} &\leq \sup_{\mu\in I}\sup_{0\leq s\leq m}\Big(\frac{s+1}{m+1}\Big)^{\frac{1}{p}-\frac{1}{q}}\sigma_s(T_\mu,\mathcal{D}_\mu)\Big(\sum_{\mu\in I}\|x_\mu|X_\mu\|^p\Big)^{\frac{1}{q}} \quad (6.2.2)\\ &\leq \sup_{\mu\in I}\sup_{0\leq s\leq m}\Big(\frac{s+1}{m+1}\Big)^{\frac{1}{p}-\frac{1}{q}}\sigma_s(T_\mu,\mathcal{D}_\mu).\end{aligned}$$

In case $m_\mu = 0$ we simply write

$$\begin{aligned}\|x_\mu|X_\mu\|^q &= \|x_\mu|X_\mu\|^{q-p}\|x_\mu|X_\mu\|^p \\ &= (\|x_\mu|X_\mu\|^p)^{\frac{q-p}{p}}\|x_\mu|X_\mu\|^p \\ &\leq \Big(\frac{1}{m+1}\Big)^{\frac{q-p}{p}}\|x_\mu|X_\mu\|^p,\end{aligned}$$

since because of $m_\mu < 1$ we have $\|x_\mu|X_\mu\|^p < \frac{1}{m+1}$. The result above holds for $\|x|X\| < 1$. It remains to consider the case $\|x|X\| = 1$. Let $x \in X$ with $\|x|X\| = 1$ and additionally $\lambda > 1$. We use a limiting argument together with (6.2.2). Obviously $\big\|\frac{x}{\lambda}|X\big\| < 1$. For that reason we obtain

$$\begin{aligned}\sigma_m(Tx,\mathcal{D})_{\ell_q(Y_\mu)} &= \lambda\sigma_m\Big(Tx/\lambda,\mathcal{D}\Big)_{\ell_q(Y_\mu)} \quad (6.2.3)\\ &\leq \lambda\sup_{\mu\in I}\sup_{0\leq s\leq m}\Big(\frac{s+1}{m+1}\Big)^{\frac{1}{p}-\frac{1}{q}}\sigma_s(T_\mu,\mathcal{D}_\mu)\Big(\sum_{\mu\in I}\Big\|\frac{x_\mu}{\lambda}\Big|X_\mu\Big\|^p\Big)^{\frac{1}{q}} \\ &\leq \lambda^{1-\frac{p}{q}}\sup_{\mu\in I}\sup_{0\leq s\leq m}\Big(\frac{s+1}{m+1}\Big)^{\frac{1}{p}-\frac{1}{q}}\sigma_s(T_\mu,\mathcal{D}_\mu) \quad (6.2.4)\end{aligned}$$

This holds for all $\lambda > 1$ arbitrary close to 1. We obtain

$$\sigma_m(Tx,\mathcal{D})_{\ell_q(Y_\mu)} \leq \sup_{\mu\in I}\sup_{0\leq s\leq m}\Big(\frac{s+1}{m+1}\Big)^{\frac{1}{p}-\frac{1}{q}}\sigma_s(T_\mu,\mathcal{D}_\mu).$$

Finally, taking the supremum over $\|x|\ell_p(X_\mu)\| \leq 1$ on both sides proves the theorem. □

This theorem has some consequences. We consider some special cases and start with Lemma 6.2. Choosing $X_\mu = Y_\mu = \mathbb{C}$ we can prove a similar result having the same convergence rate as in (6.2.1) immediately by applying Theorem 6.4. This gives us a slightly different selection procedure for the m terms in the m-term approximation.

Corollary 6.5. *Let* $0 < p < q \leq \infty$ *and*

$$A(m,a) := \{i \in \mathbb{N} : m_i := |a_i|^p(m+1) \geq 1\}.$$

Then we have

$$\sup_{\|a|\ell_p\|<1} \Big\| a - \sum_{i\in A(m,a)} a_i e_i \Big| \ell_q \Big\| \leq \Big(\frac{1}{m+1}\Big)^{\frac{1}{p}-\frac{1}{q}}$$

which implies

$$\sigma_m(\ell_p, \mathbb{E})_{\ell_q} \leq \Big(\frac{1}{m+1}\Big)^{\frac{1}{p}-\frac{1}{q}}. \tag{6.2.5}$$

Proof. Choosing $X_\mu = Y_\mu = \mathbb{C}$ we can prove the upper bound in (6.2.1) immediately by applying Theorem 6.4. Let $a \in \ell_p$ with $\|a|\ell_p\| < 1$. Then $\sum_{i\in A(m,a)} a_i e_i$ is a m-term approximation of a, since

$$|A(m,a)| \leq \sum_{i=1}^{\infty} \lfloor m_i \rfloor \leq (m+1)\sum_{i=1}^{\infty} |a_i|^p = (m+1)\|a|\ell_p\|^p < m+1.$$

The arguments provided in (6.2.4) yield the case

$$\|a|\ell_p\| = 1.$$

This gives (6.2.5). □

Remark 6.6. *The case $\|a|\ell_p\| = 1$ is based on a limiting argument. The above algorithm may not work in case $\|a|\ell_p\| = 1$. Replacing the definition of m_i by $m_i = |a_i|^p m$ in 6.5 and accepting a constant $C \geq 1$ in the approximation rates then we obtain an explicit algorithm that generates a m-term approximation for the case $\|a|\ell_p\| = 1$. Similarly m_μ in Theorem 6.4 can be replaced by $m_\mu = \|x_\mu|X_\mu\|^p m$. This gives us a more transparent approximation strategy. The price to pay is constant $C \geq 1$.*

Next we consider $T = (T, \ldots, T)$ with $T = id$, $X_\mu = \ell_u^d$, $Y_\mu = \ell_r^d$, so $X_\mu = X$, $Y_\mu = Y$ independent of μ, $u < r$. The next corollary generalizes [54, Theorem 4].

Corollary 6.7. *Let $u < r$, $p < q$ and $0 < \frac{1}{u} - \frac{1}{r} \leq \frac{1}{p} - \frac{1}{q}$. Let further $T = (T_0, \ldots, T_0) = id$ and $b, d \in \mathbb{N}$. Then we have for the dictionary $\mathcal{D} = (e_{i,j})_{i\in[b], j\in[d]}$ of unit vectors*

$$\sigma_m(id : \ell_p^b(\ell_u^d) \to \ell_q^b(\ell_r^d), \mathcal{D}) \leq \begin{cases} \Big(\frac{1}{m+1}\Big)^{\frac{1}{u}-\frac{1}{r}} & : \quad 1 \leq m < d, \\ \Big(\frac{d}{m+1}\Big)^{\frac{1}{p}-\frac{1}{q}} \Big(\frac{1}{d}\Big)^{\frac{1}{u}-\frac{1}{r}} & : \quad d \leq m < bd, \\ 0 & : \quad m \geq bd. \end{cases}$$

Proof. The case $m \geq bd$ is trivial, since $\dim \ell_p^b(\ell_u^d) = bd$. We consider the remaining cases. We apply Theorem 6.4. This gives

$$\sigma_m(T, \mathcal{D}) \leq \sup_{i=1,\dots,b} \sup_{0 \leq s \leq m} \left(\frac{s+1}{m+1}\right)^{\frac{1}{p}-\frac{1}{q}} \sigma_s(id : \ell_u^d \to \ell_r^d, \mathcal{D}).$$

By Lemma 6.4 we have

$$\sigma_s(id : \ell_u^d \to \ell_r^d, \mathcal{D}) \leq (s+1)^{-(\frac{1}{u}-\frac{1}{r})}$$

if $1 \leq s \leq d$ and

$$\sigma_s(id : \ell_u^d \to \ell_r^d, \mathcal{D}) = 0,$$

otherwise. Hence we have for $m \leq d$

$$\begin{aligned} \sigma_m(T, \mathcal{D}) &\leq \sup_{0 \leq s \leq m} \left(\frac{s+1}{m+1}\right)^{\frac{1}{p}-\frac{1}{q}} (s+1)^{-\left(\frac{1}{u}-\frac{1}{r}\right)} \\ &= (m+1)^{-(\frac{1}{p}-\frac{1}{q})} (m+1)^{(\frac{1}{p}-\frac{1}{q})-(\frac{1}{u}-\frac{1}{r})} \\ &= (m+1)^{-(\frac{1}{u}-\frac{1}{r})}. \end{aligned}$$

In case $d \leq m \leq db$ we have

$$\begin{aligned} \sigma_m(T, \mathcal{D}) &\leq \sup_{0 \leq s \leq d-1} \left(\frac{s+1}{m+1}\right)^{\frac{1}{p}-\frac{1}{q}} (d+1)^{-(\frac{1}{u}-\frac{1}{r})} \\ &= \left(\frac{d+1}{m+1}\right)^{\frac{1}{p}-\frac{1}{q}} (d+1)^{-(\frac{1}{u}-\frac{1}{r})}. \end{aligned}$$

This gives the corollary. □

Now we consider a more complicated situation. In a sense this represents a vector valued framework suitable for function space embeddings. Let

$$X_\mu = \ell_p^{b_\mu}(d_\mu^\alpha \ell_u^{d_\mu}),$$

$$Y_\mu = \ell_q^{b_\mu}(d_\mu^\beta \ell_r^{d_\mu}), \tag{6.2.6}$$

where d_μ, b_μ are natural numbers which are growing with μ and $\alpha, \beta \geq 0$. Let us now study

$$\sigma_m(id : \ell_p(X_\mu) \to \ell_q(Y_\mu), \mathcal{D}),$$

where

$$\|x\|_{\ell_p(X_\mu)} = \Big(\sum_{\mu \in I} \|x_\mu|X_\mu\|^p\Big)^{\frac{1}{p}},$$

$$\|y\|_{\ell_p(Y_\mu)} = \Big(\sum_{\mu \in I} \|y_\mu|Y_\mu\|^q\Big)^{\frac{1}{q}}.$$

We define $\mathcal{D}$ as the set of unit vectors in $\ell_p(X_\mu)$. We are interested in the following special case

$$\frac{1}{u} - \frac{1}{r} \le \frac{1}{p} - \frac{1}{q} \tag{6.2.7}$$

with

$$-\Big(\frac{1}{u} - \frac{1}{r}\Big) < \alpha - \beta \le \frac{1}{p} - \frac{1}{q} - \Big(\frac{1}{u} - \frac{1}{r}\Big). \tag{6.2.8}$$

Corollary 6.8. *Let $\alpha - \beta \le \frac{1}{p} - \frac{1}{q} - \Big(\frac{1}{u} - \frac{1}{r}\Big)$ with $\frac{1}{u} - \frac{1}{r} \le \frac{1}{p} - \frac{1}{q}$. Then we have*

$$\sigma_m(id : \ell_p(X_\mu, I) \to \ell_q(Y_\mu, I), \mathcal{D}) \le \Big(\frac{1}{m+1}\Big)^{(\alpha-\beta)+(\frac{1}{u}-\frac{1}{r})}$$

if $m+1 \ge \sup_{\mu \in I} d_\mu$. $I \subseteq \mathbb{N}_0^d$ denotes the index set of the outer sequence spaces.

Proof. Due to $m \ge \sup_{\mu\in I} d_\mu$ we have

$$\begin{aligned}
\sigma_m(id : \ell_p(X_\mu, I) \to \ell_q(Y_\mu, I)) &\le \sup_{\mu\in I}\sup_{0\le s\le m} \Big(\frac{s+1}{m+1}\Big)^{\frac{1}{p}-\frac{1}{q}} \sigma_s(id : X_\mu \to Y_\mu, \mathcal{D}_\mu) \\
&\le \sup_{\mu\in I}\sup_{0\le s< d_\mu} \Big(\frac{s+1}{m+1}\Big)^{\frac{1}{p}-\frac{1}{q}} \sigma_s(id : X_\mu \to Y_\mu, \mathcal{D}_\mu) \\
&\quad + \sup_{\mu\in I}\sup_{d_\mu\le s\le m} \Big(\frac{s+1}{m+1}\Big)^{\frac{1}{p}-\frac{1}{q}} \sigma_s(id : X_\mu \to Y_\mu, \mathcal{D}_\mu).
\end{aligned} \tag{6.2.9}$$

Inserting the result from Corollary 6.7 the first term can be estimated by

$$\begin{aligned}
&\sup_{\mu\in I}\sup_{0\le s< d_\mu} \Big(\frac{s+1}{m+1}\Big)^{\frac{1}{p}-\frac{1}{q}} \Big(\frac{1}{s+1}\Big)^{\frac{1}{u}-\frac{1}{r}} d_\mu^{-(\alpha-\beta)} \\
&= \Big(\frac{1}{m+1}\Big)^{\frac{1}{p}-\frac{1}{q}} d_\mu^{\frac{1}{p}-\frac{1}{q}-(\frac{1}{u}-\frac{1}{r})-(\alpha-\beta)}.
\end{aligned}$$

Taking the supremum with respect to μ we obtain

$$\sup_{\mu\in I}\sup_{0\le s\le d_\mu} \Big(\frac{s+1}{m+1}\Big)^{\frac{1}{p}-\frac{1}{q}} \sigma_s(id : X_\mu \to Y_\mu, \mathcal{D}_\mu) \le (m+1)^{-[(\alpha-\beta)+(\frac{1}{u}-\frac{1}{r})]}.$$

Finally estimating the second term in (6.2.9) gives

$$\begin{aligned}
&\sup_{\mu\in I}\sup_{d_\mu\le s\le m} \Big(\frac{s+1}{m+1}\Big)^{\frac{1}{p}-\frac{1}{q}} \sigma_s(id : X_\mu \to Y_\mu, \mathcal{D}_\mu) \\
\le\ &\sup_{\mu\in I}\sup_{d_\mu\le s\le m} \Big(\frac{s+1}{m+1}\Big)^{\frac{1}{p}-\frac{1}{q}} \Big(\frac{d_\mu}{s+1}\Big)^{\frac{1}{p}-\frac{1}{q}} \Big(\frac{1}{d_\mu}\Big)^{\frac{1}{u}-\frac{1}{r}} d_\mu^{-(\alpha-\beta)} \\
\le\ &\sup_{\mu\in I} \Big(\frac{1}{m+1}\Big)^{\frac{1}{p}-\frac{1}{q}} d_\mu^{\frac{1}{p}-\frac{1}{q}-(\frac{1}{u}-\frac{1}{r})-(\alpha-\beta)} \\
\le\ &(m+1)^{-(\alpha-\beta)-(\frac{1}{u}-\frac{1}{r})}.
\end{aligned}$$

□

Corollary 6.9. *Let $0 \leq \alpha - \beta$ with $\frac{1}{u} - \frac{1}{r} \leq \frac{1}{p} - \frac{1}{q}$. Then we have*

$$\sigma_m(id : \ell_p(X_\mu, I) \to \ell_q(Y_\mu, I), \mathcal{D}) \leq \Big(\frac{1}{m+1}\Big)^{(\alpha-\beta)+(\frac{1}{u}-\frac{1}{r})}$$

if $m + 1 \leq \inf_{\mu \in I} d_\mu$. $I \subset \mathbb{N}_0^d$ denotes the index set of the outer sequence spaces.

Proof. We fix $\mu \in I$. Due to $m + 1 < d_\mu$ Corollary 6.7 gives

$$\sup_{0\leq s\leq m} \Big(\frac{s+1}{m+1}\Big)^{\frac{1}{p}-\frac{1}{q}} \Big(\frac{1}{s+1}\Big)^{\frac{1}{u}-\frac{1}{r}} d_\mu^{-(\alpha-\beta)} \leq (m+1)^{-(\frac{1}{u}-\frac{1}{r})} d_\mu^{-(\alpha-\beta)} \leq \Big(\frac{1}{m+1}\Big)^{(\alpha-\beta)+(\frac{1}{u}-\frac{1}{r})}.$$

□

Corollary 6.10. *Let p, q, u, r as in (6.2.7), (6.2.8) and X_μ, Y_μ as in (6.2.6). If $\alpha - \beta \geq 0$ we have*

$$\sigma_m(id : \ell_p(X_\mu) \to \ell_q(Y_\mu), \mathcal{D}) \lesssim \Big(\frac{1}{m}\Big)^{(\alpha-\beta)+(\frac{1}{u}-\frac{1}{r})}$$

for all $m \in \mathbb{N}$.

Proof. The proof follows immediately by Corollary 6.8 and 6.9 using for $u = \min\{q, 1\}$ the decomposition

$$\begin{aligned}\sigma_{2m}(id : \ell_p(X_\mu) \to \ell_q(Y_\mu), \mathcal{D})^q &\leq \sigma_m(id : \ell_p(X_\mu, I_1) \to \ell_q(Y_\mu, I_1), \mathcal{D})^q \\ &\quad + \sigma_m(id : \ell_p(X_\mu, I_2) \to \ell_q(Y_\mu, I_2), \mathcal{D})^q\end{aligned}$$

where

$$I_1 := \{\mu \in \mathbb{N}_0^d : d_\mu \leq m+1\} \quad \text{and} \quad I_2 := \{\mu \in \mathbb{N}_0^d : d_\mu > m+1\}.$$

□

Next we consider best m-term approximation for discrete function spaces $s^{r,\Omega}_{p,\theta}f$ and $s^{r,\Omega}_{p,\theta}b$ with $\Omega = [0,1]^d$. We need some further notation. We introduce for $\mu \in \mathbb{N}_0$ the following sets and quantities

$$\begin{aligned} M(\mu,d) &:= \Big\{\boldsymbol{j} \in \mathbb{N}_{-1}^d : \sum_{i=1}^d \max\{j_i, 0\} = \mu\Big\}, \qquad (6.2.10)\\ S(\mu,d) &= |M(\mu,d)|,\\ \nabla_\mu &:= \{(\boldsymbol{j},\boldsymbol{k}) : \boldsymbol{j} \in M(\mu,d), \boldsymbol{k} \in D_{\boldsymbol{j}}\},\\ N(\mu,d) &:= |\nabla_\mu|.\end{aligned}$$

Definition 6.11. *We define the projection of the sequence $a := (a_{\boldsymbol{j},\boldsymbol{k}})_{(\boldsymbol{j},\boldsymbol{k})\in\nabla}$ to indices of the hyperbolic cross layer $M(\mu,d)$ by*

$$R_\mu a := \sum_{\boldsymbol{j}\in M(\mu,d)} \sum_{\boldsymbol{k}\in \boldsymbol{D}_{\boldsymbol{j}}} a_{\boldsymbol{j},\boldsymbol{k}} e_{\boldsymbol{j},\boldsymbol{k}},$$

where $e_{\boldsymbol{j},\boldsymbol{k}}$ are the unit vectors with index $(\boldsymbol{j},\boldsymbol{k})$. Such a projection fulfills the following properties.

Lemma 6.12. *Let $x \in \{b, f\}$, $0 < p < q < \infty$ ($q = \infty : x = b$,), $0 < \theta < \nu \leq \infty$ and $t \leq r$. Then the following inequalities hold*

(i)

$$\|R_\mu a|s^{r,\Omega}_{p,\theta}x\| \leq S(\mu,d)^{\frac{1}{\theta}-\frac{1}{\nu}}\|R_\mu a|s^{r,\Omega}_{p,\nu}x\|,$$

(ii)

$$\|R_\mu a|s^{r,\Omega}_{p,\theta}x\| \leq \|R_\mu a|s^{r,\Omega}_{q,\theta}x\|,$$

(iii)

$$\|R_\mu a|s^{r,\Omega}_{p,\theta}x\| \leq \|a|s^{r,\Omega}_{p,\theta}x\|.$$

(iv) Additionally the identity

$$id : s^{r,\Omega}_{p,\theta}x \to s^{t,\Omega}_{p,\theta}x = \sum_{\mu=0}^{\infty} R_\mu$$

holds.

Proof. (i) and (ii) can be proven using Hölder's inequality. The estimates in *(iii)* and *(iv)* are obvious. □

The case of large smoothness

The following result is due to Hansen and Sickel [56, Proposition 5.4]. We provide an alternative proof using pseudo s-number properties. $\mathcal{D}$ denotes the set of unit vectors in sequence spaces $s^{r,\Omega}_{p,\theta}f$.

Theorem 6.13. *Let $x, y \in \{b, f\}$, $0 < p, q \leq \infty$ ($p = \infty : x = b$, $q = \infty : y = b$) and $0 < \theta, \nu \leq \infty$. We denote by $\gamma_0 = \min\{p, \theta\}$ and $\delta_1 = \max\{\nu, q\}$. Further let $r, t \geq 0$ with $r - t > \max\left\{0, \frac{1}{\gamma_0} - \frac{1}{\delta_1}\right\}$ then*

$$\sigma_m(s^{r,\Omega}_{p,\theta}x, \mathcal{D})_{s^{t,\Omega}_{q,\nu}y} \asymp C(m^{-1}\log^{d-1} m)^{r-t}(\log^{d-1} m)^{\frac{1}{\nu}-\frac{1}{\theta}}. \tag{6.2.11}$$

Proof. For the lower bound we refer to [56]. We give a new proof for the upper bound. We start with the case $\gamma_0 < \delta_1$. We denote by D_μ the set of unit vectors in $s^{r,\Omega}_{p,\theta}f$ restricted to the hyperbolic layer with $|\boldsymbol{j}|_1 = \mu$. Let $a \in s^r_{p,\theta}x$ with $\|a|s^r_{p,\theta}x\| \leq 1$ then Lemma 6.1, (ii) with $u := \min\{q, \nu, 1\}$ provides the decomposition

$$\begin{aligned}\sigma_m(a, \mathcal{D})^u_{s^{t,\Omega}_{q,\nu}y} \leq & \sum_{\mu=0}^{M}\sigma_{m_\mu}(R_\mu a, \mathcal{D}_\mu)^u_{s^{t,\Omega}_{q,\nu}y} + \sum_{\mu=M+1}^{L}\sigma_{m_\mu}(R_\mu a, \mathcal{D}_\mu)^u_{s^{t,\Omega}_{q,\nu}y} \qquad (6.2.12)\\ & + \sum_{\mu=L+1}^{\infty}\sigma_{m_\mu}(R_\mu a, \mathcal{D}_\mu)^u_{s^{t,\Omega}_{q,\nu}y}, \qquad (6.2.13)\end{aligned}$$

where

$$m_\mu \asymp \begin{cases} 2^\mu \mu^{d-1} & : \quad 0 \le \mu \le M, \\ \lfloor 2^\mu 2^{(M-\mu)\kappa} \mu^{d-1} \rfloor & : \quad M+1 \le \mu \le L, \\ 0 & : \quad \text{otherwise}, \end{cases}$$

with $\kappa > 1$,

$$m \asymp 2^M M^{d-1}$$

and

$$L = \left\lceil \frac{M\kappa + (d-1)\log M - 1}{\kappa - 1} \right\rceil.$$

First show that $m \gtrsim \sum_{\mu=0}^{L} m_\mu$. Obviously

$$\sum_{\mu=0}^{\infty} m_\mu \lesssim \sum_{\mu=0}^{M} \mu^{d-1} 2^\mu + 2^{M\kappa} \sum_{\mu=M+1}^{L} 2^{(1-\kappa)\mu} \mu^{d-1} \lesssim M^{d-1} 2^M \asymp m.$$

The first sum in (6.2.13) vanishes, since $|\nabla_\mu| \asymp \mu^{d-1} 2^\mu$. So this part can be approximated exactly. We continue dealing with the second sum. Applying Lemma 6.12, (i) gives

$$\begin{aligned} \sigma_{m_\mu}(R_\mu a, \mathcal{D}_\mu)_{s^{t,\Omega}_{q,\nu} y} &\le S(\mu,d)^{\frac{1}{\nu} - \frac{1}{\delta_1}} \sigma_{m_\mu}(R_\mu a, \mathcal{D}_\mu)_{s^{t}_{\delta_1,\delta_1} b} \\ &\lesssim S(\mu,d)^{\frac{1}{\nu} - \frac{1}{\delta_1}} \sigma_{m_\mu}(R_\mu : s^{r,\Omega}_{\gamma_0,\gamma_0} b \to s^{t,\Omega}_{\delta_1,\delta_1} b, \mathcal{D}_\mu) \| R_\mu a | s^{r,\Omega}_{\gamma_0,\gamma_0} x \| \\ &\asymp S(\mu,d)^{\frac{1}{\nu} - \frac{1}{\delta_1}} 2^{-\mu(r-t-(\frac{1}{\gamma_0} - \frac{1}{\delta_1}))} \sigma_{m_\mu}(R_\mu : \ell^{\mu^{d-1}2^\mu}_{\gamma_0} \to \ell^{\mu^{d-1}2^\mu}_{\delta_1}, \mathcal{D}_\mu) \\ &\quad \| R_\mu a | s^{r,\Omega}_{\gamma_0,\gamma_0} x \|. \end{aligned}$$

Corollary 6.5 yields

$$\sigma_{m_\mu}(R_\mu a, \mathcal{D}_\mu)_{s^{t,\Omega}_{q,\nu} y} \lesssim S(\mu,d)^{\frac{1}{\nu} - \frac{1}{\delta_1}} 2^{-\mu(r-t-(\frac{1}{\gamma_0} - \frac{1}{\delta_1}))} m_\mu^{-(\frac{1}{\gamma_0} - \frac{1}{\delta_1})} \| R_\mu a | s^{r,\Omega}_{\gamma_0,\gamma_0} x \|.$$

Lemma 6.12 allows to estimate this by

$$\sigma_{m_\mu}(R_\mu a, \mathcal{D}_\mu)_{s^{t,\Omega}_{q,\nu} y} \lesssim S(\mu,d)^{\frac{1}{\nu} - \frac{1}{\delta_1} + \frac{1}{\gamma_0} - \frac{1}{\theta}} 2^{-\mu(r-t-(\frac{1}{\gamma_0} - \frac{1}{\delta_1}))} m_\mu^{-(\frac{1}{\gamma_0} - \frac{1}{\delta_1})} \| R_\mu a | s^{r,\Omega}_{p,\theta} x \|.$$

Choosing $\kappa > 1$ close to 1 such that $\kappa(\frac{1}{\gamma_0} - \frac{1}{\delta}_1) < r - t$ then summing up yields

$$\begin{aligned} \sum_{\mu=M+1}^{L} \sigma_{m_\mu}(R_\mu a, \mathcal{D}_\mu)_{s^{t,\Omega}_{q,\nu} y} &\lesssim 2^{-M\kappa(\frac{1}{\gamma_0} - \frac{1}{\delta_1})u} \sum_{\mu=M+1}^{L} \mu^{(d-1)(\frac{1}{\nu} - \frac{1}{\theta})u} 2^{\mu(\kappa(\frac{1}{\gamma_0} - \frac{1}{\delta_1}) - (r-t))u} \\ &\lesssim 2^{-M(r-t)u} M^{(d-1)(\frac{1}{\nu} - \frac{1}{\theta})u}. \end{aligned}$$

We estimate the last sum in (6.2.13). The choice of L yields $m_\mu = 0$ for $\mu > L$. Lemma 6.1 together with 6.12 gives

$$\begin{aligned}
\sigma_{m_\mu}(R_\mu a, \mathcal{D}_\mu)_{s^{t,\Omega}_{q,\nu}y} &\lesssim \|R_\mu a|s^{t,\Omega}_{q,\nu}y\| \le \mu^{(d-1)(\frac{1}{\nu}-\frac{1}{q})_+}\|R_\mu a|s^{t,\Omega}_{q,q}b\| \\
&\lesssim \mu^{(d-1)(\frac{1}{\nu}-\frac{1}{q})_+}2^{-(r-t-\frac{1}{p}+\frac{1}{q}))\mu}\|R_\mu a|s^{r,\Omega}_{p,p}b\| \\
&\lesssim \mu^{(d-1)[(\frac{1}{\nu}-\frac{1}{q})_+ + (\frac{1}{p}-\frac{1}{\theta})_+]}2^{-(r-t-\frac{1}{p}+\frac{1}{q}))\mu}\|R_\mu a|s^{r,\Omega}_{p,\theta}x\|.
\end{aligned}$$

Summing up with $L > M$ yields

$$\begin{aligned}
\sum_{\mu=L+1}^{\infty} \sigma_{m_\mu}(R_\mu a, \mathcal{D}_\mu)^u_{s^{t,\Omega}_{q,\nu}y} &\lesssim \sum_{\mu=L+1}^{\infty} \mu^{(d-1)[(\frac{1}{\nu}-\frac{1}{q})_+ + (\frac{1}{p}-\frac{1}{\theta})_+]u}2^{-(r-\frac{1}{p}+\frac{1}{q}))\mu u} \\
&\lesssim 2^{-(r-(\frac{1}{p}-\frac{1}{q}))Lu}L^{(d-1)[(\frac{1}{\nu}-\frac{1}{q})_+ + (\frac{1}{p}-\frac{1}{\theta})_+]u}.
\end{aligned}$$

Finally, if κ is choosen close enough to 1 then L is sufficent large such that

$$\sum_{\mu=L+1}^{\infty} \sigma_{m_\mu}(R_\mu a, \mathcal{D}_\mu)^u_{s^{t,\Omega}_{q,\nu}y} \lesssim 2^{-(r-(\frac{1}{p}-\frac{1}{q}))Lu}L^{(d-1)[(\frac{1}{\nu}-\frac{1}{q})_+ + (\frac{1}{p}-\frac{1}{\theta})_+]u} \lesssim 2^{-Mru}M^{(d-1)(\frac{1}{\nu}-\frac{1}{\theta})u}$$

holds. Altogether, we obtain

$$\sigma_m(a, \mathcal{D})_{s^{t,\Omega}_{q,\nu}y} \lesssim 2^{-M(r-t)}M^{(d-1)(\frac{1}{\nu}-\frac{1}{\theta})} \asymp (m^{-1}\log^{d-1} m)^{r-t}(\log^{d-1} m)^{\frac{1}{\nu}-\frac{1}{\theta}}.$$

Finally we consider the case $\delta_1 < \gamma_0$. Here we use (linear) hyperbolic cross approximation. Again we choose M such that

$$m \asymp 2^M M^{d-1}$$

and

$$m_\mu \asymp \begin{cases} \mu^{d-1}2^\mu & : \quad \mu \le M, \\ 0 & : \quad \text{otherwise.} \end{cases}$$

Obviously

$$\sum_{\mu=0}^{\infty} m_\mu \lesssim 2^M M^{d-1}.$$

Lemma 6.12 yields

$$\begin{aligned}
\|R_\mu a|s^{t,\Omega}_{q,\nu}y\| &\lesssim \mu^{(d-1)(\frac{1}{\nu}-\frac{1}{\delta_1})}\|R_\mu a|s^{t,\Omega}_{\delta_1,\delta_1}b\| \\
&= \mu^{(d-1)(\frac{1}{\nu}-\frac{1}{\delta_1})}2^{\mu(t-\frac{1}{\delta_1})}\|R_\mu a|\ell^{\mu^{d-1}2^\mu}_{\delta_1}\| \\
&\lesssim \mu^{(d-1)(\frac{1}{\nu}-\frac{1}{\delta_1})}2^{\mu(t-\frac{1}{\delta_1})}(2^\mu\mu^{d-1})^{\frac{1}{\delta_1}-\frac{1}{\gamma_0}}\|R_\mu a|\ell^{\mu^{d-1}2^\mu}_{\gamma_0}\| \\
&= \mu^{(d-1)(\frac{1}{\nu}-\frac{1}{\gamma_0})}2^{-\mu(r-t)}\|R_\mu a|s^{r,\Omega}_{\gamma_0,\gamma_0}b\| \\
&\lesssim \mu^{(d-1)(\frac{1}{\nu}-\frac{1}{\theta})}2^{-\mu(r-t)}\|R_\mu a|s^{r,\Omega}_{p,\theta}x\|
\end{aligned}$$

Inserting this into the following estimate gives

$$\begin{aligned}
\sigma_m(a,\mathcal{D})^u_{s^{t,\Omega}_{q,\nu}y} &\leq \sum_{\mu=0}^{M} \underbrace{\sigma_{m_\mu}(R_\mu a,\mathcal{D}_\mu)^u_{s^{t,\Omega}_{q,\nu}y}}_{=0} + \sum_{\mu=M+1}^{\infty} \underbrace{\sigma_0(R_\mu a,\mathcal{D}_\mu)^u_{s^{t,\Omega}_{q,\nu}y}}_{=\|R_\mu a|s^{t,\Omega}_{q,\nu}y\|^u} \\
&\lesssim \sum_{\mu=M+1}^{\infty} \mu^{(d-1)(\frac{1}{\nu}-\frac{1}{\theta})} 2^{-\mu(r-t)} \|R_\mu a|s^{r,\Omega}_{p,\theta}x\| \\
&\lesssim M^{(d-1)(\frac{1}{\nu}-\frac{1}{\theta})} 2^{-M(r-t)} \\
&\asymp (m^{-1}\log^{d-1} m)^{r-t} (\log^{d-1} m)^{\frac{1}{\nu}-\frac{1}{\theta}}.
\end{aligned}$$

This concludes the proof. □

Remark 6.14. *Compared to [56] the analysis here uses ideas known from the Maiorov discretization technique, cf. [75], which is very well known for estimates on several s-numbers of classical function space embeddings (F and B spaces).The choice of parameters is borrowed from [127, Theorem 3.19] where entropy numbers have been studied.*

The case of small smoothness

In this section we consider the so called case of small smoothness. The small smoothness range is given by $\frac{1}{p}-\frac{1}{q} \leq r \leq \frac{1}{\theta}-\frac{1}{\nu}$. Here we recover some interesting effects concerning the logarithm. The next result originally goes back to [56]. In fact, it was obtained in a non-constructive way using interpolation theory. We contribute a constructive approximation method.

Theorem 6.15. *Let $0<p<q\leq\infty$, $0<\theta<\nu\leq\infty$ and $\frac{1}{p}-\frac{1}{q}\leq r-t\leq\frac{1}{\theta}-\frac{1}{\nu}$. Then*

$$\sigma_m(id: s^{r,\Omega}_{p,\theta}b \to s^{t,\Omega}_{q,\nu}b,\mathcal{D}) \asymp m^{-(r-t)}. \tag{6.2.14}$$

Proof. For the lower bound we refer to [56, Corollary 5.11]. We prove the upper bound with a constructive method in case of the compact embedding with $0<\varepsilon<r-t-(\frac{1}{p}-\frac{1}{q})$. For the non-compact embedding $r-t=\frac{1}{p}-\frac{1}{q}$ we refer to the comments in Remark 6.19. We set

$$L = \left\lceil \frac{(r-t)\log m}{r-t-\frac{1}{p}+\frac{1}{q}-\varepsilon} \right\rceil. \tag{6.2.15}$$

Defining $u:=\min\{q,\nu,1\}$ then Lemma 6.1, (i) together with Lemma 6.12, (iv) yields

$$\begin{aligned}
\sigma_m(id: s^{r,\Omega}_{p,\theta}b \to s^{t,\Omega}_{q,\nu}b,\mathcal{D})^u &\leq \sigma_m\Big(\sum_{\mu=0}^{L} R_\mu : s^{r,\Omega}_{p,\theta}b \to s^{t,\Omega}_{q,\nu}b,\mathcal{D}\Big)^u \\
&\quad + \sum_{\mu=L+1}^{\infty} \sigma_0(R_\mu : s^{r,\Omega}_{p,\theta}b \to s^{t,\Omega}_{q,\nu}b,\mathcal{D}_\mu)^u.
\end{aligned} \tag{6.2.16}$$

We consider the first term in (6.2.16). By Corollary 6.10 we have

$$\sigma_{m_\mu}\Big(\sum_{\mu=0}^{L} R_\mu, \mathcal{D}\Big)_{s^{t,\Omega}_{q,\nu}b} \asymp \sigma_m(id : \ell_\theta(X_\mu) \to \ell_\nu(X_\mu)) \lesssim m^{-(r-t)} \tag{6.2.17}$$

where

$$X_\mu = \ell_\theta^{\mu^{d-1}}(2^{\mu(r-\frac{1}{p})}\ell_p^{2^\mu}) \quad \text{and} \quad Y_\mu = \ell_\nu^{\mu^{d-1}}(2^{\mu(t-\frac{1}{q})}\ell_q^{2^\mu}).$$

Finally we deal with the last sum in (6.2.16). Let $a \in s^{r,\Omega}_{p,\theta}b$ with $\|a|s^{r,\Omega}_{p,\theta}b\| \leq 1$, then applying Lemma 6.1, (i) together with Lemma 6.12 gives

$$\begin{aligned}\sigma_0(R_\mu a, \mathcal{D}_\mu)_{s^{t,\Omega}_{q,\nu}b} &= \|R_\mu a|s^{t,\Omega}_{q,\nu}b\| \leq \mu^{(d-1)(\frac{1}{\theta}-\frac{1}{\nu})}\|R_\mu a|s^{t,\Omega}_{q,\theta}b\| \\ &\lesssim \mu^{(d-1)(\frac{1}{\theta}-\frac{1}{\nu})}2^{-(r-t-\frac{1}{p}+\frac{1}{q})\mu}\|R_\mu a|s^{r,\Omega}_{p,\theta}b\| \\ &\lesssim \mu^{(d-1)(\frac{1}{\theta}-\frac{1}{\nu})}2^{-(r-t-\frac{1}{p}+\frac{1}{q})\mu}.\end{aligned}$$

Taking supremum over $\|a|s^{r,\Omega}_{p,\theta}b\| \leq 1$ and summing up yields

$$\sum_{\mu=L+1}^{\infty} \sigma_0(R_\mu : s^{r,\Omega}_{p,\theta}b \to s^{t,\Omega}_{q,\nu}b, \mathcal{D})^u \leq \sum_{\mu=L+1}^{\infty} \mu^{u(d-1)(\frac{1}{\theta}-\frac{1}{\nu})}2^{-(r-t-\frac{1}{p}+\frac{1}{q})\mu u} \tag{6.2.18}$$

$$\lesssim L^{u(d-1)(\frac{1}{\theta}-\frac{1}{\nu})}2^{-(r-t-(\frac{1}{p}-\frac{1}{q}))Lu}. \tag{6.2.19}$$

The choice of L in (6.2.15) gives

$$\sum_{\mu=L+1}^{\infty} \sigma_0(R_\mu : s^{r,\Omega}_{p,\theta}b \to s^{t,\Omega}_{q,\nu}b, \mathcal{D})^u \lesssim m^{-(r-t)}. \tag{6.2.20}$$

Inserting the estimates from (6.2.20), (6.2.17) into (6.2.16) yields

$$\sigma_m(id : s^{r,\Omega}_{p,\theta}b \to s^{t,\Omega}_{q,\nu}b, \mathcal{D}) \lesssim m^{-(r-t)}.$$

That proves the claim. □

Theorem 6.16. *Let $0 < p < q \leq \infty$ and $0 < \theta < \nu \leq \infty$ with $\frac{1}{p} - \frac{1}{q} \leq r - t \leq \frac{1}{\theta} - \frac{1}{\nu}$ and $q \geq \nu$ or $\theta \leq p < q < \nu$. Then*

$$\sigma_m(id : s^{r,\Omega}_{p,\theta}b \to s^{t,\Omega}_{q,\nu}f, \mathcal{D}) \asymp m^{-(r-t)}. \tag{6.2.21}$$

Proof. For the lower bound we refer to [54], Proposition 5.1. Similarly to Theorem 6.15 we prove the upper bound with a constructive method in case of the compact embedding with $r > \frac{1}{p} - \frac{1}{q}$. For the non-compact embedding $r = \frac{1}{p} - \frac{1}{q}$ we refer to the comments in Remark 6.19. The case $q \geq \nu$ follows from Theorem 6.15 using the decomposition provided in Figure 6.1 with Lemma 6.1, (iii). We prove the case $q < \nu$ with $p \geq \theta$. Let

$$N = \lfloor \log m \rfloor \text{ and } L \text{ as in (6.2.15).}$$

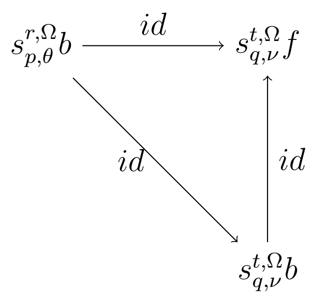


Figure 6.1: Trivial embedding in case $q \geq \nu$.

Defining $u := \min\{q, \nu, 1\}$ then Lemma 6.1, (ii) yields

$$\sigma_{2m}(id : s^{r,\Omega}_{p,\theta}b \to s^{t,\Omega}_{q,\nu}f, \mathcal{D})^u \leq \sigma_m\Big(\sum_{\mu=0}^{N} R_\mu : s^{r,\Omega}_{p,\theta}b \to s^{t,\Omega}_{q,\nu}f, \mathcal{D}\Big)^u \tag{6.2.22}$$

$$+\sigma_m\Big(\sum_{\mu=N+1}^{L} R_\mu : s^{r,\Omega}_{p,\theta}b \to s^{t,\Omega}_{q,\nu}f, \mathcal{D}\Big)^u$$

$$+\sum_{\mu=L+1}^{\infty} \sigma_0(R_\mu : s^{r,\Omega}_{p,\theta}b \to s^{t,\Omega}_{q,\nu}f, \mathcal{D})^u. \tag{6.2.23}$$

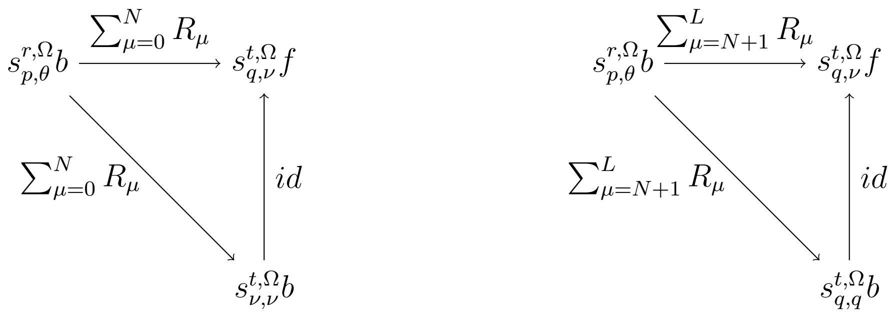


Figure 6.2: Decomposition of $\sum R_\mu$ in the $b - f$ case.

In the first sum we apply the decomposition provided in the left commutative diagram of Figure 6.2. Lemma 6.1, (iii) yields

$$\sigma_m\Big(\sum_{\mu=0}^{N} R_\mu : s^{r,\Omega}_{p,\theta}b \to s^{t,\Omega}_{q,\nu}f, \mathcal{D}\Big) \lesssim \sigma_m\Big(\sum_{\mu=0}^{N} R_\mu : s^{r,\Omega}_{p,\theta}b \to s^{t,\Omega}_{\nu,\nu}b, \mathcal{D}\Big) \underbrace{\|id : s^{t,\Omega}_{\nu,\nu}b \to s^{t,\Omega}_{q,\nu}f\|}_{\leq 1}.$$

The choice of N gives

$$\sup_{\mu=0,\dots,N} d_\mu = \sup_{\mu=0,\dots,N} 2^\mu \leq m \quad \text{and} \quad \inf_{\mu=N+1,\dots,L} d_\mu = \inf_{\mu=N+1,\dots,L} 2^\mu \geq m.$$

Additionally, we have $r-t<\frac{1}{\theta}-\frac{1}{\nu}$. This allows us to apply Corollary 6.8 which yields

$$\sigma_m\Big(\sum_{\mu=0}^{N} R_\mu : s^{r,\Omega}_{p,\theta}b \to s^{t,\Omega}_{\nu,\nu}b, \mathcal{D}\Big) \asymp \sigma_m(id : \ell_p(X_\mu) \to \ell_\nu(Y_\mu), \mathcal{D}) \lesssim m^{-(r-t)}, \tag{6.2.24}$$

where

$$X_\mu = \ell_\theta^{\mu^{d-1}}(2^{\mu(r-\frac{1}{p})}\ell_p^{2^\mu}) \quad \text{and} \quad Y_\mu = \ell_\nu^{\mu^{d-1}}(2^{\mu(t-\frac{1}{\nu})}\ell_\nu^{2^\mu}).$$

We estimate the second sum in (6.2.23) by using the right commutative diagram in Figure 6.2. Notice, that $\frac{1}{p}-\frac{1}{q}<r-t<\frac{1}{\theta}-\frac{1}{q}$. This allows us to apply Corollary 6.10 which yields

$$\sigma_m\Big(\sum_{\mu=0}^{N} R_\mu : s^{r,\Omega}_{p,\theta}b \to s^{t,\Omega}_{q,q}b, \mathcal{D}\Big) \asymp \sigma_m(id : \ell_p(X_\mu) \to \ell_q(Y_\mu), \mathcal{D}) \lesssim m^{-(r-t)}, \tag{6.2.25}$$

where

$$X_\mu = \ell_\theta^{\mu^{d-1}}(2^{\mu(r-\frac{1}{p})}\ell_p^{2^\mu}) \quad \text{and} \quad Y_\mu = \ell_q^{\mu^{d-1}}(2^{\mu(t-\frac{1}{q})}\ell_q^{2^\mu}).$$

Finally we deal with the last sum in (6.2.23). Let $a \in s^{r,\Omega}_{p,\theta}b$ with $\|a|s^{r,\Omega}_{p,\theta}b\| \leq 1$. Proceeding by applying Lemma 6.1, (i) and Lemma 6.12 gives

$$\begin{aligned}
\sigma_0(R_\mu a, \mathcal{D}_\mu)^u_{s^{t,\Omega}_{q,\nu}f} &\lesssim \|R_\mu a|s^{t,\Omega}_{q,q}f\| \leq \mu^{(d-1)(\frac{1}{\theta}-\frac{1}{q})}\|R_\mu a|s^{t,\Omega}_{q,\theta}b\| \\
&\lesssim \mu^{(d-1)(\frac{1}{\theta}-\frac{1}{q})}2^{-(r-t-\frac{1}{p}+\frac{1}{q})\mu}\|R_\mu a|s^{r,\Omega}_{p,\theta}b\| \\
&\lesssim \mu^{(d-1)(\frac{1}{\theta}-\frac{1}{q})}2^{-(r-t-\frac{1}{p}+\frac{1}{q})\mu}.
\end{aligned}$$

Taking supremum over a and summing up shows

$$\begin{aligned}
\sum_{\mu=L+1}^{\infty} \sigma_0(R_\mu : s^{r,\Omega}_{p,\theta}b \to s^{t,\Omega}_{q,\nu}f, \mathcal{D}_\mu)^u &\leq \sum_{\mu=L+1}^{\infty} \mu^{u(d-1)(\frac{1}{\theta}-\frac{1}{q})}2^{-(r-t-\frac{1}{p}+\frac{1}{q})\mu u} \\
&\lesssim L^{u(d-1)(\frac{1}{\theta}-\frac{1}{q})}2^{-(r-t-(\frac{1}{p}-\frac{1}{q}))Lu}.
\end{aligned}$$

Finally, the choice of L yields

$$\sum_{\mu=L+1}^{\infty} \sigma_0(R_\mu : s^{r,\Omega}_{p,\theta}b \to s^{t,\Omega}_{q,\nu}f, \mathcal{D}_\mu)^u \lesssim L^{u(d-1)(\frac{1}{\theta}-\frac{1}{q})}2^{-(r-t-(\frac{1}{p}-\frac{1}{q})Lu} \lesssim m^{-(r-t)}. \tag{6.2.26}$$

Inserting the estimates from (6.2.25), (6.2.24), (6.2.26) into (6.2.23) yields

$$\sigma_m(id : s^{r,\Omega}_{p,\theta}b \to s^{t,\Omega}_{q,\nu}f, \mathcal{D}) \lesssim m^{-(r-t)}.$$

This proves the claim. □

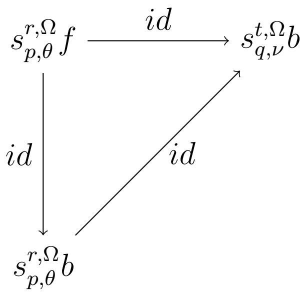


Figure 6.3: Trivial embedding in case $\theta \geq p$.

Theorem 6.17. *Let* $0 < p < q \leq \infty$ *and* $0 < \theta < \nu \leq \infty$ *with* $\frac{1}{p} - \frac{1}{q} < r - t \leq \frac{1}{\theta} - \frac{1}{\nu}$ *and* $\theta \geq p$ *or* $\theta < p < q \leq \nu$. *Then*

$$\sigma_m(id : s^{r,\Omega}_{p,\theta}f \to s^{t,\Omega}_{q,\nu}b, \mathcal{D}) \asymp m^{-(r-t)}. \tag{6.2.27}$$

Proof. For the lower bound we refer to [54], Proposition 5.1. Again, we prove the upper bound with a constructive method in case of the compact embedding with $r > \frac{1}{p} - \frac{1}{q}$. For the non-compact embedding $r = \frac{1}{p} - \frac{1}{q}$ we refer to the comments in Remark 6.19.. The case $\theta \geq p$ follows from Theorem 6.15 using the decomposition provided in Figure 6.3 with Lemma 6.1, (iii). We prove the case $q \leq \nu$ with $p \geq \theta$. Let

$$N = \lfloor \log m \rfloor \text{ and } L \text{ as in (6.2.15).}$$

Defining $u := \min\{q, \nu, 1\}$ Lemma 6.1, (ii) yields

$$\begin{aligned}\sigma_{2m}(id : s^{r,\Omega}_{p,\theta}f \to s^{t,\Omega}_{q,\nu}b, \mathcal{D})^u \quad \leq \quad & \sigma_m\Big(\sum_{\mu=0}^{N} R_\mu : s^{r,\Omega}_{p,\theta}f \to s^{t,\Omega}_{q,\nu}b, \mathcal{D}\Big)^u \\ & + \sigma_m\Big(\sum_{\mu=N+1}^{L} R_\mu : s^{r,\Omega}_{p,\theta}f \to s^{t,\Omega}_{q,\nu}b, \mathcal{D}\Big)^u \\ & + \sum_{\mu=L+1}^{\infty} \sigma_0(R_\mu : s^{r,\Omega}_{p,\theta}f \to s^{t,\Omega}_{q,\nu}b, \mathcal{D}_\mu)^u.\end{aligned} \tag{6.2.28}$$

We consider the first sum where we use the decomposition presented in the left commutative diagram of Figure 6.4.

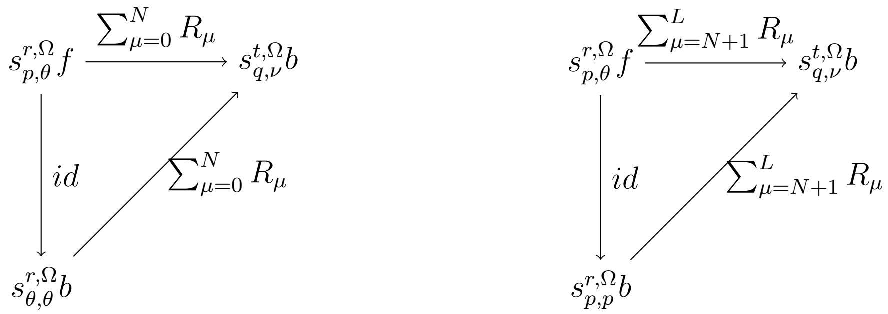


Figure 6.4: Decomposition of $\sum R_\mu$ in the $f-b$ case.

Then Lemma 6.1, (iii) provides

$$\sigma_m\Big(\sum_{\mu=0}^{N} R_\mu : s^{r,\Omega}_{p,\theta}f \to s^{t,\Omega}_{q,\nu}b, \mathcal{D}\Big) \lesssim \sigma_m\Big(\sum_{\mu=0}^{N} R_\mu : s^{r,\Omega}_{\theta,\theta}b \to s^{t,\Omega}_{q,\nu}b, \mathcal{D}\Big) \underbrace{\|id : s^{r,\Omega}_{\theta,\theta}f \to s^{r,\Omega}_{\theta,\theta}b\|}_{\leq 1}.$$

The choice of N gives

$$\sup_{\mu=0,\dots,N} d_\mu = \sup_{\mu=0,\dots,N} 2^\mu \leq m \quad \text{and} \quad \inf_{\mu=N+1,\dots,L} d_\mu = \inf_{\mu=N+1,\dots,L} 2^\mu \geq m.$$

Additionally we have $r-t < \frac{1}{\theta} - \frac{1}{\nu}$. This allows us to apply Corollary 6.8 which yields

$$\sigma_m\Big(\sum_{\mu=0}^{N} R_\mu : s^{r,\Omega}_{p,\theta}f \to s^{t,\Omega}_{q,\nu}b, \mathcal{D}\Big) \lesssim \sigma_m(id : \ell_\theta(X_\mu) \to \ell_\nu(Y_\mu), \mathcal{D}) \lesssim m^{-(r-t)}, \quad (6.2.29)$$

where

$$X_\mu = \ell_\theta^{\mu^{d-1}}(2^{\mu(r-\frac{1}{\theta})}\ell_\theta^{2^\mu}) \quad \text{and} \quad Y_\mu = \ell_\nu^{\mu^{d-1}}(2^{\mu(t-\frac{1}{q})}\ell_q^{2^\mu}).$$

We estimate the second sum in (6.2.28) by using the decomposition provided in the right commutative diagram in Figure 6.4. Note that $\frac{1}{p} - \frac{1}{q} < r-t$. This allows us to apply Corollary 6.9 that yields

$$\sigma_m\Big(\sum_{\mu=0}^{N} R_\mu : s^{r,\Omega}_{p,p}b \to s^{t,\Omega}_{q,\nu}b, \mathcal{D}\Big) \lesssim \sigma_m(id : \ell_p(X_\mu) \to \ell_q(Y_\mu), \mathcal{D}) \lesssim m^{-(r-t)}, \quad (6.2.30)$$

where

$$X_\mu = \ell_p^{\mu^{d-1}}(2^{\mu(r-\frac{1}{p})}\ell_p^{2^\mu}) \quad \text{and} \quad Y_\mu = \ell_\nu^{\mu^{d-1}}(2^{\mu(t-\frac{1}{q})}\ell_q^{2^\mu}).$$

Finally we deal with the third sum in (6.2.23). Let $a \in s^{r,\Omega}_{p,\theta}f$ with $\|a|s^{r,\Omega}_{p,\theta}f\| \leq 1$. Applying Lemma 6.1, (i) and Lemma 6.12 gives

$$\begin{aligned}
\sigma_0(R_\mu a, \mathcal{D})^u_{s^{t,\Omega}_{q,\nu}b} &\lesssim \|R_\mu a|s^{t,\Omega}_{q,\nu}f\| \leq \mu^{(d-1)(\frac{1}{p}-\frac{1}{\nu})}\|R_\mu a|s^{t,\Omega}_{q,p}b\| \\
&\lesssim \mu^{(d-1)(\frac{1}{p}-\frac{1}{\nu})}2^{-(r-t-\frac{1}{p}+\frac{1}{q})\mu}\|R_\mu a|s^{r,\Omega}_{p,p}b\| \\
&\lesssim \mu^{(d-1)(\frac{1}{p}-\frac{1}{\nu})}2^{-(r-t-\frac{1}{p}+\frac{1}{q})\mu}\underbrace{\|R_\mu a|s^{r,\Omega}_{p,\theta}f\|}_{\leq 1}.
\end{aligned}$$

Taking supremum over a and summing up shows

$$\begin{aligned}\sum_{\mu=L+1}^{\infty} \sigma_0(R_\mu : s^{r,\Omega}_{p,\theta} f \to s^{t,\Omega}_{q,\nu} b, \mathcal{D})^u &\le \sum_{\mu=L+1}^{\infty} \mu^{u(d-1)(\frac{1}{p}-\frac{1}{\nu})} 2^{-(r-t-\frac{1}{p}+\frac{1}{q})\mu u} \\ &\lesssim L^{u(d-1)(\frac{1}{p}-\frac{1}{\nu})} 2^{-(r-t-(\frac{1}{p}-\frac{1}{q}))Lu}.\end{aligned}$$

Finally, the choice of L yields

$$\sum_{\mu=L+1}^{\infty} \sigma_0(R_\mu : s^{r,\Omega}_{p,\theta} f \to s^{t,\Omega}_{q,\nu} b, \mathcal{D})^u \lesssim L^{u(d-1)(\frac{1}{p}-\frac{1}{\nu})} 2^{-(r-t-(\frac{1}{p}-\frac{1}{q})Lu} \lesssim m^{-(r-t)}. \quad (6.2.31)$$

Inserting the estimates from (6.2.30), (6.2.29), (6.2.31) into (6.2.28) yields

$$\sigma_m(id : s^{r,\Omega}_{p,\theta} b \to s^{t,\Omega}_{q,\nu} f, \mathcal{D}) \lesssim m^{-(r-t)},$$

which concludes the proof. □

Theorem 6.18. *Let* $0 < p < q \le \infty$ *and* $0 < \theta < \nu \le \infty$ *with* $\frac{1}{p} - \frac{1}{q} < r - t \le \frac{1}{\theta} - \frac{1}{\nu}$ *and* $p \le \theta$, $\nu \le q$ *or* $\theta \le p < q \le \nu$. *Then*

$$\sigma_m(id : s^{r,\Omega}_{p,\theta} f \to s^{t,\Omega}_{q,\nu} f, \mathcal{D}) \asymp m^{-(r-t)}. \quad (6.2.32)$$

Proof. For the lower bound we refer to [54], Proposition 5.1. We prove the upper bound in case $r > \frac{1}{p} - \frac{1}{q}$ with a constructive way. For the case $r = \frac{1}{p} - \frac{1}{q}$ we refer to Remark 6.19. The case $\theta \ge p$, $q \ge \nu$ follows from Theorem 6.15 using the the commutative diagram provided in Figure 6.5 together with Lemma 6.1, (iii). We prove the case

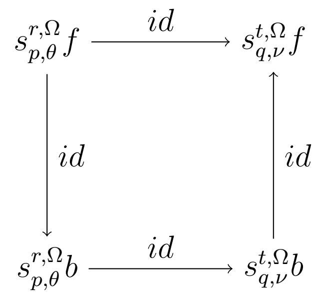


Figure 6.5: Trivial embeddings in case $p \le \theta, \nu \le q$.

$q \le \nu$ with $p \ge \theta$. Let

$$N = \lfloor \log m \rfloor \text{ and } L \text{ as in (6.2.15).}$$

Defining $u := \min\{q,\nu,1\}$ Lemma 6.1, (ii) yields

$$\begin{aligned}\sigma_{2m}(id : s^{r,\Omega}_{p,\theta}f \to s^{t,\Omega}_{q,\nu}f,\mathcal{D})^u \;\le\; & \sigma_m\Big(\sum_{\mu=0}^{N} R_\mu : s^{r,\Omega}_{p,\theta}f \to s^{t,\Omega}_{q,\nu}f,\mathcal{D}\Big)^u \qquad (6.2.33)\\ & +\sigma_m\Big(\sum_{\mu=N+1}^{L} R_\mu : s^{r,\Omega}_{p,\theta}f \to s^{t,\Omega}_{q,\nu}f,\mathcal{D}\Big)^u\\ & +\sum_{\mu=L+1}^{\infty} \sigma_0(R_\mu : s^{r,\Omega}_{p,\theta}f \to s^{t,\Omega}_{q,\nu}f,\mathcal{D})^u. \qquad (6.2.34)\end{aligned}$$

Concerning the first sum we consider the left commutative diagram in Figure 6.6.

$$\begin{array}{ccc} s^{r,\Omega}_{p,\theta}f & \xrightarrow{\sum_{\mu=0}^{N} R_\mu} & s^{t,\Omega}_{q,\nu}f \\ \downarrow id & & \uparrow id \\ s^{r,\Omega}_{\theta,\theta}b & \xrightarrow{\sum_{\mu=0}^{N} R_\mu} & s^{t,\Omega}_{\nu,\nu}b \end{array} \qquad\qquad \begin{array}{ccc} s^{r,\Omega}_{p,\theta}f & \xrightarrow{\sum_{\mu=N+1}^{L} R_\mu} & s^{t,\Omega}_{q,\nu}f \\ \downarrow id & & \uparrow id \\ s^{r,\Omega}_{p,p}b & \xrightarrow{\sum_{\mu=N+1}^{L} R_\mu} & s^{t,\Omega}_{q,q}b \end{array}$$

Figure 6.6: Decomposition of $\sum R_\mu$ in the $f-f$ case.

Lemma 6.1, (iii) provides

$$\begin{aligned}\sigma_m\Big(\sum_{\mu=0}^{N} R_\mu : s^{r,\Omega}_{p,\theta}b \to s^{t,\Omega}_{q,\nu}f,\mathcal{D}\Big) \;\lesssim\; & \underbrace{\|id : s^{t,\Omega}_{\nu,\nu}f \to s^{r,\Omega}_{\theta,\theta}b\|}_{\le 1}\,\sigma_m\Big(\sum_{\mu=0}^{N} R_\mu : s^{r,\Omega}_{\theta,\theta}b \to s^{t,\Omega}_{\nu,\nu}b,\mathcal{D}\Big)\\ & \times \underbrace{\|id : s^{r,\Omega}_{p,\theta}f \to s^{r,\Omega}_{\theta,\theta}b\|}_{\le 1}.\end{aligned}$$

The choice of N gives

$$\sup_{\mu=0,\dots,N} d_\mu = \sup_{\mu=0,\dots,N} 2^\mu \le m \quad \text{and} \quad \inf_{\mu=N+1,\dots,L} d_\mu = \inf_{\mu=N+1,\dots,L} 2^\mu \ge m.$$

Additionally we have $r-t \le \frac{1}{\theta}-\frac{1}{\nu}$. This allows us to apply Corollary 6.8 that yields

$$\sigma_m\Big(\sum_{\mu=0}^{N} R_\mu : s^{r,\Omega}_{p,\theta}f \to s^{t,\Omega}_{q,\nu}f,\mathcal{D}\Big) \lesssim \sigma_m(id : \ell_\theta(X_\mu) \to \ell_\nu(Y_\mu),\mathcal{D}) \lesssim m^{-(r-t)}, \qquad (6.2.35)$$

where

$$X_\mu = \ell_\theta^{\mu^{d-1}}(2^{\mu(r-\frac{1}{\theta})}\ell_\theta^{2^\mu}) \quad \text{and} \quad Y_\mu = \ell_\nu^{\mu^{d-1}}(2^{\mu(t-\frac{1}{\nu})}\ell_\nu^{2^\mu}).$$

We estimate the second sum in (6.2.34) by using the right commutative diagram in Figure 6.6. We recognize $\frac{1}{p}-\frac{1}{q}<r-t$. This allows us to apply Corollary 6.9 which yields

$$\sigma_m\Big(\sum_{\mu=N+1}^{L} R_\mu : s^{r,\Omega}_{p,p}b \to s^{t,\Omega}_{q,q}b,\mathcal{D}\Big) \lesssim \sigma_m(id:\ell_p(X_\mu)\to\ell_q(Y_\mu),\mathcal{D}) \lesssim m^{-(r-t)}, \quad (6.2.36)$$

where

$$X_\mu = \ell_p^{\mu^{d-1}}(2^{\mu(r-\frac{1}{p})}\ell_p^{2^\mu}) \quad \text{and} \quad Y_\mu = \ell_q^{\mu^{d-1}}(2^{\mu(t-\frac{1}{q})}\ell_q^{2^\mu}).$$

Finally we deal with the last sum in (6.2.23). Let $a \in s^{r,\Omega}_{p,\theta}f$ with $\|a|s^{r,\Omega}_{p,\theta}f\| \le 1$. Applying Lemma 6.1, (i) and Lemma 6.12 gives

$$\begin{aligned}\sigma_0(R_\mu a,\mathcal{D})^u_{s^{t,\Omega}_{q,\nu}f} &\lesssim \|R_\mu a|s^{t,\Omega}_{q,\nu}f\| \le \mu^{(d-1)(\frac{1}{\theta}-\frac{1}{\nu})}\|R_\mu a|s^{t,\Omega}_{q,p}f\| \\ &\lesssim \mu^{(d-1)(\frac{1}{\theta}-\frac{1}{\nu})}2^{-(r-t-\frac{1}{p}+\frac{1}{q})\mu}\|R_\mu a|s^{r,\Omega}_{p,\theta}f\|.\end{aligned}$$

Taking supremum over a and summing up shows

$$\begin{aligned}\sum_{\mu=L+1}^{\infty}\sigma_0(R_\mu : s^{r,\Omega}_{p,\theta}f \to s^{t,\Omega}_{q,\nu}f,\mathcal{D})^u &\lesssim \sum_{\mu=L+1}^{\infty}\mu^{u(d-1)(\frac{1}{\theta}-\frac{1}{\nu})}2^{-(r-t-\frac{1}{p}+\frac{1}{q})\mu u} \\ &\lesssim L^{u(d-1)(\frac{1}{\theta}-\frac{1}{\nu})}2^{-(r-t-(\frac{1}{p}-\frac{1}{q}))Lu}.\end{aligned}$$

Finally, the choice of L yields

$$\sum_{\mu=L+1}^{\infty}\sigma_0(R_\mu : s^{r,\Omega}_{p,\theta}f \to s^{t,\Omega}_{q,\nu}b,\mathcal{D})^u \lesssim L^{u(d-1)(\frac{1}{p}-\frac{1}{\nu})}2^{-(r-t-(\frac{1}{p}-\frac{1}{q})Lu} \lesssim m^{-(r-t)}. \quad (6.2.37)$$

Inserting the estimates from (6.2.36), (6.2.35), (6.2.37) into (6.2.34) yields

$$\sigma_m(id: s^{r,\Omega}_{p,\theta}f \to s^{t,\Omega}_{q,\nu}f,\mathcal{D}) \lesssim m^{-(r-t)},$$

which concludes the proof. □

Remark 6.19. *Setting $L=\infty$ the proofs of Theorems 6.15, 6.16, 6.17 and 6.18 work also in case $r-t=\frac{1}{p}-\frac{1}{q}$ (non-compact embedding). The price to pay is the constructivity of the underlying algorithm. The algorithm needs full knowledge of the coefficients on infinitely many hyperbolic layers $M(\mu,d)$.*

6.3 Explicit algorithms

The results in Theorems 6.13, 6.15, 6.16, 6.17 and 6.18 are constructive. A constructive algorithm that approximates $a\in s^{r,\Omega}_{p,\theta}x$ is an algorithm, that needs for its evaluation only partial (finite) knowledge of the coefficients of a. Algorithm 1 and Algorithm

2 describe the method that constructs the approximants for the approximation rates provided in the mentioned theorems for series $a \in s^{r,\Omega}_{x,\theta}b$, $x \in \{f,b\}$ with $\|a|s^{r,\Omega}_{p,\theta}x\| \le 1$. These algorithms are obtained by inserting the approximation methods from Section 6.2 into the corresponding estimates in this section. Note, that the algorithms presented here are modified due to Remark 6.6, so that they can handle $\|a|s^{r,\Omega}_{p,\theta}b\| = 1$ directly. Algorithm 2 considers the small smoothness $b-b$ situation, which means where the source and the target space is a Besov type sequence space. The underlying methods of Theorems 6.17, 6.16 and 6.18 are based on Algorithm 2. Here one has to divde a into two parts

$$a = \sum_{|\boldsymbol{j}|_1=0}^{N} \sum_{\boldsymbol{k}\in D_{\boldsymbol{j}}} a_{\boldsymbol{j},\boldsymbol{k}} e_{\boldsymbol{j},\boldsymbol{k}} + \sum_{|\boldsymbol{j}|_1=N+1}^{\infty} \sum_{\boldsymbol{k}\in D_{\boldsymbol{j}}} a_{\boldsymbol{j},\boldsymbol{k}} e_{\boldsymbol{j},\boldsymbol{k}}$$

where Algorithm 2 is applied to each of these parts with different embedding parameters p, q, θ, ν.

6.4 Best m-term approximation with respect to the Faber-Schauder system

We denote by

$$\mathbb{F}^d := \{v_{\boldsymbol{j},\boldsymbol{k}} : (\boldsymbol{j},\boldsymbol{k}) \in \nabla\}$$

the Faber-Schauder dictionary on $[0,1]^d$. In this section we consider best m-term approximation in function spaces with respect to $\mathbb{F}^d$

$$\sigma_m(S^r_{p,\theta}X, \mathbb{F}^d)_{S^t_{q,\nu}Y} = \sigma_m(id : S^r_{p,\theta}X \to S^t_{q,\nu}Y, \mathbb{F}^d).$$

Lower bounds

The next theorem is our main result concering lower bounds for best m-term approximation with respect to the Faber-Schauder dictionary.

Theorem 6.20. *Let* $0 < p < q \le \infty$, $0 < \theta \le \infty$ *(B-case:* $p \le q = \infty$*) and* $r > \frac{1}{p}$ *(F-case:* $r > \max\{\frac{1}{p}, \frac{1}{\theta} - 1\}$*. Then*

$$\sigma_m(S^r_{p,\theta}B([0,1]^d), \mathbb{F}^d)_{L_q} \gtrsim m^{-r}$$

and

$$\sigma_m(S^r_{p,\theta}F([0,1]^d), \mathbb{F}^d)_{L_q} \gtrsim m^{-r}$$

for all $m \in \mathbb{N}$.

Proof. We consider the bump function

$$b(x) = e^{-\frac{1}{x(1-x)}} e^{\frac{1}{4}}$$

Algorithm 1 Large smoothnes algorithm

Input:	m, r, t	degrees of freedom, smoothness
	$\gamma_0 := \min\{p, \theta\}$,	parameters
	$\delta_1 := \max\{q, \nu\}$	
	$(a_{\boldsymbol{j},\boldsymbol{k}}), \lvert\boldsymbol{j}\rvert_1 \leq L, \boldsymbol{k} \in D_{\boldsymbol{j}}$	finite part of a.

choose κ with

$$1 < \kappa < \frac{r-t}{\frac{1}{\gamma_0} - \frac{1}{\delta_1}}.$$

choose M such that

$$m \asymp M^{d-1}2^M.$$

set

$$L := \left\lceil \frac{M\kappa + (d-1)\log M - 1}{\kappa - 1} \right\rceil,$$

$$a_m := 0.$$

for each $\mu \in 0, \ldots, L$ ***do***
  set

$$m_\mu \asymp \begin{cases} 2^\mu \mu^{d-1} & : \quad 0 \leq \mu \leq M, \\ \lfloor 2^\mu 2^{(M-\mu)\kappa} \mu^{d-1} \rfloor & : \quad M+1 \leq \mu \leq L. \end{cases}$$

  for each $\boldsymbol{j}$ ***with*** $\lvert\boldsymbol{j}\rvert_1 = \mu$ ***do***
    for each $\boldsymbol{k} \in D_{\boldsymbol{j}}$ ***do***
      if $\lvert a_{\boldsymbol{j},\boldsymbol{k}}\rvert^{\gamma_0} m_\mu \geq 1$ ***then***
        set

$$a_m \quad := \quad a_m + a_{\boldsymbol{j},\boldsymbol{k}} e_{\boldsymbol{j},\boldsymbol{k}}$$

      end if
    end for
  end for
end for

Output: a_m m-term approximation to a

Algorithm 2 Small smoothness algorithm b-b-case

Input:	m, r, t	degrees of freedom, smoothness
	$p, q, \theta, \nu,$	parameters
	$(a_{\boldsymbol{j},\boldsymbol{k}}), \lvert\boldsymbol{j}\rvert_1 \leq L, \boldsymbol{k} \in D_{\boldsymbol{j}}$	finite part of a.

choose

$$L \text{ as in } (6.2.15)$$

Set

$$a_m := 0$$

for each $\mu \in 0, \ldots, L$ ***do***

$$m_\mu := \|R_\mu a | s^{r,\Omega}_{p,\theta} b\|^\theta m$$

for each j with $|\boldsymbol{j}|_1 = \mu$ ***do***

$$m_{\mu,\boldsymbol{j}} := 2^{\theta\mu(r-\frac{1}{p})} \|a|\ell_p^{D_{\boldsymbol{j}}}\|^\theta m_\mu$$

for each $\boldsymbol{k} \in D_{\boldsymbol{j}}$ ***do***
if $|a_{\boldsymbol{j},\boldsymbol{k}}|^p m_{\mu,\boldsymbol{j}} \geq 1$ ***then***
set

$$a_m := a_m + a_{\boldsymbol{j},\boldsymbol{k}} e_{\boldsymbol{j},\boldsymbol{k}}$$

end if
end for
end for
end for

Output: a_m best m-term approximation to a

which is a L_∞-normalized C_0^∞ function. We denote by

$$b_{j,k} = b(2^j x - k)$$

its j-th dilation and k-th translation. Taking a linear function it is obvious that

$$\|b(x) - ax + m|L_q([0,1])\| \geq C_1 \tag{6.4.1}$$

holds. The bumps $(b_{j,k})_{\boldsymbol{j}\in\mathbb{N}_0^d,\boldsymbol{k}\in\mathbb{Z}^d}$ are $(K,-1)$-atoms. According to Theorem 4.34 the relation

$$\Big\| \sum_{\boldsymbol{j}\in\mathbb{N}_{-1}^d} \sum_{\boldsymbol{k}\in D_{\boldsymbol{j}}} \lambda_{\boldsymbol{j},\boldsymbol{k}} b_{\boldsymbol{j},\boldsymbol{k}} \Big| S^r_{p,\theta}X([0,1]^d)\Big\| \lesssim \|\lambda|s^{r,\Omega}_{p,\theta}x\| \tag{6.4.2}$$

holds for every sequence $(\lambda_{\boldsymbol{j},\boldsymbol{k}})$ of complex numbers with finite (RHS) in (6.4.2). Let us define

$$f_M(\boldsymbol{x}) = \sum_{\boldsymbol{k}\in D_j} 2^{-(M+1)r} b_{M+1,k}(x_1).$$

The relation in (6.4.2) easily allows to prove

$$\|f_M|S^r_{p,\theta}X([0,1]^d)\| \leq \|f_M|S^r_{p,\theta}X(\mathbb{R}^d)\| \leq \|(2^{-(M+1)r})_{\boldsymbol{j}=j(M),\boldsymbol{k}\in D_{j(M)}}|s^r_{p,\theta}x\| \leq 1.$$

We consider a 2^M-term from the dictionary $\mathbb{F}^d$ given by

$$g_M = \sum_{(\boldsymbol{j},\boldsymbol{k})\in\Lambda_M} \lambda_{\boldsymbol{j},\boldsymbol{k}} v_{\boldsymbol{j},\boldsymbol{k}}$$

where $\Lambda_M \subset \nabla$ (cf. (4.4.1)) with $|\Lambda_M| = 2^M$. We decompose the approximation of f_M by g_M as follows

$$\|f_M - g_M|L_q([0,1]^d)\| = \|u_M - a_M|L_q([0,1]^d)\| \tag{6.4.3}$$

where

$$u_M = f_M - \sum_{\substack{(\boldsymbol{j},\boldsymbol{k})\in\Lambda_M \\ |\boldsymbol{j}|_1\geq M+1}} \lambda_{\boldsymbol{j},\boldsymbol{k}} v_{\boldsymbol{j},\boldsymbol{k}} \quad \text{and} \quad a_M = \sum_{\substack{(\boldsymbol{j},\boldsymbol{k})\in\Lambda_M \\ |\boldsymbol{j}|_1< M+1}} \lambda_{\boldsymbol{j},\boldsymbol{k}} v_{\boldsymbol{j},\boldsymbol{k}}.$$

Furthermore, let

$$I_{\boldsymbol{j},\boldsymbol{k}} = \operatorname{supp} v_{\boldsymbol{j},\boldsymbol{k}}.$$

Additionally, we decompose the domain

$$\Omega = [0,1]^d = \bigcup_{\boldsymbol{k}\in D_{(M+1,\ldots,M+1)}} E_{\boldsymbol{k}}$$

into elementary cells $E_{\boldsymbol{k}} := I_{(M+1,\ldots,M+1),\boldsymbol{k}}$. Simple volume arguments for the support of u_m yield that u_M can differ from f_M only in $2^{(M+1)d-1}$ elementary cells. As a consquence we find a set $A \subset D_{(M+1,\ldots,M+1)}$, $|A| \geq 2^{d(M+1)-1}$ such that

$$f_M(\boldsymbol{x}) = u_M(\boldsymbol{x}) \quad \text{for all} \quad \boldsymbol{x} \in \bigcup_{\boldsymbol{k}\in A} E_{\boldsymbol{k}}.$$

We continue estimating (6.4.3) by

$$\|f_M - g_M|L_q([0,1]^d)\|^q = \sum_{\boldsymbol{k}\in A} \|f_M - a_M|L_q(E_{\boldsymbol{k}})\|^q. \tag{6.4.4}$$

Considering a single summand we obtain

$$\begin{aligned}\|f_M - a_M|L_q(E_{\boldsymbol{k}})\|^q &= \int_{E_{\boldsymbol{k}}} |f_M(\boldsymbol{x}) - a_M(\boldsymbol{x})|^q d\boldsymbol{x}\\ &= \int_{x_d} \cdots \int_{x_1} |2^{-(M+1)r} b_{M+1,k^*}(x_1) - (a(x_2,\ldots,x_d)x_1 - m(x_2,\ldots,x_d))|^q dx_1 \ldots dx_d,\end{aligned}$$

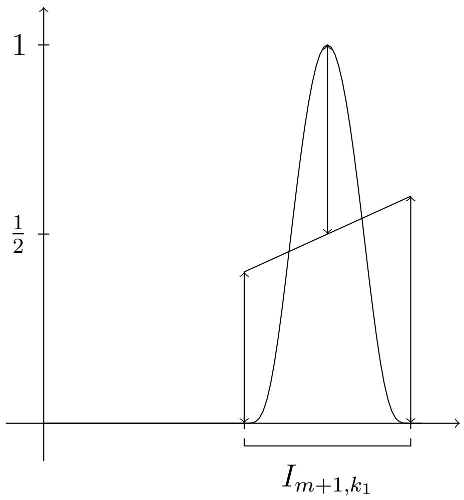
where a and m are functions mapping $\mathbb{R}^{d-1} \to \mathbb{C}$. This representation is possible since a_M consists of frequencies smaller $M+1$ in every direction, which means it is piecewise linear in every single direction of an elementary cell $E_{\boldsymbol{k}}$. Change of variable gives

$$\begin{aligned}&\|f_M - a_M|L_q(E_{\boldsymbol{k}})\|^q\\ =\ & \int_{x_d} \cdots \int_{x_2} \|b(\cdot) - (a^*(x_2,\ldots,x_d)\cdot - m^*(x_2,\ldots,x_d))|L_q(I_{M+1,k_1})\|^q dx_2 \ldots dx_d\\ &\times 2^{-(M+1)r-M}\end{aligned}$$

Applying the observation in (6.4.1) yields

$$\begin{aligned}\|f_M - a_M|L_q(E_{\boldsymbol{k}})\|^q &\gtrsim C_1 2^{-(M+1)r-M} \int_{x_d} \cdots \int_{x_2} 1 dx_2 \ldots dx_d\\ &\asymp 2^{-(M+1)d} 2^{-(M+1)r}.\end{aligned}$$

Inserting this into (6.4.4) gives



$$\begin{aligned}\|f_M - g_M|L_q([0,1]^d)\|^q &\gtrsim \sum_{\boldsymbol{k}\in A} 2^{(M+1)(d-1)} 2^{-(M+1)r}\\ &= |A| 2^{(M+1)(d-1)} 2^{-(M+1)r}\\ &\asymp 2^{-(M+1)r}\end{aligned}$$

This yields

$$\sigma_{2^M}(S^r_{p,\theta}X,\mathbb{F})_{L_q} \gtrsim \inf_{\Lambda_M,\lambda_{\boldsymbol{j},\boldsymbol{k}}} \left\| f_M - \sum_{(\boldsymbol{j},\boldsymbol{k})\in\Lambda_M} \lambda_{\boldsymbol{j},\boldsymbol{k}} v_{\boldsymbol{j},\boldsymbol{k}} \middle| L_q([0,1]^d) \right\| \gtrsim 2^{-Mr}.$$

Simple monotonicity arguments finally show,

$$\sigma_m(S^r_{p,\theta}X,\mathbb{F})_{L_q} \gtrsim m^{-r}.$$

□

Upper bounds

In this subsection we apply the sequence space results from the last section to obtain estimates for best m-term approximations in function spaces.

Theorem 6.21. *Let* $\frac{1}{2} < p < q \leq \infty$, $0 < \theta, \nu \leq \infty$ *and* $\frac{1}{p} < r < \min\{\frac{1}{\theta} - \frac{1}{\min\{q,1\}}, 2\}$ *or* $\frac{1}{p} < r = \frac{1}{\theta} - \frac{1}{\min\{q,1\}} < 2$. *Then*

$$\sigma_m(S^r_{p,\theta}B([0,1]^d),\mathbb{F}^d)_{L_q} \asymp m^{-r}.$$

Proof. The lower bound is due to Theorem 6.20. We prove the upper bound. Theorem 4.25 allows us to write $f \in S^r_{p,\theta}B([0,1]^d)$ as a Faber-Schauder series. Let $u = \min\{q,1\}$. This gives

$$\begin{aligned}
&\sigma_m(S^r_{p,\theta}B([0,1]^d),\mathbb{F}^d)^u_{L_q([0,1]^d)} \\
&= \sup_{\|f|S^r_{p,\theta}B([0,1]^d)\|\leq 1} \inf_{\substack{\Lambda\subset\nabla \\ |\Lambda|\leq m}} \inf_{\substack{(\lambda_{\boldsymbol{j},\boldsymbol{k}})\subset\mathbb{C} \\ \lambda_{\boldsymbol{j},\boldsymbol{k}}\neq 0 \Longrightarrow (\boldsymbol{j},\boldsymbol{k})\in\Lambda}} \left\| \sum_{\boldsymbol{j}\in\mathbb{N}^d_{-1}} \sum_{\boldsymbol{k}\in D_{\boldsymbol{j}}} (d_{\boldsymbol{j},\boldsymbol{k}}(f) - \lambda_{\boldsymbol{j},\boldsymbol{k}}) v_{\boldsymbol{j},\boldsymbol{k}} \middle| L_q([0,1]^d) \right\|^u.
\end{aligned}$$

u-triangle inequality yields

$$\begin{aligned}
&\sigma_m(S^r_{p,\theta}B([0,1]^d),\mathbb{F}^d)^u_{L_q([0,1]^d)} \\
&\lesssim \sup_{\|f|S^r_{p,\theta}B([0,1]^d)\|\leq 1} \inf_{\substack{\Lambda\subset\nabla \\ |\Lambda|\leq m}} \inf_{\substack{(\lambda_{\boldsymbol{j},\boldsymbol{k}})\subset\mathbb{C} \\ \lambda_{\boldsymbol{j},\boldsymbol{k}}\neq 0 \Longrightarrow (\boldsymbol{j},\boldsymbol{k})\in\Lambda}} \sum_{\boldsymbol{j}\in\mathbb{N}^d_{-1}} \left\| \sum_{\boldsymbol{k}\in D_{\boldsymbol{j}}} (d_{\boldsymbol{j},\boldsymbol{k}}(f) - \lambda_{\boldsymbol{j},\boldsymbol{k}}) \chi_{\boldsymbol{j},\boldsymbol{k}} \middle| L_q([0,1]^d) \right\|^u \\
&\asymp \sup_{\|f|S^r_{p,\theta}B([0,1]^d)\|\leq 1} \inf_{\substack{\Lambda\subset\nabla \\ |\Lambda|\leq m}} \inf_{\substack{(\lambda_{\boldsymbol{j},\boldsymbol{k}})\subset\mathbb{C} \\ \lambda_{\boldsymbol{j},\boldsymbol{k}}\neq 0 \Longrightarrow (\boldsymbol{j},\boldsymbol{k})\in\Lambda}} \sum_{\boldsymbol{j}\in\mathbb{N}^d_{-1}} 2^{-\frac{u|\boldsymbol{j}|_1}{q}} \Big(\sum_{\boldsymbol{k}\in D_{\boldsymbol{j}}} (d_{\boldsymbol{j},\boldsymbol{k}}(f) - \lambda_{\boldsymbol{j},\boldsymbol{k}})^q \Big)^{\frac{u}{q}}.
\end{aligned}$$

Applying Theorem 4.26 yields

$$\begin{aligned}
&\sigma_m(S^r_{p,\theta}B([0,1]^d),\mathbb{F}^d)_{L_q([0,1]^d)} \\
&\lesssim \sup_{\|a|s^{r,\Omega}_{p,\theta}f\|\lesssim 1} \inf_{\substack{\Lambda\subset\nabla \\ |\Lambda|\leq m}} \inf_{\substack{(\lambda_{\boldsymbol{j},\boldsymbol{k}})\subset\mathbb{C} \\ \lambda_{\boldsymbol{j},\boldsymbol{k}}\neq 0 \Longrightarrow (\boldsymbol{j},\boldsymbol{k})\in\Lambda}} \sum_{\boldsymbol{j}\in\mathbb{N}^d_{-1}} 2^{-\frac{u|\boldsymbol{j}|_1}{q}} \Big(\sum_{\boldsymbol{k}\in D_{\boldsymbol{j}}} (a_{\boldsymbol{j},\boldsymbol{k}} - \lambda_{\boldsymbol{j},\boldsymbol{k}}))^q_q \Big)^{\frac{u}{q}} \\
&\asymp \sigma_m(s^{r,\Omega}_{p,\theta}b,\mathcal{D})^u_{s^{0,\Omega}_{q,u}b}.
\end{aligned}$$

Inserting the estimate from Theorem 6.15 proves the claim. □

Remark 6.22. *This is one of the very rarely known situations where one knows the exact rate for non-linear approximation with target space* $L_\infty([0,1]^d)$.

Theorem 6.23. *Let* $\frac{1}{2} < p < q \le \infty$, $0 < \theta \le \infty$ *and*

(i) $\max\{\frac{1}{p}, \frac{1}{\theta} - \frac{1}{\max\{q,1\}}\} < r < 2$ *then*

$$m^{-r} \lesssim \sigma_m(S^r_{p,\theta}B([0,1]^d), \mathbb{F}^d)_{L_q} \lesssim m^{-r}(\log m)^{(d-1)(r+1-\frac{1}{\theta})}.$$

(ii) In case $\max\{\frac{1}{p}, \frac{1}{\theta}\} < r < 2$ *we have*

$$m^{-r} \lesssim \sigma_m(S^r_{p,\theta}F([0,1]^d), \mathbb{F}^d)_{L_q} \lesssim m^{-r}(\log m)^{(d-1)(r+1-\frac{1}{\theta})}.$$

Proof. The lower bound is due to Theorem 6.20. We prove the upper bound. Theorem 4.25 allows us to write $f \in S^r_{p,\theta}X([0,1]^d)$ as a Faber-Schauder series. This gives

$$\begin{aligned}&\sigma_m(S^r_{p,\theta}X([0,1]^d), \mathbb{F}^d)_{L_q}\\ &= \sup_{\|f|S^r_{p,\theta}X([0,1]^d)\|\le 1} \inf_{\substack{\Lambda\subset\nabla\\ |\Lambda|\le m}} \inf_{\substack{(\lambda_{\boldsymbol{j},\boldsymbol{k}})\subset\mathbb{C}\\ \lambda_{\boldsymbol{j},\boldsymbol{k}}\neq 0\Longrightarrow(\boldsymbol{j},\boldsymbol{k})\in\Lambda}} \Big\| \sum_{\boldsymbol{j}\in\mathbb{N}^d_{-1}} \sum_{\boldsymbol{k}\in D_{\boldsymbol{j}}} (d_{\boldsymbol{j},\boldsymbol{k}}(f) - \lambda_{\boldsymbol{j},\boldsymbol{k}}) v_{\boldsymbol{j},\boldsymbol{k}} \Big| L_q([0,1]^d)\Big\|.\end{aligned}$$

Lemma 4.27 provides

$$\begin{aligned}&\sigma_m(S^r_{p,\theta}X([0,1]^d), \mathbb{F}^d)_{L_q}\\ &\lesssim \sup_{\|f|S^r_{p,\theta}X([0,1]^d)\|\le 1} \inf_{\substack{\Lambda\subset\nabla\\ |\Lambda|\le m}} \inf_{\substack{(\lambda_{\boldsymbol{j},\boldsymbol{k}})\subset\mathbb{C}\\ \lambda_{\boldsymbol{j},\boldsymbol{k}}\neq 0\Longrightarrow(\boldsymbol{j},\boldsymbol{k})\in\Lambda}} \Big\| \sum_{\boldsymbol{j}\in\mathbb{N}^d_{-1}} \Big| \sum_{\boldsymbol{k}\in D_{\boldsymbol{j}}} (d_{\boldsymbol{j},\boldsymbol{k}}(f) - \lambda_{\boldsymbol{j},\boldsymbol{k}}) \chi_{\boldsymbol{j},\boldsymbol{k}} \Big| \Big| L_q([0,1]^d)\Big\|.\end{aligned}$$

Applying Theorem 4.26 yields

$$\begin{aligned}&\sigma_m(S^r_{p,\theta}X([0,1]^d), \mathbb{F}^d)_{L_q}\\ \le\; &\sup_{\|a|s^{r,\Omega}_{p,\theta}f\|\lesssim 1} \inf_{\substack{\Lambda\subset\nabla\\ |\Lambda|\le m}} \inf_{\substack{(\lambda_{\boldsymbol{j},\boldsymbol{k}})\subset\mathbb{C}\\ \lambda_{\boldsymbol{j},\boldsymbol{k}}\neq 0\Longrightarrow(\boldsymbol{j},\boldsymbol{k})\in\Lambda}} \Big\| \sum_{\boldsymbol{j}\in\mathbb{N}^d_{-1}} \Big| \sum_{\boldsymbol{k}\in D_{\boldsymbol{j}}} (a_{\boldsymbol{j},\boldsymbol{k}} - \lambda_{\boldsymbol{j},\boldsymbol{k}}) \chi_{\boldsymbol{j},\boldsymbol{k}} \Big| \Big| L_q([0,1]^d)\Big\|\\ \asymp\; &\sigma_m(s^{r,\Omega}_{p,\theta}x, \mathcal{D})_{s^{0,\Omega}_{q,1}f}.\end{aligned}$$

Inserting the estimate from Theorem 6.13 proves the claim. □

Theorem 4.30 allows us to state the following result for the limiting case $r = 2$.

Theorem 6.24. *Let* $1 < p < \infty$. *Then*

$$m^{-2} \le \sigma_m(S^2_pW([0,1]^d), \mathbb{F}^d)_{L_\infty} \lesssim m^{-2}(\log^{d-1} m)^3$$

Proof. The lower bound is due to Theorem 6.20. We prove the upper bound. Analogously to the proof of Theorem 6.21 applying Theorem 4.30 and Lemma 4.27 gives

$$\sigma_m(S^2_pW([0,1]^d, \mathbb{F}^d))_{L_\infty} \lesssim \sigma_m(s^{2,\Omega}_{p,\infty}b, \mathcal{D})_{s^{0,\Omega}_{\infty,1}}.$$

Theorem 6.13 allows to bound this from above by

$$\sigma_m(S^2_pW([0,1]^d, \mathbb{F}^d))_{L_\infty} \lesssim m^{-2}(\log^{d-1} m)^3$$

which concludes the proof. □

Finally we consider situations with smoothness in the target spaces.

Theorem 6.25. *Let* $0 < p, q \le \infty$, $0 < \theta, \nu \le \infty$ *with* $\max\left\{0, \frac{1}{\min\{p,\theta\}} - \frac{1}{\max\{q,\nu\}}\right\} < r - t$ *and*

(i) $\max\{\frac{1}{p}, \frac{1}{\theta}\} < r < 1 + \min\{\frac{1}{p}, \frac{1}{\theta}\}$, $\max\{\frac{1}{q}, \frac{1}{\nu}\} < t < 1 + \min\{\frac{1}{q}, \frac{1}{\nu}\}$. *Then*

$$\sigma_m(S^r_{p,\theta}F([0,1]^d), \mathbb{F}^d)_{S^t_{q,\nu}F([0,1]^d)} \asymp m^{-(r-t)}(\log m)^{(d-1)((r-t)-(\frac{1}{\theta}-\frac{1}{\nu}))}.$$

(ii) $\frac{1}{p} < r < 1 + \frac{1}{p}$, $\frac{1}{q} < t < 1 + \frac{1}{q}$. *Then*

$$\sigma_m(S^r_{p,\theta}F([0,1]^d), \mathbb{F}^d)_{S^t_{q,\nu}B([0,1]^d)} \asymp m^{-(r-t)}(\log m)^{(d-1)((r-t)-(\frac{1}{\theta}-\frac{1}{\nu}))}.$$

(iii) $\max\{\frac{1}{p}, \frac{1}{\theta}\} < r < 1 + \min\{\frac{1}{p}, \frac{1}{\theta}\}$, $\frac{1}{q} < t < 1 + \frac{1}{q}$. *Then*

$$\sigma_m(S^r_{p,\theta}F([0,1]^d), \mathbb{F}^d)_{S^t_{q,\nu}B([0,1]^d)} \asymp m^{-(r-t)}(\log m)^{(d-1)((r-t)-(\frac{1}{\theta}-\frac{1}{\nu}))}.$$

(iv) $\frac{1}{p} < r < 1 + \frac{1}{p}$, $\max\{\frac{1}{q}, \frac{1}{\nu}\} < t < 1 + \min\{\frac{1}{q}, \frac{1}{\nu}\}$. *Then*

$$\sigma_m(S^r_{p,\theta}B([0,1]^d), \mathbb{F}^d)_{S^t_{q,\nu}F([0,1]^d)} \asymp m^{-(r-t)}(\log m)^{(d-1)((r-t)-(\frac{1}{\theta}-\frac{1}{\nu}))}.$$

Proof. Let $X, Y \in \{F, B\}$ and $x, y \in \{f, b\}$. Theorem 4.25 allows us to write $f \in S^r_{p,\theta}X([0,1]^d)$ as a Faber-Schauder series. The equivalent norms in Theorem 4.25 provide

$$\begin{aligned}
&\sigma_m(S^r_{p,\theta}X([0,1]^d), \mathbb{F}^d)_{S^t_{q,\nu}Y([0,1]^d))} \\
&= \sup_{\|f|S^r_{p,\theta}X([0,1]^d)\|\le 1} \inf_{\substack{\Lambda\subset\nabla, |\Lambda|\le m \\ (\lambda_{\boldsymbol{j},\boldsymbol{k}})\subset\mathbb{C} \\ \lambda_{\boldsymbol{j},\boldsymbol{k}}\neq 0 \Longrightarrow (\boldsymbol{j},\boldsymbol{k})\in\Lambda}} \Big\| \sum_{\boldsymbol{j}\in\mathbb{N}^d_{-1}} \sum_{\boldsymbol{k}\in D_{\boldsymbol{j}}} (d_{\boldsymbol{j},\boldsymbol{k}}(f) - \lambda_{\boldsymbol{j},\boldsymbol{k}}) v_{\boldsymbol{j},\boldsymbol{k}} \Big\|_{S^t_{q,\nu}Y([0,1]^d)} \\
&\asymp \sup_{\|a|s^{r,\Omega}_{p,\theta}x\|\lesssim 1} \inf_{\substack{\Lambda\subset\nabla, |\Lambda|\le m \\ (\lambda_{\boldsymbol{j},\boldsymbol{k}})\subset\mathbb{C} \\ \lambda_{\boldsymbol{j},\boldsymbol{k}}\neq 0 \Longrightarrow (\boldsymbol{j},\boldsymbol{k})\in\Lambda}} \|a - \lambda|s^{t,\Omega}_{q,\nu}y\| \\
&\asymp \sigma_m(s^{r,\Omega}_{p,\theta}x, \mathcal{D})_{s^{t,\Omega}_{q,1}y}.
\end{aligned}$$

Inserting the estimate from Theorem 6.13 proves the claim. □

Remark 6.26. *Applying Theorems 6.15, 6.16, 6.17 and 6.18 similar results can be formulated for the case of small smoothness, i.e.* $\frac{1}{p} - \frac{1}{q} < r - t < \frac{1}{\theta} - \frac{1}{\nu}$. *Since this translation is straight forward we leave this to the reader.*

Remark 6.27. *Applying Algorithm 1 and 2 to approximate a function* $f \in S^r_{p,\theta}X([0,1]^d)$ *means approximating a finite part of the sequence of Faber-Schauder coefficients of* f *by* m *Faber-Schauder coefficients. Faber-Schauder coefficients are build on point evaluations, cf. (4.1.2). For that reason our results can be interpreted as non-linear adaptive sampling approximations of* f. *The input of our method is a finite number of samples of* f *from which we choose the Faber-Schauder coefficients.*

6.5 Important special cases

Let us explicitely discuss some special cases hidden in the scales of Besov-Triebel-Lizorkin spaces in the last section. First of all we discuss the probably most natural case $S_2^rW([0,1]^d) \to L_2([0,1]^d)$ where both, the model and the target spaces are Hilbert spaces. The space $S_2^rW([0,1]^d)$ equals the space $H^r_{\mathrm{mix}}([0,1]^d)$ which is well known in numerical analysis, cf. for instance [8], [9]. Theorem 6.23 yields the following:

Corollary 6.28. *Let* $\frac{1}{2} < r < 2$. *Then*

$$m^{-r} \lesssim \sigma_m(S_2^rW([0,1]^d), \mathbb{F}^d)_{L_2} \lesssim (m^{-1}\log^{d-1} m)^r(\log^{d-1} m)^{\frac{1}{2}} \tag{6.5.1}$$

holds for all $m \in \mathbb{N}$.

This means we can prove the same upper bound as for sparse grid approximation in Theorem 5.14. For dictionaries consisting out of Daubechies wavelets $\mathbb{D}^d$ it is known that

$$\sigma_m(S_2^rW([0,1]^d), \mathbb{D}^d)_{L_2} \asymp (m^{-1}\log^{d-1} m)^r$$

holds, cf. [56]. Due to missing moment conditions of the Faber-Schauder system which go together with missing L_2-orthogonality we expect slower or at least equal approximation rates as in the case of Daubechies wavelets. So the open problem for the gap in the corollary above reduces in some sense to the question whether $(\log^{d-1} m)^{\frac{1}{2}}$ is neccesary for the upper bound in (6.5.1) or not. This is closely related to an open problem for linear sampling recovery discussed in Section 9.1. For $\mathbb{D}^d$ being the dictionary of Daubechies wavelets it is well known that one does not benefit from the available non-linearity in the algorithms. The rate can be obtained by simple hyperbolic cross approximation [19]. Next we discuss the embedding $S_2^rW([0,1]) \to L_\infty([0,1]^d)$. Again, Theorem 6.23 yields

Corollary 6.29. *Let* $\max\{\frac{1}{p}, \frac{1}{2}\} < r < 2$. *Then*

$$m^{-r} \lesssim \sigma_m(S_2^rW([0,1]^d), \mathbb{F}^d)_{L_\infty} \lesssim (m^{-1}\log^{d-1} m)^r(\log^{d-1} m)^{\frac{1}{2}} \tag{6.5.2}$$

holds for all $m \in \mathbb{N}$.

Obviously we have the same bounds as above, where we measure the error in $L_2([0,1]^d)$ with the difference that $L_\infty([0,1]^d)$ is a much stronger error criterion. The comparison of both corollaries shows us a general effect for non-linear approximation in the sense of best m-term widths. The asymptotic main rates do not depend on the integrability in the source and target spaces as it is the case for linear approximation (cf. Section 5.2, where we have

$$g_m^{SG}(S_2^rW([0,1]^d), L_\infty([0,1]^d)) \asymp (m^{-1}\log^{d-1} m)^{r-\frac{1}{2}}(\log^{d-1} m)^{\frac{1}{2}}$$

for sparse grid widths. In case of best m-term approximation with respect to the Faber-Schauder dictionary the main rate depends only on the difference of the smoothness between both spaces. Studying Sobolev spaces $S_p^rW([0,1]^d)$, $p \neq 2$ the last corollary extends to:

Corollary 6.30. *Let* $1 < p < \infty$ *(*$q = \infty$*) with* $\max\{\frac{1}{p}, \frac{1}{2}\} < r < 2$*. Then*

$$m^{-r} \lesssim \sigma_m(S^r_pW([0,1]^d), \mathbb{F}^d)_{L_\infty} \lesssim (m^{-1}\log^{d-1} m)^r(\log^{d-1} m)^{\frac{1}{2}}$$

It turns out that it is important to study best m-term approximation with sequence spaces of Triebel-Lizorkin type directly. In case $p > 2$ the simple embedding $S^r_pW([0,1]^d) \hookrightarrow S^r_{p,p}B([0,1]^d)$ would yield

$$\sigma_m(S^r_pW([0,1]^d), \mathbb{F}^d)_{L_\infty} \leq \sigma_m(S^r_{p,p}B([0,1]^d), \mathbb{F}^d)_{L_\infty} \lesssim (m^{-1}\log^{d-1} m)^r(\log^{d-1} m)^{1-\frac{1}{p}}$$

which can be improved as stated in the corollary above. Finally we leave behind even the Banach space setting in the model space and consider smaller spaces $S^r_{p,\theta}B([0,1]^d)$ with $\theta < 1$. Remember, for fine index $\theta = 2$ we have the identity $S^r_2W([0,1]^d) = S^r_{2,2}B([0,1]^d) = H^r_{\text{mix}}([0,1]^d)$ in the sense of equivalent norms. Due to Lemma 3.4 modifications in the fine index cause the smallest changes within the scale of Besov-Triebel-Lizorkin spaces. Theorem 6.21 provides for this spaces:

Corollary 6.31. *Let* $0 < \theta < 1$ *with* $\frac{1}{2} < r < \min\{\frac{1}{\theta} - 1, 2\}$ *or* $\frac{1}{2} < r = \frac{1}{\theta} - 1 < 2$*. Then*

$$\sigma_m(S^r_{2,\theta}B([0,1]^d), \mathbb{F}^d)_{L_\infty} \asymp m^{-r}.$$

We observe two important effects. First there is no d-dependent logarithm in the rate which means this result behaves asymptotically like a univariate one. Second, our lower bound in Theorem 6.20 becomes sharp. For sampling recovery or even linear approximation sharp rates are unknown in literature for this parameter constellation. For that reason we compare to the sampling width for sparse grid approximation obtained in [35, Theorem 5.1]:

$$g^{SG}_m(S^r_{2,\theta}B([0,1]^d), L_\infty([0,1]^d)) \asymp (m^{-1}\log^{d-1} m)^{r-\frac{1}{2}}, \quad 0 < \theta \leq 1.$$

Here we have a main rate that depends on the integrability in the target space and additionally a d-dependent logarithm in m. In fact, the non-periodic approximation in the sense of best m-term approximation guarantees much faster approximation rates than sparse grid approximation. Last but not least we obtain from Theorem 6.24 the following corollary for the well known space $S^r_2W([0,1]^d) = H^r_{\text{mix}}([0,1]^d)$ with smoothness $r = 2$.

Corollary 6.32. *We have*

$$m^{-2} \lesssim \sigma_m(S^2_2W([0,1]^d), \mathbb{F}^d)_{L_\infty} \lesssim m^{-2}(\log^{d-1} m)^3$$

for all $m \in \mathbb{N}$.

In fact, we proved an upper bound with a worse behaving d-dependent logarithm compared to the situation where $r < 2$. In Section 6.3 we presented approximation strategies in sequence spaces. Finally, let us present for the special case $S^2_pW([0,1]^d \to L_\infty([0,1]^d)$ the corresponding sampling strategy which generates the m-term approximations for functions $f \in S^2_pW([0,1]^d$ with $\|f|S^2_pW([0,1]^d)\| \leq 1$.

Algorithm 3 $S^2_2 W \to L_\infty$ m-term approximation

Input: m degrees of freedom,

choose M such that

$$m \asymp M^{d-1} 2^M.$$

set

$$L := \left\lceil 2M + (d-1)\log M - 1 \right\rceil,$$
$$f_m := 0.$$

sample

$$\{f(x_{\boldsymbol{j},\boldsymbol{k}}) : |\boldsymbol{j}|_1 \le L, \boldsymbol{k} \in D_{\boldsymbol{j}}\}.$$

compute

$$\{d_{\boldsymbol{j},\boldsymbol{k}}(f) : |\boldsymbol{j}|_1 \le L, \boldsymbol{k} \in D_{\boldsymbol{j}}\}.$$

according to (4.1.2).

for each $\mu \in 0, \ldots, L$ ***do***

set

$$m_\mu \asymp \begin{cases} 2^\mu \mu^{d-1} & : \quad 0 \le \mu \le M, \\ \lfloor 2^\mu 2^{2(M-\mu)} \mu^{d-1} \rfloor & : \quad M+1 \le \mu \le L. \end{cases}$$

for each $\boldsymbol{j}$ ***with*** $|\boldsymbol{j}|_1 = \mu$ ***do***

for each $\boldsymbol{k} \in D_{\boldsymbol{j}}$ ***do***

if $|d_{\boldsymbol{j},\boldsymbol{k}}(f)|^2 m_\mu \ge 1$ ***then***

set

$$f_m \ := \ f_m + d_{\boldsymbol{j},\boldsymbol{k}}(f) v_{\boldsymbol{j},\boldsymbol{k}}.$$

end if

end for

end for

end for

Output: f_m best m-term approximation to f.

Chapter 7

Discrete Littlewood-Paley type characterizations of multivariate periodic functions

In the last part we had to deal with several restrictions concerning regularity and integrability of the considered model and target spaces. In the upcoming chapter we restrict to the periodic setting and prove a new kind of trigonometric sampling representation that is able to overcome most of this restrictions. This part is already published in [11].

7.1 Univariate fundamental interpolants

In this section we construct univariate sampling operators of type (1.5.2) based on bandlimited kernels $K : \mathbb{R} \to \mathbb{C}$ with suitable decay. Here $K^L_{\pi,j}$ denotes the 2π-periodization of $K^L(2^j(\cdot))$ which we will call fundamental interpolant. The following construction allows to arrange any prescribed polynomial decay (of order L) of the kernel K, which is crucial for our analysis. In addition the operator I^L_j is supposed to reproduce trigonometric polynomials of a degree related to $\asymp 2^j$. The sampling kernels we study are constructed from a finite product of dilated sinc functions. As a starting point we define for $L \in \mathbb{N}$,

$$K^L(x) := \prod_{\ell=1}^{L} \operatorname{sinc}(2^{-\ell}x), \quad x \in \mathbb{R},$$

with

$$\operatorname{sinc}(x) := \begin{cases} \frac{\sin(x)}{x} & : \quad x \neq 0, \\ 1 & : \quad \textit{otherwise}. \end{cases}$$

The next step is a 2π-periodization of dyadic dilations of $K^L(x)$ given by

$$K^L_{\pi,j}(x) := \sum_{k=-\infty}^{\infty} K^L(2^j(x + 2\pi k)). \tag{7.1.1}$$

In case $L = 1$ the summation in (7.1.1) is replaced by

$$K^1_{\pi,j}(x) := 2^{-j}\mathcal{D}^1_j(x) := 2^{-j}\sum_{k=-2^{j-1}}^{2^{j-1}-1} e^{ikx}.$$

This kernel represents an exception and requires some extra attention in the next chapters. It is a convenient modification of the classical Dirichlet kernel that provides a nested set of zeros as j increases, cf. [33, (2.6)]. For $j \in \mathbb{N}_0$ we define the interpolation operator

$$I^L_j[f](x) := \sum_{u=-2^{j-1}}^{2^{j-1}-1} f\Big(\frac{2\pi u}{2^j}\Big)K^L_{\pi,j}\Big(x - \frac{2\pi u}{2^j}\Big),$$

where in case $j = 0$ we put $I^L_0[f](x) := f(0)K^L_{\pi,0}(x)$. The kernel defined in (7.1.1) consists of a sum with infinitely many summands. For practical reasons such a definition is not useful. For every fixed $L \in \mathbb{N}$ we can compute an explicit representation of the kernel. Beginning from the definition we obtain the following identity

$$K^L_{\pi,j}(x) = \sum_{k=-\infty}^{\infty}\prod_{\ell=1}^{L}\frac{\sin(2^{j-\ell}(x+2\pi k))}{2^{j-\ell}(x+2\pi k)}.$$

Obviously, in case $x \mod \pi = 0$ we obtain

$$K^L_{\pi,j}(0) = K^L(0) = 1.$$

In case $0 < |x| < \pi$ an elementary calculation shows

$$\begin{aligned} K^L_{\pi,j}(x) &= \sum_{k=-\infty}^{\infty}\prod_{\ell=1}^{L}\frac{\sin(2^{j-\ell}x)\cos(2^{\ell+j+1}\pi k) + \sin(2^{\ell+j+1}\pi k)\cos(2^{j-\ell}x)}{2^{j-\ell}(x+2\pi k)} \\ &= \sum_{k=-\infty}^{\infty}\prod_{\ell=1}^{L}\frac{\sin(2^{j-\ell}x)}{2^{j-\ell}(x+2\pi k)} = \frac{\sin(2^{j-1}x)\dots\sin(2^{j-L}x)}{2^{jL}2^{-\frac{(L+1)L}{2}}}\sum_{k=-\infty}^{\infty}\frac{1}{(x+2\pi k)^L}. \end{aligned}$$

Using the so-called Herglotz-trick (Eisenstein series) (cf. [1]) we find the identity

$$\frac{1}{2}\cot\Big(\frac{x}{2}\Big) = \sum_{k=-\infty}^{\infty}\frac{1}{x+2\pi k}. \tag{7.1.2}$$

Taking $L-1$ derivatives yields

$$\Big[\frac{1}{2}\cot\Big(\frac{\cdot}{2}\Big)\Big]^{(L-1)}(x) = (-1)^{L-1}(L-1)!\sum_{k=-\infty}^{\infty}\frac{1}{(x+2\pi k)^L}.$$

Computing $\Big[\frac{1}{2}\cot\Big(\frac{\cdot}{2}\Big)\Big]^{(L-1)}$ and inserting this identity in (7.1.2) gives us a closed representation of the kernel $K^L_{\pi,j}(x)$. For $L = 2$ and $L = 3$ we obtain the explicit representations

$$K^2_{\pi,j}(x) = \begin{cases} \frac{2\sin(2^{j-1}x)\sin(2^{j-2}x)}{2^{2j}\sin^2(\frac{x}{2})} & : \quad x \mod 2\pi \neq 0, \\ 1 & : \quad \textit{otherwise}, \end{cases} \tag{7.1.3}$$

and

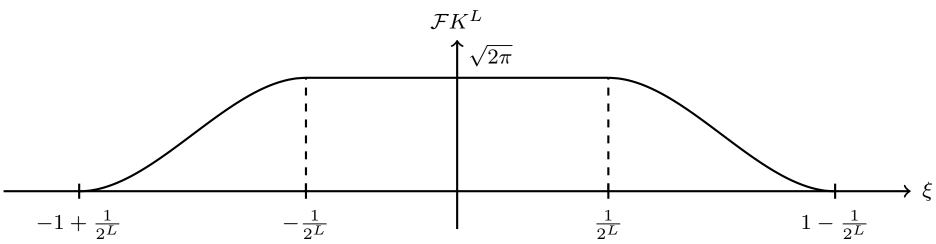
$$K^3_{\pi,j}(x) = \begin{cases} 8\frac{\sin(2^{j-1}x)\sin(2^{j-2}x)\sin(2^{j-3}x)\cos(\frac{x}{2})}{2^{3j}\sin^3(\frac{x}{2})} & : \quad x \mod 2\pi \neq 0, \\ 1 & : \quad \textit{otherwise}. \end{cases}$$

Remark 7.1. *K^L, $L > 1$, consists of products of dilated* sinc *functions. The convolution property of the Fourier transform yields*

$$K^L(x) = \prod_{\ell=1}^{L} \operatorname{sinc}(2^{-\ell}x) = \sqrt{2\pi}\mathcal{F}\Big[\chi^*_{[-2^{-1},2^{-1}]} * \ldots * 2^{L-1}\chi^*_{[-2^{-L},2^{-L}]}(\cdot)\Big](x) \qquad (7.1.4)$$

Altogether $\mathcal{F}K^L$ is a locally supported $L-2$ times continuously differentiable function fulfilling

$$\mathcal{F}K^L(\xi) = \begin{cases} \sqrt{2\pi} & : |\xi| \leq \frac{1}{2^L}, \\ 0 & : |\xi| \geq 1 - \frac{1}{2^L}. \end{cases}$$



Lemma 7.2. *Let $L \geq 1$, $j \in \mathbb{N}_0$ and $f \in C(\mathbb{T})$.*

(i) Then for $\ell \in \mathbb{Z}$

$$\widehat{I^L_j[f]}(\ell) = \frac{1}{\sqrt{2\pi}}\mathcal{F}K^L\Big(\frac{\ell}{2^j}\Big)\sum_{u=-2^{j-1}}^{2^{j-1}-1} f\Big(\frac{2\pi u}{2^j}\Big)e^{-i\frac{2\pi u}{2^j}\ell} \qquad (7.1.5)$$

holds true.

(ii) If additionally $\sum_{\ell\in\mathbb{Z}}|\widehat{f}(\ell)| < \infty$ is fulfilled. Then

$$\widehat{I^L_j[f]}(\ell) = \frac{1}{\sqrt{2\pi}}\mathcal{F}K^L\Big(\frac{\ell}{2^j}\Big)\sum_{k\in\mathbb{Z}}\widehat{f}(\ell+2^jk)$$

holds.

Proof. We compute the ℓ-th Fourier coefficient of f and obtain by the translation property the following identity

$$\widehat{I^L_j f}(\ell) \;=\; \sum_{u=-2^{j-1}}^{2^{j-1}-1} f\Big(\frac{2\pi u}{2^j}\Big)\widehat{K^L_{\pi,j}\Big(\cdot - \frac{2\pi u}{2^j}\Big)}(\ell) = \widehat{K^L_{\pi,j}}(\ell)\sum_{u=-2^{j-1}}^{2^{j-1}-1} f\Big(\frac{2\pi u}{2^j}\Big)e^{-i\frac{2\pi u}{2^j}\ell}.$$

Lemma B.19 together with the dilation property of the Fourier transform yields

$$\widehat{I_j^L f}(\ell) = \frac{1}{\sqrt{2\pi}2^j}\mathcal{F}K^L\Big(\frac{\ell}{2^j}\Big)\sum_{u=-2^{j-1}}^{2^{j-1}-1} f\Big(\frac{2\pi u}{2^j}\Big)e^{-i\frac{2\pi u}{2^j}\ell}.$$

If the Fourier coefficients are absolutely summable we get

$$\widehat{I_j^L f}(\ell) = \frac{1}{\sqrt{2\pi}2^j}\mathcal{F}K^L\Big(\frac{\ell}{2^j}\Big)\sum_{u=-2^{j-1}}^{2^{j-1}-1}\left(\sum_{k\in\mathbb{Z}}\widehat{f}(k)e^{ik\frac{2\pi u}{2^j}}\right)e^{-i\frac{2\pi u}{2^j}\ell}.$$

Interchanging the order of summation yields

$$\widehat{I_j^L f}(\ell) = \frac{1}{\sqrt{2\pi}2^j}\mathcal{F}K^L\Big(\frac{\ell}{2^j}\Big)\sum_{k\in\mathbb{Z}}\widehat{f}(k)\sum_{u=-2^{j-1}}^{2^{j-1}-1} e^{i\frac{2\pi u}{2^j}(k-\ell)}.$$

The formula for geometric partial sums tells us

$$\sum_{u=-2^{j-1}}^{2^{j-1}-1} e^{i\frac{2\pi u}{2^j}(k-\ell)} = \begin{cases} 2^j & : \quad k-\ell \mod 2^j = 0 \\ 0 & : \quad \text{otherwise.}\end{cases}$$

Finally, we obtain

$$\widehat{I_j^L f}(\ell) = \frac{1}{\sqrt{2\pi}}\mathcal{F}K^L\Big(\frac{\ell}{2^j}\Big)\sum_{k\in\mathbb{Z}}\widehat{f}(\ell+2^jk).$$

□

Definition 7.3. *We define for $j, L \in \mathbb{N}_0$ the dyadic blocks*

$$\mathcal{P}_j^L := \Big\{k\in\mathbb{Z} : |k| \le \frac{1}{2^L}2^j\Big\}. \tag{7.1.6}$$

Additionally, we denote the set of trigonometric polynomials with frequencies in $\mathcal{P}_j^L$ by

$$\mathcal{T}_j^L := span\{e^{ikx} : k\in\mathcal{P}_j^L\}.$$

Corollary 7.4. *Let $L\in\mathbb{N}$ and $f\in C(\mathbb{T})$.*

(i) Then it holds $I_j^L[f]\in\mathcal{T}_j^0$.

(ii) If additionally $f\in\mathcal{T}_j^L$ then $I_j^L[f] = f$.

Proof. Assertion (i) is an easy consequence of (7.1.5) together with the support properties of K^L. For assertion (ii) we may use

$$\widehat{I_j^L[f]}(\ell) = \frac{1}{\sqrt{2\pi}}\mathcal{F}K^L\Big(\frac{\ell}{2^j}\Big)\sum_{k\in\mathbb{Z}}\widehat{f}(\ell+2^jk)$$

which equals $\widehat{f}(\ell)$ for all ℓ if $f\in\mathcal{T}_j^L$. □

The next lemma provides the reason for calling $K^L_{\pi,j}$ a fundamental interpolant for the equidistant grid $\mathcal{G}^1_j := \{\frac{-2\pi 2^{j-1}}{2^j}, \dots, \frac{2\pi(2^{j-1}-1)}{2^j}\}$.

Lemma 7.5. *Let $f \in C(\mathbb{T})$ and $L \geq 1$. Then*

$$f\Big(\frac{2\pi u}{2^j}\Big) = I^L_j f\Big(\frac{2\pi u}{2^j}\Big) \quad , \quad u \in \{-2^{j-1}, \dots, 2^{j-1}-1\}\,.$$

Proof. Obviously, it is sufficient to proof

$$K^L_{\pi,j}\Big(\frac{2\pi u}{2^j}\Big) = \delta_{0,u}.$$

In case $L = 1$ this is a trivial consequence of (1.5.3). In case $L > 1$ we have according to our definition for $u \in \{-2^{j-1}, \dots, 2^{j-1}-1\}$

$$K^L_{\pi,j}\Big(\frac{2\pi u}{2^j}\Big) = \sum_{k=-\infty}^{\infty} K^L(2\pi(u + 2^j k)) = \delta_{0,u}\,.$$

□

Lemma 7.6. *Let $j \in \mathbb{N}_0$ and $L > 1$. Then there are constants $C, C^* > 0$ (independent of x and j) such that*

$$|K^L_{\pi,j}(x)| \leq C \min\Big\{\frac{1}{|2^j x|^L}, 1\Big\} \leq C^* \frac{1}{(1 + 2^j|x|)^L}$$

holds for all $x \in [-\pi, \pi]$.

Proof. The second inequality of the chain is trivial. We prove the first one. Starting for $x \in [-\pi, \pi]$ the estimate with

$$\begin{aligned} |K^L_{\pi,j}(x)| &= \Big|\sum_{k=-\infty}^{\infty} K^L(2^j(x + 2\pi k))\Big| \leq |K^L(2^j x)| + \sum_{|k|>0} |K^L(2^j(x + 2\pi k))| \\ &\lesssim \prod_{\ell=1}^{L} |\mathrm{sinc}\,(2^{j-\ell}x)| + \sum_{|k|>0} \frac{1}{2^{jL}|x + 2\pi k|^L}. \end{aligned} \qquad (7.1.7)$$

Clearly, the first summand is uniformly bounded. Estimating the second summand in (7.1.7) we use the fact that $|x| \leq \pi$ implies $|2\pi k + x| \geq |\pi k|$ for every integer $k \in \mathbb{Z}$ and obtain

$$\sum_{|k|>0} \frac{1}{2^{jL}|x + 2\pi k|^L} \leq \frac{1}{2^{jL}\pi^L} \sum_{|k|>0} \frac{1}{|k|^L},$$

which is known to be finite for $L \geq 2$. Using $|x| \leq \pi$ yields

$$\sum_{|k|>0} \frac{1}{2^{jL}|x + 2\pi k|^L} \lesssim \frac{1}{2^{jL}|x|^L}.$$

Considering again the first summand in (7.1.7) gives

$$\prod_{\ell=1}^{L} |\mathrm{sinc}\,(2^{j-\ell}x)| \quad \leq \quad \frac{1}{2^{(L+1)\frac{L}{2}}} \frac{1}{2^{jL}|x|^L},$$

which concludes the proof. □

7.2 Multivariate interpolation

Based on the univariate interpolation scheme from the previous subsection we are now able to define the building blocks used for the Smolyak algorithm, cf. (1.2.4),

$$q_{\boldsymbol{j}}^L[f](\boldsymbol{x}) := \Big(\bigotimes_{i=1}^{d} \eta_{j_i}^L\Big)[f](\boldsymbol{x}) \quad \text{with} \quad \eta_{j_i}^L := \begin{cases} I_{j_i}^L - I_{j_i-1}^L & : \quad j_i > 0, \\ I_0^L & : \quad j_i = 0. \end{cases} \tag{7.2.1}$$

We may write $q_{\boldsymbol{j}}^L[f]$ as follows

$$q_{\boldsymbol{j}}^L[f] = \sum_{\boldsymbol{b}\in\{-1,0\}^d} \varepsilon_{\boldsymbol{b}} I_{\boldsymbol{j}+\boldsymbol{b}}^L[f] \tag{7.2.2}$$

with suitable signs $\varepsilon_{\boldsymbol{b}}$. The definition of the operators $I_{\boldsymbol{j}+\boldsymbol{b}}^L[f](\boldsymbol{x})$ requires some more notation.

$$\boldsymbol{x}_{\boldsymbol{u}}^{\boldsymbol{j}} = \left(x_{u_1}^{j_1}, \ldots, x_{u_d}^{j_d}\right) \quad , \quad \boldsymbol{u} \in \mathbb{Z}^d,$$

where $x_u^j = 2\pi u/2^j$ for $u \in \mathbb{Z}$. For $\boldsymbol{x} \in \mathbb{R}^d$ let further

$$A_{\boldsymbol{j}}(\boldsymbol{x}) := A_{j_1}(x_1) \times \ldots \times A_{j_d}(x_d) \tag{7.2.3}$$

with $A_j(x) = \{u \in \mathbb{Z} \ : \ x_u^j \in [x-\pi, x+\pi)\}$ and put $A_{\boldsymbol{j}} := A_{\boldsymbol{j}}(\boldsymbol{0})$. We further let

$$K_{\pi^d,\boldsymbol{j}}^L := \prod_{i=1}^{d} K_{\pi,j_i}^L(x_i)$$

and define the tensorized interpolation operator by

$$I_{\boldsymbol{j}}^L[f] = \sum_{\boldsymbol{u}\in A_{\boldsymbol{j}}} f(\boldsymbol{x}_{\boldsymbol{u}}^{\boldsymbol{j}}) K_{\pi^d,\boldsymbol{j}}^L(\boldsymbol{x} - \boldsymbol{x}_u^{\boldsymbol{j}})\,.$$

Lemma 7.7. *Let $\Delta \subset \mathbb{N}_0^d$ be a solid finite set meaning that $\boldsymbol{j} \in \Delta$ and $\boldsymbol{k} \leq \boldsymbol{j}$ implies $\boldsymbol{k} \in \Delta$. Then $\sum_{\boldsymbol{j}\in\Delta} q_{\boldsymbol{j}}^L[f]$ reproduces trigonometric polynomials with frequencies in*

$$\mathcal{H}_\Delta^L := \bigcup_{\boldsymbol{j}\in\Delta} \mathcal{P}_{\boldsymbol{j}}^L\,,$$

Proof. We refer to [9, Lem. 6.1]. □

Lemma 7.8. *Let $\Delta \subset \mathbb{N}_0^d$ be a solid finite set (i.e. $\boldsymbol{k} \le \boldsymbol{j}$ and $\boldsymbol{j} \in \Delta$ implies $\boldsymbol{k} \in \Delta$). Then $T_\Delta^L f := \sum_{\boldsymbol{j}\in\Delta} q_{\boldsymbol{j}}^L[f]$ interpolates f on the grid*

$$G_\Delta^d := \bigcup_{\boldsymbol{j}\in\Delta} \{\boldsymbol{x}_{\boldsymbol{u}}^{\boldsymbol{j}} \ : \ \boldsymbol{u} \in A_{\boldsymbol{j}} \ \} \quad , \tag{7.2.4}$$

that means

$$f(\boldsymbol{x}) = T_\Delta^L f(\boldsymbol{x})$$

for all $\boldsymbol{x} \in G_\Delta^d$.

Proof. The interpolation property of the univariate operator I_j^L in Lemma 7.5 immediately gives an interpolation property of the multivariate sampling operator $I_{(m_1,\dots,m_d)}^L$ on a "full grid" $\mathcal{G}_{\{\boldsymbol{j}\le\boldsymbol{m}\}}^d$. Choosing $\boldsymbol{m}$ such that $\Delta \subset \{\boldsymbol{j} \le \boldsymbol{m}\}$ and arguing similar as in Lemma [107, Lem. 4.3] gives the result. □

Definition 7.9. *For $\boldsymbol{j} \in \mathbb{N}_0^d$ and $L \in \mathbb{N}$ we tensorize the dyadic blocks defined in (7.1.6) by*

$$\mathcal{P}_{\boldsymbol{j}}^L := \mathcal{P}_{j_1}^L \cdot \ldots \cdot \mathcal{P}_{j_d}^L,$$

and define the set of trigonometric polynomials with frequencies in $\mathcal{P}_{\boldsymbol{j}}^L$ by

$$\mathcal{T}_{\boldsymbol{j}}^L := span\ \{e^{i\boldsymbol{k}\cdot\boldsymbol{x}} : \boldsymbol{k} \in \mathcal{P}_{\boldsymbol{j}}^L\}.$$

Proposition 7.10. *Let $L \in \mathbb{N}$ and $f \in \mathcal{T}_{\boldsymbol{\ell}}^L$ then $q_{\boldsymbol{j}}^L[f] \neq 0$ implies $\boldsymbol{\ell} \ge \boldsymbol{j}$.*

Proof. The proof follows immediately from the definition of $q_{\boldsymbol{j}}^L[f]$ in (7.2.1) and the univariate reproduction property in Corollary 7.4. □

7.3 Superposition of trigonometric polynomials

In this section we provide periodic counterparts for Theorems 4.16 and 4.18. We want to estimate the norm of a superposition of trigonometric polynomials

$$f = \sum_{\boldsymbol{j}\in\mathbb{N}_0^d} f_{\boldsymbol{j}}$$

where $f_{\boldsymbol{j}}$ are trigonometric polynomials of degree $\asymp \boldsymbol{2^j}$. In contrast to the usual Littlewood-Paley building blocks $\delta_{\boldsymbol{j}}^\pi[f]$ which are 'almost' orthogonal, we only need to restrict the degree of the polynomial in the sequel.

As a main tool we introduce the following componentwise variant of the Hardy-Littlewood maximal operator, see [127, (1.14),(1.15)], [122, (10)] and the references therein.

Let us now state the main result of this subsection.

Theorem 7.11. *Let* $0 < p < \infty$, $0 < \theta \leq \infty$, $\boldsymbol{r} \in \mathbb{R}^d$ *with* $\boldsymbol{r} > \sigma_{p,\theta}$ *and* $(f_{\boldsymbol{j}})_{\boldsymbol{j}\in\mathbb{N}_0^d}$ *such that* $f_{\boldsymbol{j}} \in \mathcal{T}_{\boldsymbol{j}}^0$ *and* $\|2^{\boldsymbol{r}\cdot\boldsymbol{j}} f_{\boldsymbol{j}} | L_p(\ell_\theta)\| < \infty$. *Then*

(i) $\sum_{\boldsymbol{j}\in\mathbb{N}_0^d} f_{\boldsymbol{j}}$ *converges unconditionally in* $S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d)$ *if* $\theta < \infty$ *and in every* $S^{\tilde{\boldsymbol{r}}}_{p,\nu}F(\mathbb{T}^d)$ *with* $0 < \nu \leq \infty$ *and* $\tilde{\boldsymbol{r}} < \boldsymbol{r}$.

(ii) There is a constant $C > 0$ *(independent of* f*) such that*

$$\Big\| \sum_{\boldsymbol{j}\in\mathbb{N}_0^d} f_{\boldsymbol{j}} \Big| S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d)\Big\| \leq C \big\| 2^{\boldsymbol{r}\cdot\boldsymbol{j}} f_{\boldsymbol{j}} \big| L_p(\ell_\theta)\big\|$$

holds.

Proof. Step 1. We assume the unconditional convergence of $\sum_{\boldsymbol{\ell}\in\mathbb{N}_0^d} f_{\boldsymbol{\ell}}$ in $S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d)$ (or in case $\theta = \infty$ at least in $S^{\tilde{\boldsymbol{r}}}_{p,\nu}F(\mathbb{T}^d)$) and prove the inequality

$$\Big\| \sum_{\boldsymbol{\ell}\in\mathbb{N}_0^d} f_{\boldsymbol{\ell}} \Big| S^{r}_{p,\theta}F(\mathbb{T}^d)\Big\| \lesssim \big\| 2^{\boldsymbol{r}\cdot\boldsymbol{j}} f_{\boldsymbol{j}} \big| L_p(\ell_\theta(\mathbb{N}_0^d))\big\|.$$

We mimic Step 1 of the proof of [122, Theorem 3.4.1]. This is rather technical in the multivariate situation. For that reason we give a proof for the univariate situation first. Later we explain the necessary modifications for the multivariate situation. We prove

$$\|f|F^r_{p,\theta}(\mathbb{T})\| \lesssim \big\| 2^{rj} f_j \big| L_p(\ell_\theta(\mathbb{N}_0))\big\|$$

by using methods from difference characterization of Triebel-Lizorkin spaces. We start by switching to the difference norm in $F^r_{p,\theta}(\mathbb{T})$ with $m > r$

$$\Big\| \sum_{\ell\in\mathbb{N}_0} f_\ell \Big| F^r_{p,\theta}(\mathbb{T})\Big\| \asymp \Big\| \sum_{\ell\in\mathbb{N}_0} f_\ell \Big| L_p(\mathbb{T})\Big\| + \Big\| \Big[\sum_{j=0}^\infty 2^{\theta jr} \Big(2^j \int_{-2^{-j}}^{2^{-j}} \Big| \Delta_h^m \Big[\sum_{\ell\in\mathbb{N}_0} f_\ell \Big] \Big| dh \Big)^\theta \Big]^{\frac{1}{\theta}} \Big| L_p(\mathbb{T})\Big\|. \tag{7.3.1}$$

First we estimate the L_p-norm of f and obtain trivially using either Hölder's inequality (in case $\theta \geq 1$) or the embedding $\ell_\theta \hookrightarrow \ell_1$ (in case $0 < \theta < 1$) the estimate

$$\Big\| \sum_{\ell\in\mathbb{N}_0} f_\ell \Big| L_p(\mathbb{T})\Big\| \lesssim \Big\| \Big(\sum_{j\in\mathbb{N}_0} 2^{\theta rj} |f_j|^\theta \Big)^{\frac{1}{\theta}} \Big| L_p(\mathbb{T})\Big\|.$$

Let $a > 0$ be a positive real number such that $a > \max\{\frac{1}{p}, \frac{1}{\theta}\}$ is fulfilled. Additionally choose in case $\min\{p,\theta\} \leq 1$

$$0 < \lambda < \min\{p,\theta\} \tag{7.3.2}$$

such that

$$r > (1-\lambda)a > \sigma_{p,\theta}. \tag{7.3.3}$$

This is possible since

$$\begin{aligned} (1-\lambda)a &> (1-\lambda)\max\Big\{\frac{1}{p}, \frac{1}{\theta}\Big\} \\ &\geq (1-\min\{p,\theta,1\})\max\Big\{\frac{1}{p}, \frac{1}{\theta}\Big\} = \sigma_{p,\theta}. \end{aligned}$$

In case $\min\{p,\theta\} > 1$ we simply choose $\lambda = 1$. Fix $j \in \mathbb{N}_0$ and use the identity

$$\sum_{\ell\in\mathbb{N}_0} f_\ell = \sum_{\ell\in\mathbb{Z}} f_{j+\ell}$$

with $f_{j+\ell} = 0$ for $j+\ell < 0$. The unconditional convergence of $\sum_{\ell\in\mathbb{Z}} f_{j+\ell}$ in $F^r_{p,\theta}(\mathbb{T})$ implies (by Lemma 3.4) an unconditional convergence also in $L_1(\mathbb{T})$. Therefore we can estimate the integral means as follows

$$2^j \int_{-2^{-j}}^{2^{-j}} \Big|\Delta_h^m\Big[\sum_{\ell\in\mathbb{N}_0} f_{j+\ell}\Big]\Big| dh \le \sum_{\ell\in\mathbb{Z}} 2^j \int_{-2^{-j}}^{2^{-j}} |\Delta_h^m f_{j+\ell}(x)| dh. \tag{7.3.4}$$

We split the sum over ℓ

$$\begin{aligned} &\sum_{\ell\in\mathbb{Z}} 2^j \int_{-2^{-j}}^{2^{-j}} |\Delta_h^m f_{j+\ell}(x)| dh \\ = \;& \sum_{\ell\ge 0} 2^j \int_{-2^{-j}}^{2^{-j}} |\Delta_h^m f_{j+\ell}(x)| dh + \sum_{\ell<0} 2^j \int_{-2^{-j}}^{2^{-j}} |\Delta_h^m f_{j+\ell}(x)| dh \end{aligned} \tag{7.3.5}$$

and prove

$$2^j \int_{-2^{-j}}^{2^{-j}} |\Delta_h^m f_{j+\ell}(x)| dh \lesssim \begin{cases} 2^{\ell m} P_{2^{j+\ell},a} f_{j+\ell} & : \quad \ell \ge 0, \\ 2^{(1-\lambda)\ell a} [P_{2^{j+\ell},a} f_{j+\ell}]^{1-\lambda} M|f_{j+\ell}|^\lambda & : \quad \ell < 0. \end{cases} \tag{7.3.6}$$

First we prove the case $\ell > 0$. Applying Lemma B.10 immediately gives

$$2^j \int_{-2^{-j}}^{2^{-j}} |\Delta_h^m f_{j+\ell}(x)| dh \;\lesssim\; 2^{\ell m} P_{2^{j+\ell},a}(x).$$

In case $\ell < 0$ with $\lambda < 1$ we estimate as follows

$$2^j \int_{-2^{-j}}^{2^{-j}} |\Delta_h^m f_{j+\ell}(x)| dh \;\lesssim\; 2^j \int_{-2^{-j}}^{2^{-j}} |\Delta_h^m f_{j+\ell}(x)|^\lambda |\Delta_h^m f_{j+\ell}(x)|^{1-\lambda} dh.$$

Applying Lemma B.10 to the second factor yields

$$\begin{aligned} 2^j \int_{-2^{-j}}^{2^{-j}} |\Delta_h^m f_{j+\ell}(x)| dh \;&\lesssim\; 2^{\ell(1-\lambda)a} [P_{2^{j+\ell},a} f_{j+\ell}(x)]^{1-\lambda} 2^j \int_{-2^{-j}}^{2^{-j}} |\Delta_h^m f_{j+\ell}(x)|^\lambda dh \\ &\lesssim\; 2^{\ell(1-\lambda)a} [P_{2^{j+\ell},a} f_{j+\ell}(x)]^{1-\lambda} M|f_{j+\ell}|^\lambda(x). \end{aligned}$$

Attention in case $\min\{p,\theta\} > 1$ with $\lambda = 1$ the estimate in case $\ell < 0$ simplifies to the Hardy-Littlewood maximal function of $|f_{j+\ell}|$. Inserting the decomposition in (7.3.5) together with the estimates obtained in (7.3.6) into the last term on the right hand side of (7.3.1) then we obtain by μ-triangle inequality in $L_p(\ell_\theta(\mathbb{N}))$ with $\mu := \min\{p, \theta, 1\}$

$$\begin{aligned} &\Big\| \Big[\sum_{j=0}^{\infty} 2^{\theta j r} \Big(2^j \int_{-2^{-j}}^{2^{-j}} \Big| \Delta_h^m \Big[\sum_{\ell \in \mathbb{N}_0} f_{j+\ell} \Big] \Big| dh \Big)^{\theta} \Big]^{\frac{1}{\theta}} \Big| L_p(\mathbb{T}) \Big\| \\ &\lesssim \Big[\sum_{\ell \geq 0} 2^{\mu \ell (m-r)} \| 2^{(j+\ell)r} P_{2^{j+\ell},a} f_{j+\ell}(x) | L_p(\ell_\theta(\mathbb{N}_0)) \|^{\mu} \\ &+ \sum_{\ell < 0} 2^{\mu \ell [a(1-\lambda)-r]} \| 2^{(j+\ell)r} [P_{2^{j+\ell},a} f_{2^{j+\ell}}(x)]^{1-\lambda}(x) \\ &\quad \times M |f_{j+\ell}|^{\lambda}(x) | L_p(\ell_\theta(\mathbb{N}_0)) \|^{\mu} \Big]^{\frac{1}{\mu}}. \end{aligned} \tag{7.3.7}$$

To estimate the first summand we apply Theorem B.17, which gives

$$\| 2^{(j+\ell)r} P_{2^{j+\ell},a} f_{j+\ell}(x) | L_p(\ell_\theta(\mathbb{N}_0)) \| \lesssim \| 2^{(j+\ell)r} f_{j+\ell}(x) | L_p(\ell_\theta(\mathbb{N}_0)) \|.$$

An index shift yields

$$\| 2^{(j+\ell)r} P_{2^{j+\ell},a} f_{j+\ell}(x) | L_p(\ell_\theta(\mathbb{N}_0)) \| \lesssim \| 2^{jr} f_j(x) | L_p(\ell_\theta(\mathbb{N}_0)) \|. \tag{7.3.8}$$

In case $\min\{p, \theta\} \leq 1$ with $\lambda < 1$ we apply to the norm expression in (7.3.7) Hölder's inequality with $\frac{1}{1-\lambda}$, $\frac{1}{\lambda}$ twice and obtain

$$\begin{aligned} &\| 2^{(j+\ell)r} [P_{2^{j+\ell},a} f_{j+\ell}(x)]^{1-\lambda} M |f_{j+\ell}|^{\lambda}(x) | L_p(\ell_\theta(\mathbb{N}_0)) \| \\ &\leq \| 2^{(j+\ell)r} P_{2^{j+\ell},a} f_{j+\ell}(x) | L_p(\ell_\theta(\mathbb{N}_0)) \|^{1-\lambda} \\ &\times \| 2^{(j+\ell)r} (M |f_{j+\ell}|^{\lambda}(x))^{\frac{1}{\lambda}} | L_p(\ell_\theta(\mathbb{N}_0)) \|^{\lambda}. \end{aligned} \tag{7.3.9}$$

We skip this in case $\lambda = 1$. Considering the factors in (7.3.9) separately we obtain by applying Theorem B.17

$$\| 2^{(j+\ell)r} P_{2^{j+\ell},a} f_{j+\ell}(x) | L_p(\ell_\theta(\mathbb{N}_0)) \| \lesssim \| 2^{(j+\ell)r} f_{j+\ell}(x) | L_p(\ell_\theta(\mathbb{N}_0)) \|. \tag{7.3.10}$$

For the second factor we rewrite the $L_p(\ell_\theta(\mathbb{N}_0))$-norm as a $L_{\frac{p}{\lambda}}(\ell_{\frac{\theta}{\lambda}}(\mathbb{N}_0))$-norm. This allows for applying Theorem B.14.

$$\begin{aligned} \| 2^{(j+\ell)r} (M |f_{j+\ell}|^{\lambda}(x))^{\frac{1}{\lambda}} | L_p(\ell_\theta(\mathbb{N}_0)) \| &= \| 2^{(j+\ell)r\lambda} M |f_{j+\ell}|^{\lambda}(x) | L_{\frac{p}{\lambda}}(\ell_{\frac{\theta}{\lambda}}(\mathbb{N}_0)) \|^{\frac{1}{\lambda}} \\ &\lesssim \| 2^{(j+\ell)r\lambda} |f_{j+\ell}(x)|^{\lambda} | L_{\frac{p}{\lambda}}(\ell_{\frac{\theta}{\lambda}}(\mathbb{N}_0)) \|^{\frac{1}{\lambda}} \\ &= \| 2^{(j+\ell)r} f_{j+\ell}(x) | L_p(\ell_\theta(\mathbb{N}_0)) \|. \end{aligned} \tag{7.3.11}$$

Inserting the estimates from (7.3.10) and (7.3.11) into (7.3.9) implies

$$\| 2^{(j+\ell)r} [P_{2^{j+\ell},a} f_{j+\ell}(x)]^{1-\lambda} M |f_{j+\ell}|^{\lambda}(x) | L_p(\ell_\theta(\mathbb{N}_0)) \| \leq \| 2^{(j+\ell)r} f_{j+\ell}(x) | L_p(\ell_\theta(\mathbb{N}_0)) \|.$$

A similar index shift as above yields

$$\|2^{(j+\ell)r}[P_{2^{j+\ell},a}f_{j+\ell}(x)]^{1-\lambda}M|f_{j+\ell}|^{\lambda}(x)|L_p(\ell_\theta(\mathbb{N}_0))\| \leq \|2^{jr}f_j(x)|L_p(\ell_\theta(\mathbb{N}_0))\|. \quad (7.3.12)$$

We continue estimating (7.3.7) and insert (7.3.8) and (7.3.12) to obtain

$$\Big\|\Big[\sum_{j=0}^{\infty}2^{\theta jr}\Big(2^j\int_{-2^{-j}}^{2^{-j}}\Big|\Delta_h^m\Big[\sum_{\ell\in\mathbb{N}_0}f_{j+\ell}\Big]\Big|dh\Big)^{\theta}\Big]^{\frac{1}{\theta}}\Big|L_p(\mathbb{T})\Big\|$$
$$\lesssim \|2^{jr}f_j|L_p(\ell_\theta(\mathbb{N}_0))\|\Big[\sum_{\ell\geq 0}2^{\mu\ell(m-r)}+\sum_{\ell<0}2^{\mu\ell[a(1-\lambda)-r]}\Big]^{\frac{1}{\mu}}.$$

Finally, the choice of the parameters m, a, λ in (7.3.3) yields that the series in (7.3.13) converge to a constant. Altogether we obtain the desired bound

$$\Big\|\sum_{\ell\in\mathbb{N}_0}f_\ell\Big|F^r_{p,\theta}(\mathbb{T})\Big\| \lesssim \|2^{jr}f_j|L_p(\ell_\theta(\mathbb{N}_0))\|. \quad (7.3.13)$$

Step 2. We explain the modifications in the multivariate situation. This time we start computing the norm of $\sum_{\boldsymbol{\ell}\in\mathbb{N}_0^d}f_{\boldsymbol{\ell}} \in S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d)$ in terms of differences, cf. the periodic counterpart of Theorem 3.8,

$$\Big\|\sum_{\boldsymbol{\ell}\in\mathbb{N}_0^d}f_{\boldsymbol{\ell}}\Big|S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d)\Big\| \asymp \sum_{e\subset[d]}\|f|S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d)\|_{e,\boldsymbol{m}}.$$

For each $e\subset[d]$ we have to show that

$$\Big\|\sum_{\boldsymbol{\ell}\in\mathbb{N}_0^d}f_{\boldsymbol{\ell}}\Big|S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d)\Big\|_{e,\boldsymbol{m}} \lesssim \|2^{\boldsymbol{r}\boldsymbol{j}}f_{\boldsymbol{j}}|L_p(\ell_\theta(\mathbb{N}_0^d))\|$$

holds. A full proof consists in applying the arguments from above to every single direction contained in e. Here the directionwise Hardy-Littlewood maximal function and corresponding maximal inequality come into play, see Definition B.7 and Thms. B.15, B.17. Since this requires an extensive case study in e and $\boldsymbol{\ell}$ we refer to the proof given in detail in [122, Thm. 3.4.1, Step 1] where we have to replace the decomposition of f used there by the representation $\sum_{\boldsymbol{\ell}\in\mathbb{Z}^d}f_{\boldsymbol{j}+\boldsymbol{\ell}}$.

Step 3. We prove (i) in case $\theta<\infty$. To begin with, we define the set of sequences with finite index sets given by

$$\mathfrak{E} := \Big\{\mathcal{E}=(\mathcal{E}_n)_{n\in\mathbb{N}}:\ \mathcal{E}_n\subset\mathbb{N}_0^d,\ |\mathcal{E}_n|=n,\ \mathcal{E}_n\subset\mathcal{E}_{n+1}\text{ for all }n\in\mathbb{N},\text{ and }\bigcup_{n=1}^{\infty}\mathcal{E}_n=\mathbb{N}_0^d\Big\}.$$

Every sequence in $\mathfrak{E}$ defines an order of summation. Furthermore for $\mathcal{E}\in\mathfrak{E}$ we define $F_{\mathcal{E}_n} := \sum_{\boldsymbol{j}\in\mathcal{E}_n}f_{\boldsymbol{j}}$. We take a second sequence $A\in\mathfrak{E}$ and consider $F_{\mathcal{E}_n}-F_{\mathcal{A}_m}$. This difference can be written as a sum with finitely many $f_{\boldsymbol{j}}$. This fulfills the assumptions necessary in Step 1 and yields

$$\|F_{\mathcal{E}_n}-F_{\mathcal{A}_m}|S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d)\| \lesssim \Big\|\Big(\sum_{\boldsymbol{j}\in(\mathcal{E}_n\cup\mathcal{A}_m)\setminus(\mathcal{E}_n\cap\mathcal{A}_m)}2^{\boldsymbol{r}\cdot\boldsymbol{j}\theta}|f_{\boldsymbol{j}}|^{\theta}\Big)^{\frac{1}{\theta}}\Big|L_p(\mathbb{T})\Big\|.$$

Obviously,

$$\Big(\sum_{\boldsymbol{j}\in(\mathcal{E}_n\cup\mathcal{A}_m)\setminus(\mathcal{E}_n\cap\mathcal{A}_m)} 2^{\boldsymbol{r}\cdot\boldsymbol{j}\theta}|f_{\boldsymbol{j}}|^\theta\Big)^{\frac{1}{\theta}} \le \Big(\sum_{\boldsymbol{j}\in\mathbb{N}_0^d} 2^{\boldsymbol{r}\cdot\boldsymbol{j}\theta}|f_{\boldsymbol{j}}|^\theta\Big)^{\frac{1}{\theta}} \in L_p(\mathbb{T})$$

holds almost everywhere. Therefore Lebesgue's dominated convergence theorem yields that we find for every $\varepsilon > 0$ a $n_0 \in \mathbb{N}$ such that

$$\|F_{\mathcal{E}_n} - F_{\mathcal{A}_m}|S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d)\| \le \varepsilon$$

holds for all $m, n > n_0$. Finally this implies unconditional convergence in $S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d)$. In case $\theta = \infty$ we stress on the embeddings

$$S^{\boldsymbol{s}}_{p,1}F(\mathbb{T}^d) \hookrightarrow S^{\tilde{\boldsymbol{r}}}_{p,\nu}F(\mathbb{T}^d)$$

and

$$\|2^{\boldsymbol{s}\cdot\boldsymbol{j}}f_j|L_p(\ell_1)\| \lesssim \|2^{\boldsymbol{r}\cdot\boldsymbol{j}}f_j|L_p(\ell_\infty)\|,$$

where $\boldsymbol{r} > \boldsymbol{s} > \sigma_{p,\nu}$, $\boldsymbol{s} > \tilde{\boldsymbol{r}}$ and $0 < \nu \le \infty$. Applying the arguments from above to $S^{\boldsymbol{s}}_{p,1}F(\mathbb{T}^d)$ yields the result for $S^{\tilde{\boldsymbol{r}}}_{p,\nu}F(\mathbb{T}^d)$. □

We will also need the following diagonal embedding relation which is the periodic counterpart of [99, Prop. 2.4.1], see also the diagonal embedding in Lemma 3.4, (vi) and Lemma 4.10 above.

Lemma 7.12. *Let $0 < p < q < \infty$ and $0 < \theta, \nu \le \infty$. Then*

$$\Big\|\Big(\sum_{\boldsymbol{j}\in\mathbb{N}_0^d}|f_{\boldsymbol{j}}|^\nu\Big)^{\frac{1}{\nu}}\Big|L_q(\mathbb{T}^d)\Big\| \lesssim \Big\|\Big(\sum_{\boldsymbol{j}\in\mathbb{N}_0^d}2^{\theta|\boldsymbol{j}|_1(1/p-1/q)}|f_{\boldsymbol{j}}|^\theta\Big)^{\frac{1}{\theta}}\Big|L_p(\mathbb{T}^d)\Big\|$$

holds for all $(f_{\boldsymbol{j}})_{\boldsymbol{j}\in\mathbb{N}_0^d}$ such that $f_{\boldsymbol{j}} \in \mathcal{T}^0_{\boldsymbol{j}}$.

Let us finally state the counterpart of Theorem 7.11 for the B-case.

Theorem 7.13. *Let $0 < p \le \infty$, $0 < \theta \le \infty$, $\boldsymbol{r} \in \mathbb{R}^d$ with $\boldsymbol{r} > \sigma_p$ and $(f_{\boldsymbol{j}})_{\boldsymbol{j}\in\mathbb{N}_0^d}$ such that $f_{\boldsymbol{j}} \in \mathcal{T}^0_{\boldsymbol{j}}$ and $\big\|2^{\boldsymbol{r}\cdot\boldsymbol{j}}f_{\boldsymbol{j}}\big|\ell_\theta(L_p)\big\| < \infty$. Then*

(i) $\sum_{j\in\mathbb{N}_0^d} f_{\boldsymbol{j}}$ converges unconditionally in $S^{\boldsymbol{r}}_{p,\theta}B(\mathbb{T}^d)$ if $\max\{p,\theta\} < \infty$ and in every $S^{\tilde{\boldsymbol{r}}}_{p,\nu}B(\mathbb{T}^d)$ with $0 < \nu \le \infty$ and $\tilde{\boldsymbol{r}} < \boldsymbol{r}$.

(ii) it holds

$$\Big\|\sum_{\boldsymbol{j}\in\mathbb{N}_0^d} f_{\boldsymbol{j}}\Big|S^{\boldsymbol{r}}_{p,\theta}B(\mathbb{T}^d)\Big\| \lesssim \big\|2^{\boldsymbol{r}\cdot\boldsymbol{j}}f_{\boldsymbol{j}}\big|\ell_\theta(L_p(\mathbb{T}^d)\big\|.$$

Proof. We follow the proof of Theorem 7.11 line by line and point out the necessary modifications for the B-case. To convince the reader we explain this modifications for (ii) in the univariate case. Again, we prove

$$\Big\| \sum_{\ell\in\mathbb{N}_0} f_\ell \Big| B^r_{p,\theta}(\mathbb{T})\Big\| \lesssim \|2^{rj} f_j | L_p(\ell_\theta(\mathbb{N}_0))\|$$

by using methods from difference characterization. We start by switching to the difference norm in $B^r_{p,\theta}(\mathbb{T})$ with $m > r$

$$\Big\| \sum_{\ell\in\mathbb{N}_0} f_\ell \Big| B^r_{p,\theta}(\mathbb{T})\Big\| \asymp \Big\| \sum_{\ell\in\mathbb{N}_0} f_\ell \Big| L_p(\mathbb{T}^d)\Big\| + \Big(\sum_{j=0}^{\infty} 2^{\theta jr} \Big\| 2^j \int_{-2^{-j}}^{2^{-j}} \Big| \Delta_h^m \Big[\sum_{\ell\in\mathbb{N}_0} f_\ell \Big] \Big| dh \Big\|_p^\theta \Big)^{\frac{1}{\theta}}. \tag{7.3.14}$$

First we estimate the L_p-norm of f and obtain trivially using either Hölder's inequality (in case $\min\{p,\theta\} > 1$ or $p < \min\{1,\theta\}$) or the embedding $\ell_\theta \hookrightarrow \ell_1$ (otherwise) the estimate

$$\Big\| \sum_{\ell\in\mathbb{N}_0} f_\ell \Big| L_p(\mathbb{T}^d)\Big\| \lesssim \Big(\sum_{j\in\mathbb{N}_0} 2^{\theta rj} \|f_j | L_p(\mathbb{T}^d)\|^\theta \Big)^{\frac{1}{\theta}}.$$

Let $a > 0$ be a positive real number such that $a > \frac{1}{p}$ is fulfilled. Additionally, in case $p > 1$ we choose $\lambda = 1$. Whereas in case $p \leq 1$ we choose

$$0 < \lambda < p$$

such that

$$r > (1-\lambda)a > (1-\lambda)\frac{1}{p} \geq (1-p)\frac{1}{p} = \sigma_p.$$

For the second term in (7.3.14) the estimates in (7.3.4), (7.3.5) and (7.3.6) yield

$$\begin{aligned} &\Big(\sum_{j=0}^{\infty} 2^{\theta jr} \Big\| 2^j \int_{-2^{-j}}^{2^{-j}} \Big| \Delta_h^m \Big[\sum_{\ell\in\mathbb{Z}^d} f_{j+\ell} \Big] \Big| dh \Big\|_p^\theta \Big)^{\frac{1}{\theta}} \\ &\lesssim \Big[\sum_{\ell\geq 0} 2^{\mu\ell(m-r)} \|2^{(j+\ell)r} P_{2^{j+\ell},a} f_{j+\ell}(x) | \ell_\theta(L_p(\mathbb{T}))\|^\mu \qquad (7.3.15) \\ &\quad + \sum_{\ell<0} 2^{\mu\ell[a(1-\lambda)-r]} \|2^{(j+\ell)r} [P_{2^{j+\ell},a} f_{2^{j+\ell}}(x)]^{1-\lambda} \times M|f_{j+\ell}|^\lambda(x) | \ell_\theta(L_p(\mathbb{T}))\|^\mu \Big]^{\frac{1}{\mu}} \end{aligned}$$

with $\mu = \min\{p,\theta,1\}$. The $L_p(\mathbb{T})$-norm is now the inner norm in the sequence spaces. For that reason it suffices to use simpler (non-vector valued) maximal inequalities. We apply Theorem B.16 to the first summand, which gives

$$\|2^{(j+\ell)r} P_{2^{j+\ell},a} f_{j+\ell}(x) | \ell_\theta(L_p(\mathbb{T}))\| \lesssim \|2^{(j+\ell)r} f_{j+\ell}(x) | \ell_\theta(L_p(\mathbb{T}))\|.$$

An index shift yields

$$\|2^{(j+\ell)r} P_{2^{j+\ell},a} f_{j+\ell}(x) | \ell_\theta(L_p(\mathbb{T}))\| \lesssim \|2^{jr} f_j(x) | \ell_\theta(L_p(\mathbb{T}))\|. \tag{7.3.16}$$

In case $p \leq 1$ we apply Hölder's inequality to the second summand in (7.3.15) with $\frac{1}{1-\lambda}$, $\frac{1}{\lambda}$ and obtain

$$\begin{aligned}&\|2^{(j+\ell)r}[P_{2^{j+\ell},a}f_{j+\ell}(x)]^{1-\lambda}M|f_{j+\ell}|^{\lambda}(x)|\ell_\theta(L_p(\mathbb{T}))\|^{\theta}\\ &\leq \Big(\sum_{j\in\mathbb{N}_0} 2^{\theta(j+\ell)r}\|P_{2^{j+\ell},a}f_{j+\ell}(x)|L_p(\mathbb{T})\|^{(1-\lambda)\theta}\times\|(M|f_{j+\ell}|^{\lambda}(x))^{\frac{1}{\lambda}}|L_p(\mathbb{T})\|^{\lambda\theta}\Big)^{\frac{1}{\theta}}.\end{aligned}$$

This can be skipped in case $p > 1$. Applying the maximal inequalities stated in Theorem B.16 and Theorem B.6 (together with a trick similar to (7.3.11)) yields

$$\begin{aligned}\|2^{(j+\ell)r}[P_{2^{j+\ell},a}f_{j+\ell}(x)]^{1-\lambda}M|f_{j+\ell}|^{\lambda}(x)|\ell_\theta(L_p(\mathbb{T}))\| &\lesssim \Big(\sum_{j\in\mathbb{N}_0} 2^{\theta(j+\ell)r}\|f_{j+\ell}(x)|L_p(\mathbb{T})\|^{\theta}\Big)^{\frac{1}{\theta}}\\ &\lesssim \|2^{jr}f_j(x)|\ell_\theta(L_p(\mathbb{T}))\|.\end{aligned}$$

Hence, the estimates from (7.3.16) and (7.3.17) imply

$$\begin{aligned}&\Big(\sum_{j=0}^{\infty} 2^{\theta jr}\Big\|2^j\int_{-2^{-j}}^{2^{-j}}\Big|\Delta_h^m\Big[\sum_{\ell\in\mathbb{Z}^d} f_{j+\ell}\Big]\Big|dh\Big\|_p^{\theta}\Big)^{\frac{1}{\theta}} \qquad (7.3.17)\\ &\lesssim \Big[\sum_{\ell\geq 0} 2^{\mu\ell(m-r)} + \sum_{\ell<0} 2^{\mu\ell[a(1-\lambda)-r]}\Big]^{\frac{1}{\mu}}\|2^{jr}f_j(x)|\ell_\theta(L_p(\mathbb{T}))\|.\end{aligned}$$

The choice of λ, a and m relatively to r ensures the convergence of the series to an absolute constant. This concludes the proof in the univariate case. For the multivariate situation see the comments in Step 2 of Theorem 7.11. □

7.4 Trigonometric sampling representations

Analogously to Theorem 4.25 we provide theorems that allow for replacing the Fourier analytic building blocks $\delta_{\boldsymbol{j}}[f]$ used to define the spaces $S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d)$ and $S^{\boldsymbol{r}}_{p,\theta}B(\mathbb{T}^d)$ (cf. Definition 3.18) by building blocks $q^L_{\boldsymbol{j}}[f]$ based on function evaluation. Using the short notation $L_p(\ell_\theta) = L_p(\ell_\theta(\mathbb{N}_0^d), \mathbb{T}^d)$ we will prove the following main results.

Theorem 7.14. *Let $0 < p < \infty$, $0 < \theta \leq \infty$, $L > \max\{\frac{1}{p}, \frac{1}{\theta}\}$ ($L = 1$ requires $\theta < \infty$) and $\boldsymbol{r} > \max\{\frac{1}{p}, \frac{1}{\theta}\}$ then the (quasi-)norms*

$$\|f|S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d)\| \asymp \|2^{\boldsymbol{r}\cdot\boldsymbol{j}}q^L_{\boldsymbol{j}}(f)|L_p(\ell_\theta)\|$$

are equivalent for all $f \in S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d)$.

Proof. The result is a consequence of Theorem 7.19 together with Theorem 7.11. For the case $L = 1$ we refer to Theorem 7.26. □

For the B-case weaker conditions on $\boldsymbol{r}$ and L are sufficient.

Theorem 7.15. *Let $0 < p, \theta \le \infty$, $L > \frac{1}{p}$ ($L = 1$ requires $p < \infty$) and $\boldsymbol{r} > \frac{1}{p}$ then the (quasi-)norms*

$$\|f|S^{\boldsymbol{r}}_{p,\theta}B(\mathbb{T}^d)\| \asymp \|2^{\boldsymbol{r}\cdot\boldsymbol{j}} q^L_{\boldsymbol{j}}(f)|\ell_\theta(L_p)\|$$

are equivalent for all $f \in S^{\boldsymbol{r}}_{p,\theta}B(\mathbb{T}^d)$.

Proof. The proof is a consequence of Theorem 7.13 together with Theorem 7.21. For the case $L = 1$ we refer to Theorem 7.27. □

Remark 7.16. *In case of $S^{\boldsymbol{r}}_{p,\theta}B(\mathbb{T}^d)$ with $p \ge 1$ and $\boldsymbol{r} > 1/p$ similar characterizations were proved by Dinh Dũng [25, 26, 27] using the following variant of the de la Vallée-Poussin kernel*

$$V_j(x) = \frac{\sin(2^{j-1}x)\sin(32^{j-1}x)}{2^{2j}3\sin^2(\frac{x}{2})}\,, \tag{7.4.1}$$

which yields to an interpolation operator on $3 \cdot 2^j$ equidistant nodes. We can reproduce and extend this result to the Triebel-Lizorkin scale as well as to $p > 1/2$ with straightforward modifications of the arguments used in Theorems 7.21 below. Note, that our proof only uses a reproduction and a decay property of the kernel. Also the de la Vallée Poussin sampling operator R_m used by Temlyakov in [117, I.6] is admissible here.

7.5 The case of quadratically decaying kernels

Let us first deal with kernels providing at least a quadratic decay according to Lemma 7.6. We introduce the characteristic function $\chi^*_{j,u}$ of the dyadic interval $[2\pi u/2^j, 2\pi(u+1)/2^j]$ indexed by $j \in \mathbb{N}_0$ and $u \in \mathbb{Z}$. For $\boldsymbol{j} \in \mathbb{N}_0^d$ and $\boldsymbol{u} \in \mathbb{Z}^d$ we denote with

$$\chi^*_{\boldsymbol{j},\boldsymbol{u}}(\boldsymbol{x}) := \prod_{i=1}^d \chi^*_{j_i,u_i}(x_i)$$

the characteristic function of the respective parallelepiped. We remember the definition of $A_{\boldsymbol{j}}(\boldsymbol{x})$ in (7.2.3) and state the following lemma.

Lemma 7.17. *Let $0 < \lambda \le 1$ and $L > \frac{1}{\lambda}$. For any sequence $(\lambda_{\boldsymbol{u}})_{\boldsymbol{u}\in A_{\boldsymbol{j}}(\boldsymbol{x})}$ of complex numbers and every $\boldsymbol{j} \in \mathbb{N}_0^d$ we have*

$$\sum_{\boldsymbol{u}\in\mathbb{Z}^d} |\lambda_{\boldsymbol{j},\boldsymbol{u}}| \prod_{i=1}^d (1 + 2^j|x_i - x^{j_i}_{u_i}|)^{-L} \le C\Big[M\Big|\sum_{\boldsymbol{u}\in\mathbb{Z}^d} \lambda_{\boldsymbol{j},\boldsymbol{u}}\chi^*_{\boldsymbol{j},\boldsymbol{u}}\Big|^\lambda(\boldsymbol{x})\Big]^{\frac{1}{\lambda}} \tag{7.5.1}$$

with a constant C independent of $\boldsymbol{j}$, $(\lambda_{\boldsymbol{j},\boldsymbol{u}})_{\boldsymbol{u}}$ and $\boldsymbol{x}$.

Proof. This Lemma is a special case of Lemma B.13. We refer to the prove there. □

Proposition 7.18. *Let $\boldsymbol{\ell}, \boldsymbol{j} \in \mathbb{N}_0^d$, $0 < \lambda \le 1$, $L \in \mathbb{N}$ with $L > \frac{1}{\lambda}$ and $a > 0$. Let further $f \in C(\mathbb{T}^d)$.*

(i) Then

$$|I^L_{\boldsymbol{j}}[f](\boldsymbol{x})| \lesssim 2^{a|\boldsymbol{\ell}|_1}[M|P_{\boldsymbol{2}^{\boldsymbol{j}+\boldsymbol{\ell}},a}f|^\lambda(\boldsymbol{x})]^{\frac{1}{\lambda}}$$

(ii) and furthermore

$$|q_{\boldsymbol{j}}^L[f](\boldsymbol{x})| \lesssim 2^{a|\boldsymbol{\ell}|_1}[M|P_{\boldsymbol{2}^{\boldsymbol{j}+\boldsymbol{\ell}},a}f|^\lambda(\boldsymbol{x})]^{\frac{1}{\lambda}}$$

holds with a constant independent of $\boldsymbol{\ell}, \boldsymbol{j}, \boldsymbol{x}$ *and* f.

Proof. We start proving *(i)*. Recall the notation from (7.2.3). Periodicity of f and $K_{\pi,j}^L$ yields

$$\begin{aligned} I_{\boldsymbol{j}}^L f(\boldsymbol{x})| &= \sum_{\boldsymbol{u}\in A_{\boldsymbol{j}}} f(\boldsymbol{x}_{\boldsymbol{u}}^{\boldsymbol{j}}) K_{\pi^d,\boldsymbol{j}}^L(\boldsymbol{x}-\boldsymbol{x}_{\boldsymbol{u}}^{\boldsymbol{j}}) \\ &= \sum_{\boldsymbol{u}\in A_{\boldsymbol{j}}(\boldsymbol{x})} f(\boldsymbol{x}_{\boldsymbol{u}}^{\boldsymbol{j}}) K_{\pi^d,\boldsymbol{j}}^L(\boldsymbol{x}-\boldsymbol{x}_{\boldsymbol{u}}^{\boldsymbol{j}}). \end{aligned}$$

Lemma 7.6 with $|x_i - x_{u_i}^{j_i}| \le \pi$ gives

$$\begin{aligned} |I_{\boldsymbol{j}}^L f(\boldsymbol{x})| &\le \sum_{\boldsymbol{u}\in A_{\boldsymbol{j}}(\boldsymbol{x})} \left| f(\boldsymbol{x}_{\boldsymbol{u}}^{\boldsymbol{j}}) \prod_{i=1}^d K_{\pi,j_i}^L(x_i - x_{u_i}^{j_i}) \right| \\ &\lesssim \sum_{\boldsymbol{u}\in\mathbb{Z}^d} |\lambda_{\boldsymbol{j},\boldsymbol{u}}| \prod_{i=1}^d (1+2^{j_i}|x_i - x_{u_i}^{j_i}|)^{-L}, \end{aligned}$$

where we used the notation

$$\lambda_{\boldsymbol{j},\boldsymbol{u}} := \sup_{\substack{\boldsymbol{y}:|y_i - x_{u_i}^{j_i}| < \frac{2\pi}{2^{j_i}} \\ i\in[d]}} |f(\boldsymbol{y})|.$$

Applying Lemma 7.17 gives

$$|I_{\boldsymbol{j}}^L f(\boldsymbol{x})| \lesssim \Big[M\Big|\sum_{\boldsymbol{u}\in\mathbb{Z}^d} \lambda_{\boldsymbol{j},\boldsymbol{u}}\chi_{\boldsymbol{j},\boldsymbol{k}}^*\Big|^\lambda(\boldsymbol{x})\Big]^{\frac{1}{\lambda}}. \tag{7.5.2}$$

Taking $\boldsymbol{z} \in \operatorname{supp}\chi_{\boldsymbol{j},\boldsymbol{u}^*}^*$ gives for any $a > 0$

$$\begin{aligned} \Big|\sum_{\boldsymbol{u}\in\mathbb{Z}^d} \lambda_{\boldsymbol{j},\boldsymbol{u}}\chi_{\boldsymbol{j},\boldsymbol{u}}^*(\boldsymbol{z})\Big| = |\lambda_{\boldsymbol{j},\boldsymbol{u}^*}| &= \sup_{\substack{\boldsymbol{y}:|y_i - x_{u_i^*}^{j_i}| < \frac{2\pi}{2^{j_i}} \\ i\in[d]}} |f(\boldsymbol{y})| \\ &\lesssim \sup_{\substack{\boldsymbol{y}:|y_i - z_i| < \frac{4\pi}{2^{j_i}} \\ i\in[d]}} \frac{|f(\boldsymbol{y})|}{\prod_{i=1}^d (1+2^{j_i}|y_i - z_i|)^a}. \end{aligned}$$

Finally, Lemma B.12 yields

$$\Big|\sum_{\boldsymbol{u}\in\mathbb{Z}^d} \lambda_{\boldsymbol{j},\boldsymbol{u}}\chi_{\boldsymbol{j},\boldsymbol{u}}^*(\boldsymbol{z})\Big| \lesssim 2^{|\boldsymbol{\ell}|_1 a} P_{\boldsymbol{2}^{\boldsymbol{j}+\boldsymbol{\ell}},a} f(\boldsymbol{z}). \tag{7.5.3}$$

Inserting (7.5.3) into (7.5.2) finishes the proof of *(i)*. The bound in *(ii)* is a trivial consequence of applying triangle inequality to (7.2.2) and *(i)*

$$|q_{\boldsymbol{j}}^L[f](\boldsymbol{x})| \;\leq\; \sum_{\boldsymbol{b}\in\{-1,0\}^d} |I_{\boldsymbol{j}+\boldsymbol{b}}^L[f](\boldsymbol{x})|. \tag{7.5.4}$$

□

The following two theorems are the trigonometric counterparts of Theorem 4.16 and Theorem 4.18. There is no smoothness limitation from above as it is in the case of the Faber-Schauder expansion. Nevertheless we should mention that the sampling kernels need a certain decreasing property that depends on the integration and fine index parameters of the underlying function spaces.

Theorem 7.19. *Let $0 < p, \theta \leq \infty$ ($p < \infty$), $L > \max\{\frac{1}{p}, \frac{1}{\theta}, 1\}$ and $\boldsymbol{r} > \max\{\frac{1}{p}, \frac{1}{\theta}\}$.*

(i) Then every $f \in S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d)$ admits the representation

$$f = \sum_{\boldsymbol{j}\in\mathbb{N}_0^d} q_{\boldsymbol{j}}^L[f], \tag{7.5.5}$$

with unconditional convergence in $S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d)$ in case $0 < \theta < \infty$ and with unconditional convergence in $S^{\tilde{\boldsymbol{r}}}_{p,\nu}F(\mathbb{T}^d)$ for every $\boldsymbol{r} > \tilde{\boldsymbol{r}}$ and $0 < \nu \leq \infty$ in case $\theta = \infty$.

(ii) There is a constant $C > 0$ independent of f such that

$$\|2^{\boldsymbol{r}\cdot\boldsymbol{j}} q_{\boldsymbol{j}}^L(f)|L_p(\ell_\theta)\| \leq C\|f|S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d)\| \tag{7.5.6}$$

holds for all $f \in S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d)$.

Proof. Step 1. We prove (7.5.6). To begin with we choose $a > 0$ such that $\boldsymbol{r} > a > \max\{\frac{1}{p}, \frac{1}{q}\}$ is fulfilled. Let $f \in S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d)$. We start for $\boldsymbol{j} \in \mathbb{N}_0^d$ with the Fourier decomposition

$$f(\boldsymbol{x}) = \sum_{\boldsymbol{\ell}\in\mathbb{Z}^d} \delta^\pi_{\boldsymbol{j}+\boldsymbol{\ell}}[f](\boldsymbol{x}), \tag{7.5.7}$$

cf. (3.1.1), where $\delta_{\boldsymbol{j}}[f] := 0$ for $\boldsymbol{j} \in \mathbb{Z}^d \backslash \mathbb{N}_0^d$. This series converges unconditionally in $C(\mathbb{T}^d)$, due to the embedding $S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d) \hookrightarrow \mathbb{C}(\mathbb{T}^d)$. That yields the point-wise estimate

$$|q_{\boldsymbol{j}}^L[f](\boldsymbol{x})| \leq \sum_{\boldsymbol{\ell}\in\mathbb{Z}^d} |q_{\boldsymbol{j}}^L[\delta^\pi_{\boldsymbol{j}+\boldsymbol{\ell}}[f]](\boldsymbol{x})|.$$

For the sake of simplicity we assume that the constants A, B, C in Definition 3.1 are chosen in such a way that $\delta_{\boldsymbol{j}}[f] \in \mathcal{T}^L_{\boldsymbol{j}}$ is fulfilled for all $\boldsymbol{j} \in \mathbb{N}_0^d$. Then Proposition 7.10 implies

$$|q_{\boldsymbol{j}}^L[f](\boldsymbol{x})| \leq \sum_{\boldsymbol{\ell}\geq 0} |q_{\boldsymbol{j}}^L[\delta^\pi_{\boldsymbol{j}+\boldsymbol{\ell}}[f]](\boldsymbol{x})|.$$

Applying Proposition 7.18, *(ii)* we obtain

$$|q_{\boldsymbol{j}}^L[f](\boldsymbol{x})| \;\lesssim\; \sum_{\boldsymbol{\ell}\geq 0} 2^{a|\boldsymbol{\ell}|_1}\Big[M|P_{\boldsymbol{2}^{\boldsymbol{j}+\boldsymbol{\ell}},a}\delta^{\pi}_{\boldsymbol{j}+\boldsymbol{\ell}}[f]|^{\lambda}(\boldsymbol{x})\Big]^{\frac{1}{\lambda}}.$$

Multiplying with the weight $2^{\boldsymbol{r}\cdot\boldsymbol{j}}$ we find the point-wise estimate

$$2^{\boldsymbol{r}\cdot\boldsymbol{j}}|q_{\boldsymbol{j}}^L[f](\boldsymbol{x})| \lesssim \sum_{\boldsymbol{\ell}\geq 0} 2^{(a\boldsymbol{1}-\boldsymbol{r})\cdot\boldsymbol{\ell}}2^{\boldsymbol{r}\cdot(\boldsymbol{j}+\boldsymbol{\ell})}\Big[M|P_{\boldsymbol{2}^{\boldsymbol{j}+\boldsymbol{\ell}},a}\delta^{\pi}_{\boldsymbol{j}+\boldsymbol{\ell}}[f]|^{\lambda}(\boldsymbol{x})\Big]^{\frac{1}{\lambda}}. \tag{7.5.8}$$

where λ is chosen as $L > \frac{1}{\lambda} > \frac{1}{\min\{p,\theta\}}$ ($\lambda = 1$ in case $\min\{p,\theta\} > 1$). The parameter a will be fixed later. Now we take the $L_p(\ell_\theta)$ (quasi)-norm on both sides. Due to u-triangle inequality in $L_p(\ell_\theta)$ with $u = \min\{p,\theta,1\}$ we obtain

$$\|2^{\boldsymbol{r}\cdot\boldsymbol{j}}q_{\boldsymbol{j}}^L[f]|L_p(\ell_\theta)\| \lesssim \Big(\sum_{\boldsymbol{\ell}\geq 0} 2^{(a\boldsymbol{1}-\boldsymbol{r})\cdot\boldsymbol{\ell}u}\Big\|2^{\boldsymbol{r}\cdot(\boldsymbol{j}+\boldsymbol{\ell})}\Big[M|P_{\boldsymbol{2}^{\boldsymbol{j}+\boldsymbol{\ell}},a}\delta^{\pi}_{\boldsymbol{j}+\boldsymbol{\ell}}[f]|^{\lambda}\Big]^{\frac{1}{\lambda}}\Big|L_p(\ell_\theta)\Big\|^{u}\Big)^{\frac{1}{u}}. \tag{7.5.9}$$

Since $\lambda < \min\{p,\theta\}$ in case $\min\{p,\theta\} \leq 1$ a trick similar to 7.3.11 yields

$$\Big\|2^{\boldsymbol{r}\cdot(\boldsymbol{j}+\boldsymbol{\ell})}\Big[M|P_{\boldsymbol{2}^{\boldsymbol{j}+\boldsymbol{\ell}},a}\delta^{\pi}_{\boldsymbol{j}+\boldsymbol{\ell}}[f]|^{\lambda}\Big]^{\frac{1}{\lambda}}\Big|L_p(\ell_\theta)\Big\| \;=\; \Big\|2^{\lambda\boldsymbol{r}\cdot(\boldsymbol{j}+\boldsymbol{\ell})}M|P_{\boldsymbol{2}^{\boldsymbol{j}+\boldsymbol{\ell}},a}\delta^{\pi}_{\boldsymbol{j}+\boldsymbol{\ell}}[f]|^{\lambda}\Big|L_{\frac{p}{\lambda}}(\ell_{\frac{\theta}{\lambda}})\Big\|^{\frac{1}{\lambda}}.$$

This allows us to apply Fefferman-Stein maximal inequality (Theorem B.14)

$$\Big\|2^{\boldsymbol{r}\cdot(\boldsymbol{j}+\boldsymbol{\ell})}\Big[M|P_{\boldsymbol{2}^{\boldsymbol{j}+\boldsymbol{\ell}},a}\delta^{\pi}_{\boldsymbol{j}+\boldsymbol{\ell}}[f]|^{\lambda}\Big]^{\frac{1}{\lambda}}\Big|L_p(\ell_\theta)\Big\| \;\lesssim\; \|2^{\boldsymbol{r}\cdot(\boldsymbol{j}+\boldsymbol{\ell})}P_{\boldsymbol{2}^{\boldsymbol{j}+\boldsymbol{\ell}},a}\delta^{\pi}_{\boldsymbol{j}+\boldsymbol{\ell}}[f]|L_p(\ell_\theta)\|.$$

Next we choose a such that $\boldsymbol{r} > a > \max\{\frac{1}{p},\frac{1}{\theta}\}$ holds. Then applying Peetre maximal inequality (Theorem B.17) gives

$$\|2^{\boldsymbol{r}\cdot(\boldsymbol{j}+\boldsymbol{\ell})}P_{\boldsymbol{2}^{\boldsymbol{\ell}+\boldsymbol{j}},a}\delta_{\boldsymbol{\ell}+\boldsymbol{j}}[f]|L_p(\ell_\theta)\| \lesssim \|2^{\boldsymbol{r}\cdot(\boldsymbol{j}+\boldsymbol{\ell})}\delta_{\boldsymbol{\ell}+\boldsymbol{j}}[f]|L_p(\ell_\theta)\|.$$

Obviously, we have

$$\|2^{\boldsymbol{r}\cdot(\boldsymbol{j}+\boldsymbol{\ell})}\delta_{\boldsymbol{\ell}+\boldsymbol{j}}[f]|L_p(\ell_\theta)\| \leq \|2^{\boldsymbol{r}\cdot\boldsymbol{j}}\delta_{\boldsymbol{j}}[f]|L_p(\ell_\theta)\|.$$

Inserting this into (7.5.9) yields

$$\begin{aligned}\|2^{\boldsymbol{r}\cdot\boldsymbol{j}}q_{\boldsymbol{j}}^L[f]|L_p(\ell_\theta)\| &\lesssim \|2^{\boldsymbol{r}\cdot\boldsymbol{j}}\delta_{\boldsymbol{j}}[f]|L_p(\ell_\theta)\|\Big(\sum_{\boldsymbol{\ell}\geq 0} 2^{(a\boldsymbol{1}-\boldsymbol{r})\cdot\boldsymbol{\ell}u}\Big)^{\frac{1}{u}}\\ &\lesssim \|2^{\boldsymbol{r}\cdot\boldsymbol{j}}\delta_{\boldsymbol{j}}[f]|L_p(\ell_\theta)\|,\end{aligned}$$

where the choice of a ensures the convergence of the series to an absolute constant. *Step 2.* We prove *(i)*. The equation (7.5.6) implies

$$\|2^{\boldsymbol{r}\cdot\boldsymbol{j}}q_{\boldsymbol{j}}^L[f]|L_p(\ell_\theta)\| < \infty.$$

Then Theorem 7.11 yields unconditional convergence of the series $\sum_{\boldsymbol{j}\in\mathbb{N}_0^d} q_{\boldsymbol{j}}^L[f]$. We show in case $0<\theta<\infty$

$$\Big\| f - \sum_{|\boldsymbol{j}|_1<M} q_{\boldsymbol{j}}^L[f] \Big| S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d)\Big\| \longrightarrow 0 \quad (M\to\infty).$$

As a consequence of Definition 3.2 trigonometric polynomials are dense in $S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d)$ if $\theta<\infty$. For that reason we find for every $\varepsilon>0$ a trigonometric polynomial t such that

$$\|f-t|S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d)\|<\varepsilon.$$

The u-triangle inequality gives

$$\Big\| f - \sum_{|\boldsymbol{j}|_1<M} q_{\boldsymbol{j}}^L[f] \Big| S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d)\Big\|^u \;\leq\; \|f-t|S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d)\|^u + \Big\| t - \sum_{|\boldsymbol{j}|_1<M} q_{\boldsymbol{j}}^L[f] \Big| S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d)\Big\|^u.$$

For n sufficiently large we obtain by Lemma 7.7

$$t - \sum_{|\boldsymbol{j}|_1<M} q_{\boldsymbol{j}}^L[f] = \sum_{|\boldsymbol{j}|_1<M} q_{\boldsymbol{j}}^L(t-f).$$

Applying Theorem 7.11 we have

$$\Big\| \sum_{|\boldsymbol{j}|_1<M} q_{\boldsymbol{j}}^L(t-f) \Big| S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d)\Big\| \lesssim \Big\| \Big(\sum_{|\boldsymbol{j}|_1<M} 2^{\theta\boldsymbol{r}\cdot\boldsymbol{j}} |q_{\boldsymbol{j}}^L(t-f)|^\theta \Big)^{\frac{1}{\theta}} \Big| L_p(\mathbb{T}^d)\Big\|.$$

Finally, Step 1 yields

$$\Big\| \Big(\sum_{|\boldsymbol{j}|_1<M} 2^{\theta\boldsymbol{r}\cdot\boldsymbol{j}} |q_{\boldsymbol{j}}^L(t-f)|^\theta \Big)^{\frac{1}{\theta}} \Big\|_p \lesssim \|t-f|S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d)\|$$

and hence, there is a constant $C>0$ independent of M, f and t such that

$$\Big\| f - \sum_{|\boldsymbol{j}|_1<M} q_{\boldsymbol{j}}^L[f] \Big| S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d)\Big\| \leq C2\varepsilon.$$

The case $\theta=\infty$ is based on the embedding

$$S^{\boldsymbol{r}}_{p,\infty}F(\mathbb{T}^d) \hookrightarrow S^{\boldsymbol{s}}_{p,p}F(\mathbb{T}^d) \hookrightarrow S^{\tilde{\boldsymbol{r}}}_{p,\nu}F(\mathbb{T}^d)$$

with $\boldsymbol{r}>\boldsymbol{s}>\frac{1}{p}$, $\boldsymbol{s}>\tilde{\boldsymbol{r}}$ and $0<\nu<\infty$ where the density argument from above is applied to $S^{\boldsymbol{s}}_{p,p}F(\mathbb{T}^d)$. □

Remark 7.20. *According to Remark 4.21 the recent result in [100, Rem. 7.3], see also [101], indicates that a corresponding characterization in case of small smoothness, i.e.* $\frac{1}{p}<r\leq\frac{1}{\theta}$ *may fail.*

Theorem 7.21. *Let* $0<p,\theta\leq\infty$, $L>\max\{\frac{1}{p},1\}$, $\boldsymbol{r}>\frac{1}{p}$.

(i) Then every $f \in S^{\boldsymbol{r}}_{p,\theta}B(\mathbb{T}^d)$ can be represented by

$$f = \sum_{\boldsymbol{j} \in \mathbb{N}_0^d} q^L_{\boldsymbol{j}}(f),$$

with unconditional convergence in $S^{\boldsymbol{r}}_{p,\theta}B(\mathbb{T}^d)$ in case $\max\{p,\theta\} < \infty$, and with unconditional convergence in $S^{\tilde{\boldsymbol{r}}}_{p,\nu}B(\mathbb{T}^d)$ for every $\boldsymbol{r} > \tilde{\boldsymbol{r}}$ and $0 < \nu \leq \infty$ in case $\max\{p,\theta\} = \infty$.

(ii) There is a constant $C > 0$ independent of f such that

$$\|2^{\boldsymbol{r}\cdot\boldsymbol{j}} q^L_{\boldsymbol{j}}[f]|\ell_\theta(L_p(\mathbb{T}^d))\| \leq C\|f|S^{\boldsymbol{r}}_{p,\theta}B(\mathbb{T}^d)\|$$

holds for all $f \in S^{\boldsymbol{r}}_{p,\theta}B(\mathbb{T}^d)$.

Proof. Concerning representation and unconditional convergence we follow the proof of Theorem 7.19 line by line with the obvious modifications for the B-case. The inequality in *(ii)* can be proven by the following arguments. We take the $\ell_\theta(L_p(\mathbb{T}^d))$ (quasi)-norm on both sides of the estimate in (7.5.8). Due to u-triangle inequality in $\ell_\theta(L_p(\mathbb{T}^d))$ with $u = \min\{p,\theta,1\}$ we obtain

$$\|2^{\boldsymbol{r}\cdot\boldsymbol{j}} q^L_{\boldsymbol{j}}[f]|\ell_\theta(L_p(\mathbb{T}^d))\| \lesssim \Big(\sum_{\boldsymbol{\ell}\geq 0} 2^{(a\mathbf{1}-\boldsymbol{r})\cdot\boldsymbol{\ell}u} \Big\| 2^{\boldsymbol{r}\cdot(\boldsymbol{j}+\boldsymbol{\ell})}\Big[M|P_{\mathbf{2}^{\boldsymbol{j}+\boldsymbol{\ell}},a}\delta^\pi_{\boldsymbol{j}+\boldsymbol{\ell}}[f]|^\lambda\Big]^{\frac{1}{\lambda}}\Big|\ell_\theta(L_p(\mathbb{T}^d))\Big\|^u\Big)^{\frac{1}{u}} \tag{7.5.10}$$

with $\boldsymbol{r} > a > \frac{1}{p}$ and $0 < \lambda < p$ ($\lambda = 1$ if $p > 1$). In case $p \leq 1$ a trick similar to (7.3.11) yields

$$\begin{aligned}
&\Big\| 2^{\boldsymbol{r}\cdot(\boldsymbol{j}+\boldsymbol{\ell})}\Big[M|P_{\mathbf{2}^{\boldsymbol{j}+\boldsymbol{\ell}},a}\delta^\pi_{\boldsymbol{j}+\boldsymbol{\ell}}[f]|^\lambda\Big]^{\frac{1}{\lambda}}\Big|\ell_\theta(L_p(\mathbb{T}^d)\Big\| \\
&= \Big(\sum_{\boldsymbol{j}\in\mathbb{N}_0^d} 2^{\boldsymbol{r}\cdot(\boldsymbol{j}+\boldsymbol{\ell})\theta}\Big\|M|P_{\mathbf{2}^{\boldsymbol{j}+\boldsymbol{\ell}},a}\delta^\pi_{\boldsymbol{j}+\boldsymbol{\ell}}[f]|^\lambda\Big|L_{\frac{p}{\lambda}}(\mathbb{T}^d)\Big\|^{\frac{\theta}{\lambda}}\Big)^{\frac{1}{\theta}}.
\end{aligned}$$

This allows us to apply Hardy-Littlewood maximal inequality (Theorem B.6). We obtain

$$\begin{aligned}
&\Big\| 2^{\boldsymbol{r}\cdot(\boldsymbol{j}+\boldsymbol{\ell})}\Big[M|P_{\mathbf{2}^{\boldsymbol{j}+\boldsymbol{\ell}},a}\delta^\pi_{\boldsymbol{j}+\boldsymbol{\ell}}[f]|^\lambda\Big]^{\frac{1}{\lambda}}\Big|\ell_\theta(L_p(\mathbb{T}^d)\Big\| \\
\lesssim\ &\Big(\sum_{\boldsymbol{j}\in\mathbb{N}_0^d} 2^{\boldsymbol{r}\cdot(\boldsymbol{j}+\boldsymbol{\ell})\theta}\Big\|P_{\mathbf{2}^{\boldsymbol{j}+\boldsymbol{\ell}},a}\delta^\pi_{\boldsymbol{j}+\boldsymbol{\ell}}[f]|\Big|L_p(\mathbb{T}^d)\Big\|^{\theta}\Big)^{\frac{1}{\theta}}.
\end{aligned}$$

Inserting this into (7.5.10) and applying (non-vector valued) Peetre maximal inequality (Theorem B.16) gives

$$\begin{aligned}
\|2^{\boldsymbol{r}\cdot\boldsymbol{j}} q^L_{\boldsymbol{j}}[f]|\ell_\theta(L_p(\mathbb{T}^d))\| &\lesssim \Big(\sum_{\boldsymbol{\ell}\geq 0} 2^{(a\mathbf{1}-\boldsymbol{r})\cdot\boldsymbol{\ell}u}\|2^{\boldsymbol{r}\cdot(\boldsymbol{j}+\boldsymbol{\ell})}\delta_{\boldsymbol{\ell}+\boldsymbol{j}}[f]|\ell_\theta(L_p(\mathbb{T}^d))\|^u\Big)^{\frac{1}{u}} \\
&\leq \Big(\sum_{\boldsymbol{\ell}\geq 0} 2^{(a\mathbf{1}-\boldsymbol{r})\cdot\boldsymbol{\ell}u}\Big)^{\frac{1}{u}}\|2^{\boldsymbol{r}\cdot\boldsymbol{j}}\delta_{\boldsymbol{j}}[f]|\ell_\theta(L_p(\mathbb{T}^d))\|,
\end{aligned}$$

where the term inside the $\ell_\theta(L_p(\mathbb{T}^d))$ norm does not depend any longer on ℓ. Therefore the sum over $\boldsymbol{\ell}$ converges to a constant depending only on a, $\boldsymbol{r}$ and the dimension d. Finally, we obtain

$$\|2^{\boldsymbol{r}\cdot\boldsymbol{j}} q_{\boldsymbol{j}}^L[f]|\ell_\theta(L_p(\mathbb{T}^d))\| \lesssim \|f|S^{\boldsymbol{r}}_{p,\theta}B(\mathbb{T}^d)\|.$$

□

Remark 7.22. *We strongly conjecture the optimality of the condition on L in the above theorems, see also Remark 3.3,*(ii) *above.*

7.6 The case of the Dirichlet kernel

In this subsection we study sampling representations based on the Dirichlet kernel $K^1_{\pi,j}$. Its slow decay causes some difficulties. We define an auxiliary kernel

$$\tilde{K}^2(x) := \sqrt{2\pi}\mathcal{F}^{-1}\big(4\chi^*_{[-\frac{5}{8},\frac{5}{8}]} * \chi^*_{[-\frac{1}{8},\frac{1}{8}]}\big)(x) = 16\frac{\sin\left(\frac{5}{8}x\right)\sin\left(\frac{1}{8}x\right)}{x^2} \in L_1(\mathbb{R}^d)$$

and its periodization

$$\tilde{K}^2_{\pi,j}(x) := \sum_{k=-\infty}^{\infty} \tilde{K}^2(2^j(x+2\pi k)).$$

Similar to Lemma 7.6 we can show for $|x|<\pi$ the following decay property

$$|\tilde{K}^2_{\pi,j}(x)| \lesssim \frac{1}{(1+2^j|x|)^2}. \tag{7.6.1}$$

Note, that the corresponding operator $\tilde{I}^2_j$ defined via (1.5.2) is a sampling but not an interpolation operator. However, Lemma 7.2 still holds true. According to Subsection 7.2 we define the multivariate sampling operator $\tilde{I}^2_{\boldsymbol{j}}f$ based on the tensorized kernel $\tilde{K}^2_{\pi^d,\boldsymbol{j}}$.

The following formula is a counterpart of a similar formula used by Temlyakov in [117, Lem. I.6.2]. Taking (1.5.3) into account we denote

$$\mathcal{D}^1_{\boldsymbol{j}} = \mathcal{D}^1_{j_1}\otimes\cdots\otimes\mathcal{D}^1_{j_d} \quad , \quad \boldsymbol{j}\in\mathbb{N}^d_0\,.$$

Lemma 7.23. *Let $f\in C(\mathbb{T}^d)$. Then*

$$I^1_{\boldsymbol{j}}f = (2\pi)^{-d}\mathcal{D}^1_{\boldsymbol{j}} * \tilde{I}^2_{\boldsymbol{j}}f \tag{7.6.2}$$

for all $\boldsymbol{j}\in\mathbb{N}^d_0$.

Proof. We prove the identity by comparing the Fourier series for arbitrary continuous functions f. (7.1.5) implies

$$\widehat{I^1_{\boldsymbol{j}}f}(\boldsymbol{\ell}) = \Big(\prod_{i\in[d]}\chi^*_{[-2^{j_i-1},2^{j_i-1}-1]}(\ell_i)\Big)\sum_{\boldsymbol{u}\in A_{\boldsymbol{j}}} f(\boldsymbol{x}^{\boldsymbol{j}}_{\boldsymbol{u}})e^{-i\boldsymbol{x}^{\boldsymbol{j}}_{\boldsymbol{u}}\cdot\boldsymbol{\ell}}. \tag{7.6.3}$$

Additionally, the same computation as used in Lemma 7.2 shows

$$\widehat{\tilde{I}^2_{\boldsymbol{j}}[f]}(\boldsymbol{\ell}) = \frac{1}{(2\pi)^{d/2}}\Big(\prod_{i=1}^{d}\mathcal{F}\tilde{K}^2\Big(\frac{\ell_i}{2^{j_i}}\Big)\Big)\sum_{\boldsymbol{u}\in A_{\boldsymbol{j}}} f(\boldsymbol{x}^{\boldsymbol{j}}_{\boldsymbol{u}})e^{-i\boldsymbol{x}^{\boldsymbol{j}}_{\boldsymbol{u}}\cdot\boldsymbol{\ell}}.$$

Clearly,

$$(2\pi)^{-d}\widehat{\mathcal{D}^1_{\boldsymbol{j}} * \tilde{I}^2_{\boldsymbol{j}}f}(\boldsymbol{\ell}) = \widehat{\mathcal{D}^1_{\boldsymbol{j}}}(\boldsymbol{\ell})\widehat{\tilde{I}^2_{\boldsymbol{j}}f}(\boldsymbol{\ell}) = \widehat{\mathcal{D}^1_{\boldsymbol{j}}}(\boldsymbol{\ell})\sum_{\boldsymbol{u}\in A_{\boldsymbol{j}}} f(\boldsymbol{x}^{\boldsymbol{j}}_{\boldsymbol{u}})e^{-i\boldsymbol{x}^{\boldsymbol{j}}_{\boldsymbol{u}}\cdot\boldsymbol{\ell}} \tag{7.6.4}$$

since

$$\mathcal{F}\tilde{K}^2\Big(\frac{\ell_i}{2^{j_i}}\Big) = \sqrt{2\pi}$$

for $\ell_i \in [-2^{j_i}, 2^{j_i}), i \in [d]$. Comparing (7.6.3) and (7.6.4) yields the claim. □

Lemma 7.24. *Let $\boldsymbol{\ell}, \boldsymbol{j} \in \mathbb{N}_0^d$, $a > 0$ and $1/2 < \lambda \leq 1$. Furthermore, let $f \in C(\mathbb{T}^d)$. Then*

$$|\tilde{I}^2_{\boldsymbol{j}}[f](\boldsymbol{x})| \lesssim 2^{a|\boldsymbol{\ell}|_1}[M|P_{\boldsymbol{2}^{\boldsymbol{j}+\boldsymbol{\ell}},a}f|^\lambda(\boldsymbol{x})]^{\frac{1}{\lambda}}$$

holds with a constant independent of $\boldsymbol{\ell}, \boldsymbol{j}, \boldsymbol{x}$ and f.

Proof. We refer to the proof of Proposition 7.18. Recognizing, that the only property of $\tilde{I}^2_{\boldsymbol{j}}$ we need is the decay of the underlying kernel $\tilde{K}^2_{\pi^d,\boldsymbol{j}}$ provided in (7.6.1). □

Remark 7.25. (i) *The estimates in Lemmas 7.18, 7.24 are pointwise and very useful for $L_p(\mathbb{T}^d, \ell_\theta)$ estimates. In case one is interested in (scalar) L_p estimates, similar as in [117, Lem. I.6.2], then Lemmas 7.18 and 7.24 together with the maximal inequalities Theorems B.6, B.16 imply for $0 < p \leq \infty$, $L > \max\{1/p, 1\}$ and any $a > 1/p$*

$$\|I^L_{\boldsymbol{j}}f|L_p(\mathbb{T}^d)\| \lesssim_{L,a} 2^{|\boldsymbol{\ell}|_1 a}\|f|L_p(\mathbb{T}^d)\| \quad , \quad f \in \mathcal{T}^0_{\boldsymbol{j}+\boldsymbol{\ell}} \tag{7.6.5}$$

(similar for $\tilde{I}^2_{\boldsymbol{j}}$).

(ii) *There is a different technique based on periodic versions of Plancherel-Polya inequalities (Marcinkiewicz-Zygmund inequalities) for $0 < p \leq \infty$, see [98, Thms. 6,10]. A straight-forward modification of the argument in [98, Lem. 13,*(ii)*] gives for $0 < p \leq \infty$ and $L > \max\{1/p, 1\}$*

$$\|I^L_{\boldsymbol{j}}f|L_p(\mathbb{T}^d)\| \lesssim_p 2^{|\boldsymbol{\ell}|_1/p}\|f|L_p(\mathbb{T}^d)\| \quad , \quad f \in \mathcal{T}^0_{\boldsymbol{j}+\boldsymbol{\ell}} \tag{7.6.6}$$

(similar for $\tilde{I}^2_{\boldsymbol{j}}$). In case $L = 2$ (de la Vallée Poussin) this yields an extension of [117, Lem. I.6.2] to the range $1/2 < p \leq \infty$.

(iii) *By Lemma 7.23 and the uniform boundedness of the multivariate Fourier partial sum operator in $L_p(\mathbb{T}^d)$, $1 < p < \infty$, we obtain from* (7.6.5) *and* (7.6.6) *corresponding estimates also for $\|I^1_{\boldsymbol{j}}f\|_p$.*

Theorem 7.26. *Let $1 < p, \theta < \infty$ and $\boldsymbol{r} > \max\{\frac{1}{p}, \frac{1}{\theta}\}$.*

(i) Then every $f \in S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d)$ admits the representation

$$f = \sum_{\boldsymbol{j} \in \mathbb{N}_0^d} q^1_{\boldsymbol{j}}[f],$$

with unconditional convergence in $S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d)$.

(ii) There is a constant $C > 0$ independent of f such that

$$\|2^{\boldsymbol{r}\cdot\boldsymbol{j}} q^1_{\boldsymbol{j}}(f)|L_p(\ell_\theta)\| \le C\|f|S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d)\|$$

holds for all $f \in S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d)$.

Proof. The proof of *(i)* is similar to Theorem 7.19, *(i)*. We prove *(ii)* here. Inserting the decomposition (3.6.2), applying triangle inequality and afterwards Proposition 7.10 gives

$$\|2^{\boldsymbol{r}\cdot\boldsymbol{j}} q^1_{\boldsymbol{j}}[f]|L_p(\ell_\theta(\mathbb{N}_0^d))\| \lesssim \sum_{\boldsymbol{\ell}\ge 0} 2^{-\boldsymbol{\ell}\cdot\boldsymbol{r}} \|2^{\boldsymbol{r}\cdot(\boldsymbol{j}+\boldsymbol{\ell})} q^1_{\boldsymbol{j}}[\delta^\pi_{\boldsymbol{j}+\boldsymbol{\ell}}[f]]|L_p(\ell_\theta(\mathbb{N}_0^d))\|.$$

The relation in (7.5.4) shows

$$\|2^{\boldsymbol{r}\cdot\boldsymbol{j}} q^1_{\boldsymbol{j}}[f]|L_p(\ell_\theta(\mathbb{N}_0^d))\| \lesssim \sum_{\boldsymbol{b}\in\{-1,0\}^d} \sum_{\boldsymbol{\ell}\ge 0} 2^{-\boldsymbol{\ell}\cdot\boldsymbol{r}} \|2^{\boldsymbol{r}\cdot(\boldsymbol{j}+\boldsymbol{\ell})} I^1_{\boldsymbol{j}+\boldsymbol{b}}[\delta^\pi_{\boldsymbol{j}+\boldsymbol{\ell}}[f]]|L_p(\ell_\theta(\mathbb{N}_0^d))\|.$$

Hence, Lemma 7.23 yields

$$\|2^{\boldsymbol{r}\cdot\boldsymbol{j}} q^1_{\boldsymbol{j}}[f]|L_p(\ell_\theta(\mathbb{N}_0^d))\| \lesssim \sum_{\boldsymbol{b}\in\{-1,0\}^d} \sum_{\boldsymbol{\ell}\ge 0} 2^{-\boldsymbol{\ell}\cdot\boldsymbol{r}} \|2^{\boldsymbol{r}\cdot(\boldsymbol{j}+\boldsymbol{\ell})} \mathcal{D}^1_{\boldsymbol{j}+\boldsymbol{b}} * \tilde{I}^2_{\boldsymbol{j}+\boldsymbol{b}}[\delta^\pi_{\boldsymbol{j}+\boldsymbol{\ell}}[f]]|L_p(\ell_\theta(\mathbb{N}_0^d))\|. \tag{7.6.7}$$

Lizorkin presented in [74, p. 241, Thm. 5] a theorem on Fourier multipliers for the $L_p(\ell_\theta)$ situation. The result in [99, Thm. 3.4.2] transfers this to the periodic setting. Referring to a comment in [119, 2.5.4] the Fourier partial sum with respect to a parallelepiped fulfills the requirements of this theorem and we get rid of $\mathcal{D}^1_{\boldsymbol{j}+\boldsymbol{b}}$ in (7.6.7). This gives

$$\|2^{\boldsymbol{r}\cdot\boldsymbol{j}} q^1_{\boldsymbol{j}}[f]|L_p(\ell_\theta(\mathbb{N}_0^d))\| \lesssim \sum_{\boldsymbol{b}\in\{-1,0\}^d} \sum_{\boldsymbol{\ell}\ge 0} 2^{-\boldsymbol{\ell}\cdot\boldsymbol{r}} \|2^{\boldsymbol{r}\cdot(\boldsymbol{j}+\boldsymbol{\ell})} \tilde{I}^2_{\boldsymbol{j}+\boldsymbol{b}} \delta^\pi_{\boldsymbol{j}+\boldsymbol{\ell}}[f]|L_p(\ell_\theta(\mathbb{N}_0^d))\|.$$

Lemma 7.24 with $\lambda = 1$ yields

$$\|2^{\boldsymbol{r}\cdot\boldsymbol{j}} q^1_{\boldsymbol{j}}[f]|L_p(\ell_\theta(\mathbb{N}_0^d))\| \lesssim \sum_{\boldsymbol{\ell}\ge 0} 2^{\boldsymbol{\ell}\cdot(a\boldsymbol{1}-\boldsymbol{r})} \|2^{\boldsymbol{r}\cdot(\boldsymbol{j}+\boldsymbol{\ell})} M|P_{\boldsymbol{2}^{\boldsymbol{j}+\boldsymbol{\ell}},a} f_{\boldsymbol{j}+\boldsymbol{\ell}}|(\boldsymbol{x})|L_p(\ell_\theta(\mathbb{N}_0^d))\|.$$

We finish the proof by following the estimates in the proof of Theorem 7.19 beginning from (7.5.9). □

Theorem 7.27. *Let $1 < p < \infty, 0 < \theta \le \infty$ and $\boldsymbol{r} > \frac{1}{p}$.*

(i) Then every $f \in S^{\boldsymbol{r}}_{p,\theta}B(\mathbb{T}^d)$ *can be represented by*

$$f = \sum_{\boldsymbol{j} \in \mathbb{N}_0^d} q^1_{\boldsymbol{j}}(f),$$

with unconditional convergence in $S^{\boldsymbol{r}}_{p,\theta}B(\mathbb{T}^d)$ *in case* $\theta < \infty$, *and with unconditional convergence in* $S^{\tilde{\boldsymbol{r}}}_{p,\nu}B(\mathbb{T}^d)$ *for every* $\boldsymbol{r} > \tilde{\boldsymbol{r}}$ *and* $0 < \nu \leq \infty$ *in case* $\theta = \infty$.

(ii) There is a constant $C > 0$ *independent of* f *such that*

$$\|2^{\boldsymbol{r}\cdot\boldsymbol{j}} q^1_{\boldsymbol{j}}[f]|\ell_\theta(L_p(\mathbb{T}^d))\| \leq C\|f|S^{\boldsymbol{r}}_{p,\theta}B(\mathbb{T}^d)\|$$

holds for all $f \in S^{\boldsymbol{r}}_{p,\theta}B(\mathbb{T}^d)$.

Proof. To prove *(i)* we follow the proof of Theorem 7.21, *(i)*. The assertion *(ii)* can be obtained following the proof of Theorem 7.26 where we replace $\|\cdot|L_p(\ell_\theta(\mathbb{N}_0))\|$ by $\|\cdot|\ell_\theta(L_p(\mathbb{T}^d))\|$. Now we use the estimates in Remark (7.6.5), (7.6.6) from Remark 7.25 . □

Remark 7.28. *Similar (but not nested) Dirichlet kernels were studied in [9] connected with sampling representations in case* $p = \theta = 2$.

Chapter 8

Optimal sampling recovery

In this section we generalize the sparse grids results known from Section 5. We deal with the case of vector smoothness in the scale of Triebel-Lizorkin. We compare optimality in the sense of linear sampling recovery with optimality in the sense of the worst case error with respect to standard information known from Information Based Complexity, see [84, 85, 86] and the references therein.

8.1 Multivariate interpolation on periodic Smolyak grids

This time we study function spaces with vector smoothness $\boldsymbol{r} \in \mathbb{R}^d$ fulfilling

$$r = r_1 = \ldots = r_\mu < r_{\mu+1} \leq \ldots \leq r_d < \infty \quad , \quad \mu \leq d. \tag{8.1.1}$$

For that reason we analyze a direction-wise modified version of Smolyak's algorithm, cf. (1.2.4), given by

$$T_M^{L,\boldsymbol{\eta}} f := \sum_{\frac{1}{\eta_1}\boldsymbol{\eta}\cdot\boldsymbol{j} \leq m} q_{\boldsymbol{j}}^L[f]. \tag{8.1.2}$$

The parameter $\boldsymbol{\eta} > 0$ allows to control the level of refinement in single directions. A comparatively large value of $\boldsymbol{\eta}$ in the s-th component ends up in a small refinement in the s-th direction. The interpolation operator $T_M^{L,\boldsymbol{\eta}} f$ maps a continuous function to a trigonometric polynomial with frequencies in an anisotropic hyperbolic cross

$$AH_M^{d,\boldsymbol{\eta}} := \bigcup_{\{\boldsymbol{j}:\, \frac{1}{\eta_1}\boldsymbol{\eta}\cdot\boldsymbol{j} \leq M\}} \mathcal{P}_j^0.$$

According to Lemma 7.8 the operator $T_M^{L,\boldsymbol{\eta}}$ interpolates functions on an anisotropic sparse grid

$$AG_M^{d,\boldsymbol{\eta}} := \bigcup_{\frac{1}{\eta_1}\boldsymbol{\eta}\cdot\boldsymbol{j} \leq M} \Big\{ \boldsymbol{x}_{\boldsymbol{u}}^{\boldsymbol{j}} :\ \boldsymbol{u} \in \mathbb{Z}^d,\ -2^{j_i-1} \leq u_i \leq 2^{j_i-1} - 1,\ i \in [d] \Big\}. \tag{8.1.3}$$

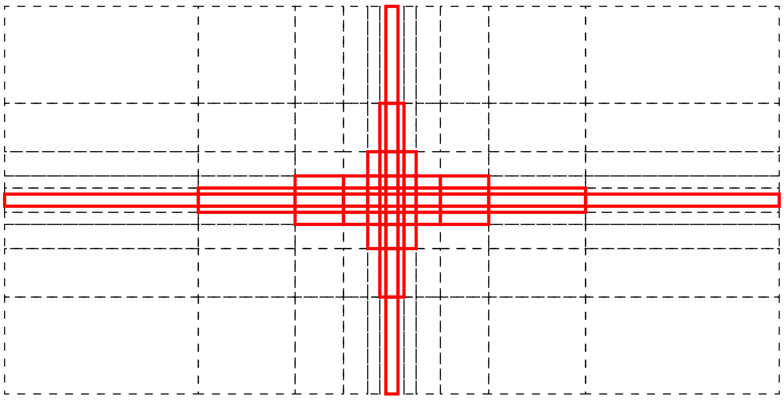


Figure 8.1: Anisotropic hyperbolic cross $AH_M^{2,(1,1.5)}$

Lemma 8.1. *Let* $\boldsymbol{\eta} \in \mathbb{R}^d$ *with*

$$0 < \eta_1 = \ldots = \eta_\mu < \eta_{\mu+1} \le \ldots \le \eta_d < \infty.$$

Then

$$|AG_M^{d,\boldsymbol{\eta}}| \asymp M^{\mu-1}2^M \tag{8.1.4}$$

holds for all $M \ge 1$.

Proof. Due to (8.1.3) an upper bound for the cardinality of $AG_M^{d,\boldsymbol{\eta}}$ is provided by

$$|AG_M^{d,\boldsymbol{\eta}}| \le \sum_{\frac{1}{\eta_1}\boldsymbol{\eta}\cdot\boldsymbol{j} \le M} 2^{|\boldsymbol{j}|_1}.$$

Hence, Lemma C.21 in the appendix provides the upper bound in (8.1.4). A trivial lower bound of 2^M is provided by simply counting the sampling nodes of $q_{\boldsymbol{j}}[f]$ of the level $j = (M, 0, \ldots, 0)$. A sharp bound can be obtained by using reproduction properties of $T_M^{L,\boldsymbol{\eta}}$ for trigonometric polynomials (cf. Lemma 7.7) with frequencies in $AH_M^{d,\boldsymbol{\eta}}$. The dimension of $AH_M^{d,\boldsymbol{\eta}}$ is given by $\sum_{\frac{1}{\eta_1}\boldsymbol{\eta}\cdot\boldsymbol{j}\le M} 2^{|\boldsymbol{j}|_1}$. □

Remark 8.2. *Comparing this estimate to uniformly refined sparse grids (*$\boldsymbol{\eta} = \mathbf{1}$*, cf. Theorems 5.6, 5.4, 5.5) we recognize that the underlying dimension of the space plays no role for the asymptotic bound. The dimension dependence is replaced by the* μ *largest refinement directions. Such effects are known at least since the 1970s in the former Soviet Union. In modern context they were rediscovered and applied in [27, 50, 49] and [32]).*

Theorem 8.3. *Let* $0 < p < q < \infty$ *and* $0 < \theta \le \infty$*. Additionally let* $L > \frac{1}{q}$ *and the smoothness vector* $\boldsymbol{r} > \frac{1}{p}$ *with* (8.1.1)*. Then*

$$\|f - T_M^{L,\boldsymbol{\eta}} f|L_q(\mathbb{T}^d)\| \lesssim 2^{-M(r_1-\frac{1}{p}+\frac{1}{q})}\|f|S_{p,\theta}^{\boldsymbol{r}}F(\mathbb{T}^d)\|$$

holds for all $M > 0$*. The operator generating vector* $\boldsymbol{\eta} \in \mathbb{R}^d$ *is chosen as* $\boldsymbol{\eta} = \boldsymbol{r} - \frac{1}{p} + \frac{1}{q}$.

Proof. We start expanding f into the series (7.5.5). This allows us to estimate

$$\begin{aligned}\|f - T_M^{L,\boldsymbol{\eta}} f|L_q(\mathbb{T}^d)\| &\le \Big\| \sum_{\frac{1}{\eta_1}\boldsymbol{\eta}\cdot\boldsymbol{j}>M} |q_{\boldsymbol{j}}[f]| \Big| L_q(\mathbb{T}^d)\Big\| \\ &\le 2^{-(r_1-\frac{1}{p}+\frac{1}{q})M} \Big\| \sum_{\boldsymbol{j}\in\mathbb{N}_0^d} 2^{(\boldsymbol{r}-\frac{1}{p}+\frac{1}{q})\cdot\boldsymbol{j}} |q_{\boldsymbol{j}}[f]| \Big| L_q(\mathbb{T}^d)\Big\|.\end{aligned}$$

We choose some parameters. Since $L > \frac{1}{q}$ we find $\tilde{q} \in \mathbb{R}$ with $p < \tilde{q} < q$ such that $L > \frac{1}{\tilde{q}}$ is fulfilled. Let $\tilde{r} := \boldsymbol{r} - \frac{1}{p} + \frac{1}{\tilde{q}}$. Applying Lemma 7.12 yields

$$\|f - T_M^{L,\boldsymbol{\eta}} f|L_q(\mathbb{T}^d)\| \lesssim 2^{-(r_1-\frac{1}{p}+\frac{1}{q})M} \Big\| \sup_{\boldsymbol{j}\in\mathbb{N}_0^d} 2^{\tilde{r}\cdot\boldsymbol{j}} |q_{\boldsymbol{j}}[f]| \Big\|_{\tilde{q}}$$

Theorem 7.19 yields

$$\|f - T_M^{L,\boldsymbol{\eta}} f|L_q(\mathbb{T}^d)\| \lesssim 2^{-(r_1-\frac{1}{p}+\frac{1}{q})M} \|f|S^{\tilde{r}}_{\tilde{q},\infty}F(\mathbb{T}^d)\|.$$

Finally, using the diagonal embedding stated in Lemma 3.4, (vi) gives

$$\|f - T_M^{L,\boldsymbol{\eta}} f|L_q(\mathbb{T}^d)\| \lesssim 2^{-(r_1-\frac{1}{p}+\frac{1}{q})M} \|f|S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d)\|,$$

which finishes the proof. □

For $\theta = 2$ we can reproduce a generalized form of a result due to Temlyakov [113].

Theorem 8.4. *Let* $0 < p < \infty$, $0 < \theta \le \infty$. *Additionally, let* $L \ge 1$ *and the smoothness vector* $\boldsymbol{r} > \frac{1}{p}$ *with* (8.1.1). *Then*

$$\|f - T_M^{L,\boldsymbol{\eta}} f|L_\infty(\mathbb{T}^d)\| \lesssim M^{(\mu-1)(1-\frac{1}{p})_+} 2^{-M(r_1-\frac{1}{p})} \|f|S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d)\|$$

holds for all $M > 0$. *The operator generating vector* $\boldsymbol{\eta} \in \mathbb{R}^d$ *is chosen as* $\boldsymbol{\eta} = \boldsymbol{\nu} - \frac{1}{p}$, *where* $\boldsymbol{\nu} \in \mathbb{R}^d$ *with*

$$r_s = \nu_s, \quad s = 1,\dots,\mu \quad \text{and} \quad r_1 < \nu_s < r_s, \quad s = \mu+1,\dots,d.$$

Proof. Step 1. We prove

$$\|f - T_M^{L,\boldsymbol{\eta}} f|L_\infty(\mathbb{T}^d)\| \lesssim \|f|S^{\boldsymbol{r}-\frac{1}{p}+\frac{1}{\tilde{p}}}_{\tilde{p},p}B(\mathbb{T}^d)\| \begin{cases} 2^{-M(r_1-\frac{1}{p})} & : \quad 0 < p \le 1, \\ M^{(\mu-1)(1-\frac{1}{p})} 2^{-M(r_1-\frac{1}{p})} & : \quad p > 1, \end{cases} \tag{8.1.5}$$

where $\tilde{p}$ is chosen such that $\max\{p,1\} < \tilde{p} < \infty$ is fulfilled. Expanding into (7.5.5) and using triangle inequality yields

$$\begin{aligned}\|f - T_M^{L,\boldsymbol{\eta}} f|L_\infty(\mathbb{T}^d)\| &= \Big\| \sum_{\frac{1}{\eta_1}\boldsymbol{\eta}\cdot\boldsymbol{j}>M} q^L_{\boldsymbol{j}}[f] \Big| L_\infty(\mathbb{T}^d)\Big\| \\ &\le \sum_{\frac{1}{\eta_1}\boldsymbol{\eta}\cdot\boldsymbol{j}>M} \|q^L_{\boldsymbol{j}}[f]|L_\infty(\mathbb{T}^d)\|.\end{aligned}$$

We have to distinguish the cases $0 < p \leq 1$ and the case $p > 1$. We start with $0 < p \leq 1$. The elementary embedding $\ell_p(\mathbb{N}_0^d) \hookrightarrow \ell_1(\mathbb{N}_0^d)$ yields

$$\begin{aligned}\|f - T_M^{L,\boldsymbol{\eta}} f|L_\infty(\mathbb{T}^d)\| &\leq 2^{-M(r_1-\frac{1}{p})} \sum_{\frac{1}{\eta_1}\boldsymbol{\eta}\cdot\boldsymbol{j}>M} 2^{(\boldsymbol{r}-\frac{1}{p})\cdot\boldsymbol{j}} \|q_{\boldsymbol{j}}^L[f]|L_\infty(\mathbb{T}^d)\| \\ &\leq 2^{-M(r_1-\frac{1}{p})}\Big(\sum_{\frac{1}{\eta_1}\boldsymbol{\eta}\cdot\boldsymbol{j}>M} 2^{p(\boldsymbol{r}-\frac{1}{p})\cdot\boldsymbol{j}} \|q_{\boldsymbol{j}}^L[f]|L_\infty(\mathbb{T}^d)\|^p\Big)^{\frac{1}{p}}.\end{aligned}$$

In case $p > 1$ we apply Hölder's inequality with $1 = \frac{1}{p} + \frac{1}{p'}$ and obtain

$$\|f - T_M^{L,\boldsymbol{\eta}} f|L_\infty(\mathbb{T}^d)\| \leq \Big(\sum_{\frac{1}{\eta_1}\boldsymbol{\eta}\cdot\boldsymbol{j}>M} 2^{-p'(\boldsymbol{r}-\frac{1}{p})\cdot\boldsymbol{j}}\Big)^{\frac{1}{p'}}\Big(\sum_{\frac{1}{\eta_1}\boldsymbol{\eta}\cdot\boldsymbol{j}>M} 2^{p(\boldsymbol{r}-\frac{1}{p})\cdot\boldsymbol{j}} \|q_{\boldsymbol{j}}^L[f]|L_\infty(\mathbb{T}^d)\|^p\Big)^{\frac{1}{p}}.$$

Lemma C.20 yields

$$\|f - T_M^{L,\boldsymbol{\eta}} f|L_\infty(\mathbb{T}^d)\| \lesssim M^{(\mu-1)(1-p)} 2^{-M(r_1-\frac{1}{p})}\Big(\sum_{\frac{1}{\eta_1}\boldsymbol{\eta}\cdot\boldsymbol{j}>M} 2^{p(\boldsymbol{r}-\frac{1}{p})\cdot\boldsymbol{j}} \|q_{\boldsymbol{j}}^L[f]|L_\infty(\mathbb{T}^d)\|^p\Big)^{\frac{1}{p}}.$$

Nikolskij's inequality (special case of Lemma 7.12) gives

$$\|f - T_M^{L,\boldsymbol{\eta}} f|L_\infty(\mathbb{T}^d)\| \lesssim M^{(\mu-1)(1-p)} 2^{-M(r_1-\frac{1}{p})}\Big(\sum_{\frac{1}{\eta_1}\boldsymbol{\eta}\cdot\boldsymbol{j}>M} 2^{p(\boldsymbol{r}-\frac{1}{p}+\frac{1}{\tilde{p}})\cdot\boldsymbol{j}} \|q_{\boldsymbol{j}}^L[f]|L_{\tilde{p}}(\mathbb{T}^d)\|^p\Big)^{\frac{1}{p}}.$$

In both cases Theorem 7.21 yields (8.1.5).
Step 2. The Jawerth-Franke type embedding implies

$$S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d) \hookrightarrow S^{\boldsymbol{r}-\frac{1}{p}+\frac{1}{\tilde{p}}}_{\tilde{p},p}B(\mathbb{T}^d)$$

(cf. Lemma 3.5). Applying this we obtain

$$\|f - T_M^{L,\boldsymbol{\eta}} f|L_\infty(\mathbb{T}^d)\| \lesssim M^{(\mu-1)(1-\frac{1}{p})_+} 2^{-M(r_1-\frac{1}{p})}\|f|S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d)\|,$$

which proves the claim. □

Remark 8.5. *It is remarkable that Theorem 8.3 allows to use the Smolyak algorithm based on the classical (nested) trigonometric interpolation (Dirichlet kernel) in case $1 < q \leq \infty$ although $p < q$ may be less than one. A similar observation has been made recently in [9, Rem. 6.12].*

In the remainder of this section we deal with Besov spaces $S^{\boldsymbol{r}}_{p,\theta}B(\mathbb{T}^d)$. A similar result as stated here was obtained by Dinh Dũng in [29], see also [27]. We contribute the case $\min\{p,\theta\} < 1$ for the Fourier analytical approach and allow the Dirichlet kernel ($L = 1$) for $q > 1$.

Theorem 8.6. *Let $0 < p < q < \infty$ and $0 < \theta \le \infty$. Additionally let $L > \frac{1}{q}$ and the smoothness vector $\boldsymbol{r} > \frac{1}{p}$ with (8.1.1). Then*

$$\|f - T_M^{L,\boldsymbol{\eta}} f|L_q(\mathbb{T}^d)\| \;\lesssim\; 2^{-M(r_1-\frac{1}{p}+\frac{1}{q})} M^{(\mu-1)(\frac{1}{q}-\frac{1}{\theta})_+} \|f|S^{\boldsymbol{r}}_{p,\theta}B(\mathbb{T}^d)\|$$

holds for all $M > 0$. The operator generating vector $\boldsymbol{\eta} \in \mathbb{R}^d$ is chosen as $\boldsymbol{\eta} = \boldsymbol{\nu} - \frac{1}{p} + \frac{1}{q}$, where $\boldsymbol{\nu} \in \mathbb{R}^d$ with

$$r_s = \nu_s, \quad s = 1, \ldots, \mu \quad \textit{and} \quad r_1 < \nu_s < r_s, \quad s = \mu+1, \ldots, d.$$

Proof. First we prove the case $q > 1$ with $\theta < \infty$. We find $\tilde{q} < q$ such that $L > \frac{1}{\tilde{q}} > \frac{1}{q}$ holds. The Jawerth-Franke embedding $S^{\frac{1}{\tilde{q}}-\frac{1}{q}}_{\tilde{q},q}B(\mathbb{T}^d) \subset L_q(\mathbb{T}^d)$ (cf. Lemma 3.5) yields

$$\|f - T_M^{L,\boldsymbol{\eta}} f|L_q(\mathbb{T}^d)\| \lesssim \|f - T_M^{L,\boldsymbol{\eta}} f|S^{\frac{1}{\tilde{q}}-\frac{1}{q}}_{\tilde{q},q}B(\mathbb{T}^d)\|$$

Expanding f into the series (7.5.5) and applying Theorem 7.13 gives

$$\|f - T_M^{L,\boldsymbol{\eta}} f|L_q(\mathbb{T}^d)\| \lesssim \Big(\sum_{\frac{1}{\eta_1}\boldsymbol{\eta}\cdot\boldsymbol{j} > M} 2^{q\boldsymbol{j}\cdot\boldsymbol{1}(\frac{1}{\tilde{q}}-\frac{1}{q})} \|q_{\boldsymbol{j}}[f]|L_{\tilde{q}}(\mathbb{T}^d)\|^q \Big)^{\frac{1}{q}}. \tag{8.1.6}$$

In case $\infty > \theta > q$ this can be estimated by using Hölder's inequality

$$\begin{aligned}\|f - T_M^{L,\boldsymbol{\eta}} f|L_q(\mathbb{T}^d)\| \;\lesssim\; & \Big(\sum_{\frac{1}{\eta_1}\boldsymbol{\eta}\cdot\boldsymbol{j} > M} 2^{-\frac{q\theta}{\theta-q}\boldsymbol{j}\cdot(\boldsymbol{r}-\frac{1}{p}+\frac{1}{q})} \Big)^{\frac{1}{q}-\frac{1}{\theta}} \\ & \times \Big(\sum_{\frac{1}{\eta_1}\boldsymbol{\eta}\cdot\boldsymbol{j} > M} 2^{\theta\boldsymbol{j}\cdot(\boldsymbol{r}-(\frac{1}{p}-\frac{1}{\tilde{q}}))} \|q_{\boldsymbol{j}}[f]|L_{\tilde{q}}(\mathbb{T}^d)\|^\theta \Big)^{\frac{1}{\theta}}.\end{aligned}$$

The estimate for the sum in Lemma C.20 gives

$$\|f - T_M^{L,\boldsymbol{\eta}} f|L_q(\mathbb{T}^d)\| \lesssim 2^{-M(r_1-\frac{1}{p}+\frac{1}{q})} M^{(\mu-1)(\frac{1}{q}-\frac{1}{\theta})} \Big(\sum_{\frac{1}{\eta_1}\boldsymbol{\eta}\cdot\boldsymbol{j} > M} 2^{\theta\boldsymbol{j}\cdot(\boldsymbol{r}-(\frac{1}{p}-\frac{1}{\tilde{q}}))} \|q_{\boldsymbol{j}}[f]|L_{\tilde{q}}(\mathbb{T}^d)\|^\theta \Big)^{\frac{1}{\theta}}.$$

In case $\theta \le q$ we use the embedding $\ell_\theta \hookrightarrow \ell_q$ and obtain

$$\|f - T_M^{L,\boldsymbol{\eta}} f|L_q(\mathbb{T}^d)\| \lesssim 2^{-M(r_1-\frac{1}{p}+\frac{1}{q})} \Big(\sum_{\frac{1}{\eta_1}\boldsymbol{\eta}\cdot\boldsymbol{j} > M} 2^{\theta\boldsymbol{j}\cdot(\boldsymbol{r}-(\frac{1}{p}-\frac{1}{\tilde{q}}))} \|q_{\boldsymbol{j}}[f]|L_{\tilde{q}}(\mathbb{T}^d)\|^\theta \Big)^{\frac{1}{\theta}}.$$

Theorem 7.21 allows to estimate

$$\|f - T_M^{L,\boldsymbol{\eta}} f|L_q(\mathbb{T}^d)\| \lesssim 2^{-M(r_1-\frac{1}{p}+\frac{1}{q})} M^{(\mu-1)(\frac{1}{q}-\frac{1}{\theta})_+} \|f|S^{\boldsymbol{r}-(\frac{1}{p}-\frac{1}{\tilde{q}})}_{\tilde{q},\theta}B(\mathbb{T}^d)\|.$$

Finally, the diagonal embedding stated in Lemma 3.4, (vi) yields

$$\|f - T_M^{L,\boldsymbol{\eta}} f | L_q(\mathbb{T}^d)\| \lesssim 2^{-M(r_1 - \frac{1}{p} + \frac{1}{q})} M^{(d-1)(\frac{1}{q} - \frac{1}{\theta})_+} \|f | S^{\boldsymbol{r}}_{p,\theta} B(\mathbb{T}^d)\|.$$

The case $q \leq 1$ is simpler. We expand f into the series (7.5.5). Then q-triangle inequality yields

$$\|f - T_M^{L,\boldsymbol{\eta}} f\|_q \lesssim \Big(\sum_{\frac{1}{\eta_1} \boldsymbol{\eta} \cdot \boldsymbol{j} > M} \|q_{\boldsymbol{j}}[f] | L_q(\mathbb{T}^d)\|^q \Big)^{\frac{1}{q}}.$$

The same case study as in the lines after (8.1.6) with $\tilde{q} = q$ finishes the proof. As usual in case $\theta = \infty$ we have to replace the corresponding sum by sup. $\square$

8.2 Linear sampling recovery

In this section we consider the optimality of convergence rates for linear sampling algorithms in case of Triebel-Lizorkin and Hölder-Nikolskij spaces with mixed smoothness, we abbreviate by $\boldsymbol{F}$. As a benchmark quantity we study linear sampling widths, cf. (1.4.1) in the introduction,

$$\varrho_n^{\mathrm{lin}}(\boldsymbol{F}, L_q(\mathbb{T}^d)) := \inf_{\substack{(\boldsymbol{\xi}_i)_{i=1}^n \subset \mathbb{T}^d \\ (\psi_i)_{i=1}^n \subset L_q(\mathbb{T}^d)}} \sup_{\|f|\boldsymbol{F}\| \le 1} \Big\| f - \sum_{i=1}^n f(\boldsymbol{\xi}_i)\psi_i \Big| L_q(\mathbb{T}^d) \Big\|.$$

This quantity can be interpreted as the minimal worst case error for the approximation of functions from the unit ball of $\boldsymbol{F}$ by linear algorithms using n function evaluations and where the error is measured in $L_q(\mathbb{T}^d)$. In case of $\boldsymbol{F} = S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d)$ with $\theta = 2$ and $1 < p < \infty$ we have the coincidence $S^{\boldsymbol{r}}_pW(\mathbb{T}^d) = S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d)$. This case is of special interest in this section because it denotes the probably most famous representative of the F-scale. Choosing m in (8.1.2) such that $n \gtrsim m^{\mu-1}2^m$ an upper bound for $\varrho_n^{\mathrm{lin}}(\boldsymbol{F}, L_q(\mathbb{T}^d))$ is provided by

$$\varrho_n^{\mathrm{lin}}(\boldsymbol{F}, L_q(\mathbb{T}^d)) \lesssim \sup_{\|f|\boldsymbol{F}\| \le 1} \|f - T_m^{L,\boldsymbol{\eta}} f | L_q(\mathbb{T}^d)\|.$$

Approximation with general linear information in case of mixed order Sobolev spaces $S^{\boldsymbol{r}}_pW(\mathbb{T}^d)$ and Hölder-Nikolskij spaces $S^{\boldsymbol{r}}_{p,\infty}B(\mathbb{T}^d)$ has been intensively studied in the past. We recall the concept of linear n-widths:

$$\lambda_n(\boldsymbol{F}, L_q(\mathbb{T}^d)) := \inf_{\substack{A:\boldsymbol{F}\to L_q(\mathbb{T}^d) \\ \operatorname{rank} A \le n}} \sup_{\|f|\boldsymbol{F}\| \le 1} \|f - A(f) | L_q(\mathbb{T}^d)\|. \tag{8.2.1}$$

In comparison to $\varrho_n^{\mathrm{lin}}(\boldsymbol{F}, L_q(\mathbb{T}^d))$ this quantity allows to benchmark linear operators using n pieces of linear information. Function evaluations are also linear information. Therefore, we have the relation

$$\lambda_n(\boldsymbol{F}, L_q(\mathbb{T}^d)) \le \varrho_n^{\mathrm{lin}}(\boldsymbol{F}, L_q(\mathbb{T}^d)).$$

That means linear n-widths can serve as lower bounds for linear sampling n-widths.

Corollary 8.7. *Let $0 < p < q < \infty$ and $0 < \theta \le \infty$. Additionally, the smoothness vector $\boldsymbol{r} > \frac{1}{p}$ is supposed to satisfy* (8.1.1). *Then*

$$\varrho_n^{lin}(S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d), L_q(\mathbb{T}^d)) \lesssim (n^{-1}\log^{(\mu-1)} n)^{r_1 - \frac{1}{p} + \frac{1}{q}}$$

holds for all $n > 0$.

Proof. The proof follows by Theorem 8.3 with the estimate from Lemma 8.1 for the number of function evaluations used by $T_M^{L,\boldsymbol{\eta}}$. □

Corollary 8.8. *Let $\boldsymbol{r} > \frac{1}{p}$ fulfilling* (8.1.1). *Furthermore, let $1 < p < q \leq 2$, $1 \leq \theta \leq \infty$ or $2 \leq p < q < \infty$, $2 \leq \theta \leq \infty$. Then*

$$\lambda_n(S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d), L_q(\mathbb{T}^d)) \asymp \varrho_n^{lin}(S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d), L_q(\mathbb{T}^d)) \asymp (n^{-1}\log^{(\mu-1)} n)^{(r_1-\frac{1}{p}+\frac{1}{q})}$$

for all $n \in \mathbb{N}$.

Proof. The upper bound for ϱ_n^{lin} follows by Corollary 8.7. The lower bound on λ_n is referred in Theorem D.25 in the Appendix. □

Remark 8.9. *The result stated above is not completely new. In case $2 \leq p, \theta < q$, ($\theta = q$) and $1 < p, q < 2$ with $\theta < q$ the upper bounds can be obtained with the help of Besov space results proven by Dinh Dũng in [29, 30] using the embedding relation $S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d) \hookrightarrow S^{\boldsymbol{r}}_{p,\max\{p,\theta\}}B(\mathbb{T}^d)$. Nevertheless, the cases $1 < p < q < 2$ with $\theta > q$ and $2 \leq p < q < \theta$ are new. Compared to Besov spaces in that range of parameters we do not observe an additional logarithmic factor in the convergence rate. This parameter range includes the situation of Sobolev spaces in case $1 < p < q < 2$.*

The following result is based on an observation by Novak/Triebel [83] for the univariate situation.

Theorem 8.10. *Let $1 < p < 2 < q < \infty$ and*

$$\boldsymbol{r} > \begin{cases} \frac{1}{p} & : \frac{1}{p} + \frac{1}{q} \geq 1,\ 1 \leq \theta \leq \infty, \\ \max\{\frac{1}{p}, 1 - \frac{1}{q}\} & : \frac{1}{p} + \frac{1}{q} \leq 1,\ 1 \leq \theta \leq \infty, \end{cases}$$

with (8.1.1). *Then*

$$\lambda_n(S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d), L_q(\mathbb{T}^d)) = o(\varrho_n^{lin}(S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d), L_q(\mathbb{T}^d))),$$

or more precisely

$$\lambda_n(S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d), L_q(\mathbb{T}^d)) \lesssim\!\!\!\!\!/\ \ n^{-(r_1-\frac{1}{p}+\frac{1}{q})} \lesssim \varrho_n^{lin}(S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d), L_q(\mathbb{T}^d))$$

holds for all $n > 0$.

Proof. The bounds for $\lambda_n(S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d), L_q(\mathbb{T}^d))$ come from the embedding $S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d) \hookrightarrow S^{\boldsymbol{r}}_{p,\infty}B(\mathbb{T}^d)$ that yields

$$\lambda_n(S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d), L_q(\mathbb{T}^d)) \leq \lambda_n(S^{\boldsymbol{r}}_{p,\infty}B(\mathbb{T}^d), L_q(\mathbb{T}^d)).$$

and the results from [42], see also [33, Thm. 4.46]. The proof for the (non-sharp) lower bound of $\varrho_n^{\text{lin}}(S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d), L_q(\mathbb{T}^d))$ follows from the univariate situation considered in [83, Theorem 23]. □

Remark 8.11. *The fact that the exponents of the main rate and the exponent of the logarithm in the upper bound obtained in Corollary 8.7 coincide and additionally the main rate is sharp seems to be a strong indication for the conjecture*

$$\varrho_n^{lin}(S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d), L_q(\mathbb{T}^d)) \asymp (n^{-1}\log^{\mu-1} n)^{r_1-\frac{1}{p}+\frac{1}{q}}$$

in case $1 < p < q < \infty$, $1 \leq \theta \leq \infty$ and $\boldsymbol{r} > \frac{1}{p}$ with (8.1.1).

Sharp lower bounds for $\lambda_n(S^{\boldsymbol{r}}_{p,\infty}B(\mathbb{T}^d), L_q(\mathbb{T}^d))$ obtained in [76] yield the following observation for Hölder-Nikolskij spaces.

Corollary 8.12. *Let* $1 < p < q \le 2$ *or* $2 \le p < q < \infty$ *and* $\boldsymbol{r} > \frac{1}{p}$ *is supposed to satisfy* (8.1.1). *Then*

$$\varrho_n^{lin}(S^{\boldsymbol{r}}_{p,\infty}B(\mathbb{T}^d), L_q(\mathbb{T}^d)) \asymp \lambda_n(S^{\boldsymbol{r}}_{p,\infty}B(\mathbb{T}^d), L_q(\mathbb{T}^d)) \asymp (n^{-1}\log^{(\mu-1)} n)^{r_1-\frac{1}{p}+\frac{1}{q}}(\log n)^{\frac{\mu-1}{q}}$$

holds for all $n > 0$.

Proof. The upper bound was originally obtained by Dinh Dũng in [25]. The lower bound for linear widths is due to Galeev [42]. In our context the upper bound for ϱ_n^{lin} follows by Theorem 8.6 with the estimate from Lemma 8.1 for the number of function evaluations used by $T_M^{L,\boldsymbol{\eta}}$. The lower bound for λ_n in the second case was proven recently by Malykhin and Ryutin [76], see also [42] and [33, Thm. 4.46]. □

Corollary 8.13. *Let* $1 < p < 2 < q < \infty$ *and* $\boldsymbol{r} > \begin{cases} \frac{1}{p} & : \frac{1}{p} + \frac{1}{q} \ge 1, \\ \max\{\frac{1}{p}, 1 - \frac{1}{q}\} & : \frac{1}{p} + \frac{1}{q} < 1, \end{cases}$ *fulfilling* (8.1.1). *Then*

$$\lambda_n(S^{\boldsymbol{r}}_{p,\infty}B(\mathbb{T}^d), L_q(\mathbb{T}^d)) = o(\varrho_n^{lin}(S^{\boldsymbol{r}}_{p,\infty}B(\mathbb{T}^d), L_q(\mathbb{T}^d))),$$

or more precisely

$$\lambda_n(S^{\boldsymbol{r}}_{p,\infty}B(\mathbb{T}^d), L_q(\mathbb{T}^d)) \not\gtrsim n^{-(r_1-\frac{1}{p}+\frac{1}{q})} \lesssim \varrho_n^{lin}(S^{\boldsymbol{r}}_{p,\infty}B(\mathbb{T}^d), L_q(\mathbb{T}^d))$$

holds for all $n > 0$.

Proof. The bounds for $\lambda_n(S^{\boldsymbol{r}}_{p,\infty}B(\mathbb{T}^d), L_q(\mathbb{T}^d))$ come from [42]. The proof for the (non-sharp) lower bounds for $\varrho_n^{\mathrm{lin}}(S^{\boldsymbol{r}}_{p,\infty}B(\mathbb{T}^d), L_q(\mathbb{T}^d))$ follow from the univariate situation considered in [83, Theorem 23]. □

Corollary 8.14. *Let* $0 < p, \theta < \infty$ *(*$\theta = \infty$*) and the smoothness vector* $\boldsymbol{r} > \frac{1}{p}$ *which is supposed to satisfy* (8.1.1) *be given. Then*

$$\varrho_n^{lin}(S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d), L_\infty(\mathbb{T}^d)) \lesssim (n^{-1}\log^{\mu-1} n)^{r_1-\frac{1}{p}} \log^{(\mu-1)(1-\frac{1}{p})_+} n$$

holds for all $n > 0$.

Proof. The upper bound follows by Theorem 8.4 with the estimate from Lemma 8.1 for the number of function evaluations. □

Based on a recent observation of Nguyen in [78, Theorem 2.15] we can state the following theorem:

Corollary 8.15. *Let* $1 < p < 2$ *and* $\boldsymbol{r} > 1$ *fulfilling* (8.1.1). *Then*

$$\lambda_n(S^r_pW(\mathbb{T}^d), L_\infty(\mathbb{T}^d)) = o(\varrho_n^{lin}(S^r_pW(\mathbb{T}^d), L_\infty(\mathbb{T}^d))),$$

or more precisely

$$\lambda_n(S^r_pW(\mathbb{T}^d), L_\infty(\mathbb{T}^d)) \asymp n^{-(r_1-\frac{1}{2})}(\log^{\mu-1} n)^{r_1} \not\gtrsim n^{-(r_1-\frac{1}{p})} \lesssim \varrho_n^{lin}(S^r_pW(\mathbb{T}^d), L_\infty(\mathbb{T}^d))$$

holds for all $n > 0$.

Proof. The bound for $\lambda_n(S^r_pW(\mathbb{T}^d), L_\infty(\mathbb{T}^d))$ comes from Theorem D.27. The proof for the (non-sharp) lower bound for $\varrho_n^{\mathrm{lin}}(S^r_pW(\mathbb{T}^d), L_\infty(\mathbb{T}^d))$ follows from the univariate situation considered in [83, Theorem 23]. □

The next result was originally observed by Temlyakov, [113]. Sampling representations allow to reproduce it.

Corollary 8.16. *Let* $\boldsymbol{r} > 1$ *fulfilling* (8.1.1). *Then*

$$\varrho_n(S^{\boldsymbol{r}}_2W(\mathbb{T}^d), L_\infty(\mathbb{T}^d)) \asymp \lambda_n(S^{\boldsymbol{r}}_2W(\mathbb{T}^d), L_\infty(\mathbb{T}^d)) \asymp n^{-(r_1-\frac{1}{2})}\log^{d-1} n,$$

holds for all $n > 0$.

Proof. The bound for $\lambda_n(S^r_pW(\mathbb{T}^d), L_\infty(\mathbb{T}^d))$ comes from Theorem D.27 and the upper bound for $\varrho_n(S^{\boldsymbol{r}}_2W(\mathbb{T}^d), L_\infty(\mathbb{T}^d))$ from Theorem 8.4 □

8.3 Sampling recovery and Gelfand n-widths

The considerations above cover linear algorithms in the classical sense. Last but not least we consider an extension of this concept, so-called approximation using standard information, cf. [84, 85]. This means we consider algorithms that are defined as a composition of a linear information map and a possibly non-linear reconstruction operator. To avoid further technicalities we restrict to Banach spaces $\boldsymbol{F}$ that are either Sobolev spaces $S^{\boldsymbol{r}}_pW(\mathbb{T}^d)$ or Hölder-Nikolskij spaces $S^{\boldsymbol{r}}_{p,\infty}B(\mathbb{T}^d)$ in this subsection. The non-linear sampling widths were defined in (1.4.2). The following relation clearly holds true

$$\varrho_n(\boldsymbol{F}, L_q(\mathbb{T}^d) \le \varrho_n^{\mathrm{lin}}(\boldsymbol{F}, L_q(\mathbb{T}^d))\,.$$

Therefore (possibly non-sharp) upper bounds for sampling widths are always provided by linear sampling widths. To consider questions on optimality of these bounds we consider Gelfand n-widths

$$c_n(\boldsymbol{F}, L_q(\mathbb{T}^d)) := \inf_{\substack{B:\,\boldsymbol{F}\to\mathbb{C}^n\\ \text{linear}}} \sup_{\substack{\|f|\boldsymbol{F}\|\le 1\\ f\in\ker B}} \|f|L_q(\mathbb{T}^d)\|. \tag{8.3.1}$$

Here B denotes a general linear mapping $B : \boldsymbol{F} \to \mathbb{C}^n$. This means c_n measures the minimal (over all information mappings) worst case distance of elements in the unit ball of $\boldsymbol{F}$ which can not be distinguished by the information mapping B. This immediately gives

$$c_n(\boldsymbol{F}, L_q(\mathbb{T}^d)) \lesssim \varrho_n(\boldsymbol{F}, L_q(\mathbb{T}^d)).$$

Note that (1.4.3) is actually the definition of the nth "Gelfand numbers", which we call "Gelfand n-width" here. For a thorough discussion on the relation between Gelfand numbers and suitable worst-case errors we refer to the recent paper [21, Rem. 2.3]. Since Gelfand widths for embeddings $id : S^{\boldsymbol{r}}_pW(\mathbb{T}^d) \to L_q(\mathbb{T}^d)$ are not studied directly we use a duality relation to Kolmogorov n-widths, cf. (D.1).

Lemma 8.17. *The following duality relation holds true*

$$d_n(T : X \to Y) = c_n(T' : Y' \to X'),$$

where T' denotes the adjoint operator of T and X', Y' the topological dual spaces of X and Y.

Proof. We refer to [91, Theorem 6.2]. $\square$

Corollary 8.18. *Let $1 < p, q < \infty$ and $\boldsymbol{r} > \begin{cases} \frac{1}{2} & : \quad 1 < p < q \leq 2, \\ 1 - \frac{1}{q} & : \quad p < 2 < q, \\ (\frac{1}{p} - \frac{1}{q})_+ & : \quad \text{otherwise,} \end{cases}$*

with

$$r_1 = \ldots = r_\mu < r_{\mu+1} \leq \ldots \leq r_d < \infty. \tag{8.3.2}$$

Then

$$c_n(S^{\boldsymbol{r}}_p W(\mathbb{T}^d), L_q(\mathbb{T}^d)) \asymp (n^{-1} \log^{\mu-1} n)^{r_1 - (\min\{\frac{1}{p}, \frac{1}{2}\} - \frac{1}{q})_+}$$

for all $n \in \mathbb{N}$.

Proof. The proof follows by the duality relation stated in Lemma 8.17 and a lifting argument. The topological dual spaces of $S^{\boldsymbol{r}}_p W(\mathbb{T}^d)$ and $L_q(\mathbb{T}^d)$ are the spaces $S^{-\boldsymbol{r}}_{p'} W(\mathbb{T}^d)$ and $L_{q'}(\mathbb{T}^d)$ with $1 = \frac{1}{p} + \frac{1}{p'} = \frac{1}{q} + \frac{1}{q'}$. Lemma 8.17 yields

$$c_n(S^{\boldsymbol{r}}_p W(\mathbb{T}^d), L_q(\mathbb{T}^d)) = d_n(L_{q'}(\mathbb{T}^d), S^{-\boldsymbol{r}}_{p'} W(\mathbb{T}^d)).$$

Finally we show the identity

$$d_n(L_{q'}(\mathbb{T}^d), S^{-\boldsymbol{r}}_{p'} W(\mathbb{T}^d)) \asymp d_n(S^{\boldsymbol{r}}_{q'} W(\mathbb{T}^d), L_{p'}(\mathbb{T}^d)).$$

For that reason we consider the lifting operator $I_{\boldsymbol{r}}$ in $D'(\mathbb{T}^d)$ given by

$$I_{\boldsymbol{r}} : f = \sum_{\boldsymbol{k} \in \mathbb{Z}^d} \widehat{f}(\boldsymbol{k}) e^{i\boldsymbol{k}\boldsymbol{x}} \mapsto \sum_{\boldsymbol{k} \in \mathbb{Z}^d} \widehat{f}(\boldsymbol{k}) \Big(\prod_{i=1}^{d} (1 + |k_i|^2)^{-\frac{r_i}{2}} \Big) e^{i\boldsymbol{k}\boldsymbol{x}}.$$

It is easy to check that this is an isometry that maps $f \in S^{\alpha}_p W$ to $I_{\boldsymbol{r}} f \in S^{\boldsymbol{\alpha}+\boldsymbol{r}}_p W$, $\alpha \in \mathbb{R}$ with $(I_{\boldsymbol{r}})^{-1} = I_{-\boldsymbol{r}}$. Therefore we may use the commutative diagram,

$$\begin{array}{ccc} L_{q'}(\mathbb{T}^d) & \xrightarrow{id_1} & S^{-\boldsymbol{r}}_{p'} W(\mathbb{T}^d) \\ \Big\downarrow {\scriptstyle I_{\boldsymbol{r}}} & & \Big\uparrow {\scriptstyle I_{-\boldsymbol{r}}} \\ S^{\boldsymbol{r}}_{q'} W(\mathbb{T}^d) & \xrightarrow{id_2} & L_{p'}(\mathbb{T}^d) \end{array}$$

which allows to describe the operators id_1, id_2 by

$$id_1 = I_{-\boldsymbol{r}} \circ id_2 \circ I_{\boldsymbol{r}} \quad \text{and} \quad id_2 = I_{\boldsymbol{r}} \circ id_1 \circ I_{-\boldsymbol{r}}.$$

Kolmogorov widths are s-numbers and fulfill a multiplicativity property that yields

$$d_n(id_1) = d_n(I_{-\boldsymbol{r}} \circ id_2 \circ I_{\boldsymbol{r}}) \leq \|I_{-\boldsymbol{r}}\| d_n(id_2) \|I_{\boldsymbol{r}}\| \asymp d_n(id_2)$$

and

$$d_n(id_2) = d_n(L_{\boldsymbol{r}} \circ id_1 \circ L_{-\boldsymbol{r}}) \leq \|L_{\boldsymbol{r}}\| d_n(id_1) \|I_{-\boldsymbol{r}}\| \asymp d_n(id_1).$$

Inserting the result from Theorem D.26 finishes the proof. □

Surprisingly, a new result in [76] allows us to prove the following results for Gelfand n-widths of Hölder spaces $S^{\boldsymbol{r}}_{p,\infty}B(\mathbb{T}^d)$.

Theorem 8.19. *Let* $1 < p < q < \infty$ *and* $\boldsymbol{r}$ *with*

$$(1/p - 1/q)_+ < r_1 = \ldots = r_\mu < r_{\mu+1} \leq \ldots \leq r_d < \infty$$

then

$$c_n(S^{\boldsymbol{r}}_{p,\infty}B(\mathbb{T}^d), L_q(\mathbb{T}^d)) \asymp \begin{cases} \left(\frac{\log^{\mu-1} n}{n}\right)^{r-\frac{1}{2}+\frac{1}{q}} (\log n)^{\frac{\mu-1}{q}} & : \ \frac{1}{p}+\frac{1}{q} < 1, \ p \leq 2, \ r_1 > 1 - \frac{1}{q}, \\ \left(\frac{\log^{\mu-1} n}{n}\right)^{r-\frac{1}{p}+\frac{1}{q}} (\log n)^{\frac{\mu-1}{q}} & : \ 2 \leq p < q. \end{cases}$$

Proof. The upper bounds follow from the results for linear widths in [42]. The lower bounds are new. Malykhin and Ryutin proved in [76] the following bound on Kolmogorov n-widths for finite dimensional normed spaces $\ell_p^M(\ell_q^N)$

$$d_{\lfloor \frac{NM}{2} \rfloor}(\ell_\infty^M(\ell_1^N), \ell_1^M(\ell_2^N)) \asymp M. \tag{8.3.3}$$

In the first case the technique for the lower bounds on linear widths presented in [42] works well also for Gelfand n-widths. The discretization stated there yields

$$c_n(S^{\boldsymbol{r}}_{p,\infty}B(\mathbb{T}^d), L_q(\mathbb{T}^d)) \gtrsim 2^{u(-r+\frac{1}{2}-\frac{1}{q})} c_n(\ell_\infty^{u^{\mu-1}}(\ell_2^{2^u}), \ell_q^{u^{\mu-1}2^u}).$$

The duality relation in Lemma 8.17 gives

$$c_n(S^{\boldsymbol{r}}_{p,\infty}B(\mathbb{T}^d), L_q(\mathbb{T}^d)) \gtrsim 2^{u(-r+\frac{1}{2}-\frac{1}{q})} d_n(\ell_{q'}^{u^{\mu-1}2^u}, \ell_1^{u^{\mu-1}}(\ell_2^{2^u})).$$

Applying Hölder's inequality in finite dimensional spaces $\ell_p^M(\ell_q^N)$ yields the following estimate

$$c_n(S^{\boldsymbol{r}}_{p,\infty}B(\mathbb{T}^d), L_q(\mathbb{T}^d)) \gtrsim 2^{u(-r+\frac{1}{2}-\frac{1}{q})} u^{-\frac{\mu-1}{q'}} d_n(\ell_\infty^{u^{\mu-1}}(\ell_1^{2^u}), \ell_1^{u^{\mu-1}}(\ell_2^{2^u})).$$

Choosing $n \asymp u^{\mu-1}2^u$ then the relation in (8.3.3) implies

$$c_n(S^{\boldsymbol{r}}_{p,\infty}B(\mathbb{T}^d), L_q(\mathbb{T}^d)) \gtrsim 2^{u(-r+\frac{1}{2}-\frac{1}{q})} u^{\frac{\mu-1}{q}} \asymp \left(\frac{\log^{\mu-1} n}{n}\right)^{r-\frac{1}{2}+\frac{1}{q}} (\log n)^{\frac{\mu-1}{q}}.$$

The second case is obtained by the embedding

$$S_{2,\infty}^{\boldsymbol{r}-(\frac{1}{2}-\frac{1}{p})} B(\mathbb{T}^d) \hookrightarrow S^{\boldsymbol{r}}_{p,\infty}B(\mathbb{T}^d)$$

together with the result from the first case. □

Corollary 8.20. *Let $2 \leq p < q < \infty$ and $\boldsymbol{r} > \frac{1}{p}$ fulfilling* (8.1.1)*. Then*

(i)

$$\begin{aligned}\varrho_n(S^{\boldsymbol{r}}_pW(\mathbb{T}^d), L_q(\mathbb{T}^d)) &\asymp c_n(S^{\boldsymbol{r}}_pW(\mathbb{T}^d), L_q(\mathbb{T}^d)) \asymp \varrho^{lin}_n(S^{\boldsymbol{r}}_pW(\mathbb{T}^d), L_q(\mathbb{T}^d)) \\ &\asymp \lambda_n(S^{\boldsymbol{r}}_pW(\mathbb{T}^d), L_q(\mathbb{T}^d)) \asymp (n^{-1}\log^{\mu-1} n)^{r_1-\frac{1}{p}+\frac{1}{q}},\end{aligned}$$

(ii)

$$\begin{aligned}\varrho_n(S^{\boldsymbol{r}}_{p,\infty}B(\mathbb{T}^d), L_q(\mathbb{T}^d)) &\asymp c_n(S^{\boldsymbol{r}}_{p,\infty}B(\mathbb{T}^d), L_q(\mathbb{T}^d)) \asymp \varrho^{lin}_n(S^{\boldsymbol{r}}_{p,\infty}B(\mathbb{T}^d), L_q(\mathbb{T}^d)) \\ &\asymp \lambda_n(S^{\boldsymbol{r}}_{p,\infty}B(\mathbb{T}^d), L_q(\mathbb{T}^d)) \\ &\asymp (n^{-1}\log^{\mu-1} n)^{r_1-\frac{1}{p}+\frac{1}{q}}(\log^{\frac{\mu-1}{q}} n),\end{aligned}$$

holds for all $n \in \mathbb{N}$.

Proof. The proof follows by Theorems 8.3, 8.6, 8.19 and Corollary 8.18. □

Remark 8.21. *In the parameter range $2 < p < q < \infty$ permitting non-linear reconstruction operators does not yield better results. Optimal rates can be achieved by completely linear sampling algorithms.*

We obtain the following counterpart of Theorem 8.10 for non-linear sampling.

Corollary 8.22. *Let $1 < p < 2 < q < \infty$ and $\boldsymbol{r} > \max\{\frac{1}{p}, 1-\frac{1}{q}\}$ fulfilling* (8.1.1)*. Additionally let $\boldsymbol{F}$ denote either $S^{\boldsymbol{r}}_pW(\mathbb{T}^d)$ or $S^{\boldsymbol{r}}_{p,\infty}B(\mathbb{T}^d)$. Then*

$$c_n(\boldsymbol{F}, L_q(\mathbb{T}^d)) = o(\varrho_n(\boldsymbol{F}, L_q(\mathbb{T}^d))),$$

or more precisely

$$c_n(\boldsymbol{F}, L_q(\mathbb{T}^d)) \lesssim\!\!\!\!\!\!/\;\; n^{-(r_1-\frac{1}{p}+\frac{1}{q})} \lesssim \varrho_n(\boldsymbol{F}, L_q(\mathbb{T}^d))$$

holds for all $n \in \mathbb{N}$.

Proof. The proof can be obtained by following the construction of the lower bound for the univariate situation in [83], where we recognize that the stronger inequality

$$\varrho_n(\boldsymbol{F}, L_q(\mathbb{T}^d)) \geq \inf_{(\xi_k)_{k=1}^n \subset \mathbb{T}^d} \sup_{\substack{\|f|\boldsymbol{F}\|\leq 1 \\ f(\xi_k)=0,\, k=1,\ldots,n}} \|f|L_q(\mathbb{T}^d)\|$$

holds. The estimates for $c_n(S^{\boldsymbol{r}}_pW(\mathbb{T}^d), L_q(\mathbb{T}^d))$ were obtained in Corollary 8.18. For $S^{\boldsymbol{r}}_{p,\infty}B(\mathbb{T}^d)$ we refer to Theorem 8.19. Gelfand numbers for more general Besov spaces were studied in [80]. □

Remark 8.23. *As a consequence of the lower bound in Corollary 8.22 for $\varrho_n(\boldsymbol{F}, L_q(\mathbb{T}^d))$, we obtain that in the parameter range $1 < p < 2 < q < \infty$ even linear approximation behaves significantly better than sampling recovery with a possibly non-linear reconstructing operator.*

Chapter 9

Outlook and open problems

We discuss some research aspects and questions that were left open at the end of our studies and require further research.

9.1 Sampling: same integrability in target and source space

Using a trigonometric sparse grid sampling operator Temlyakov [117] proved for $r > \frac{1}{p}$, $1 < p < \infty$ that

$$\varrho_n^{\text{lin}}(S^r_{p,\infty}B(\mathbb{T}^d), L_p(\mathbb{T}^d)) \lesssim (n^{-1}\log^{d-1} n)^r(\log^{d-1} n) \tag{9.1.1}$$

holds. Later, Sickel [102, 103] contributed to the 2-dimensional case and Sickel, Ullrich [104] for general $d > 1$ with $1 \le \theta \le \infty$ the (best) today known upper bounds

$$\varrho_n^{\text{lin}}(S^r_{p,\theta}B(\mathbb{T}^d), L_p(\mathbb{T}^d)) \lesssim (n^{-1}\log^{d-1} n)^r(\log^{d-1} n)^{1-\frac{1}{\theta}}, \quad r > \frac{1}{p},$$

$$\varrho_n^{\text{lin}}(S^r_{p}W(\mathbb{T}^d), L_p(\mathbb{T}^d)) \lesssim (n^{-1}\log^{d-1} n)^r(\log^{d-1} n)^{\frac{1}{2}}, \quad r > \max\Big\{\frac{1}{p}, \frac{1}{2}\Big\}. \tag{9.1.2}$$

The upper bounds in (9.1.1) and (9.1.2) have in common that the sharp estimates for linear widths λ_n (defined in (8.2.1))

$$\begin{aligned} \lambda_n(S^r_{p,\infty}B(\mathbb{T}^d), L_p(\mathbb{T}^d)) &\asymp (n^{-1}\log^{d-1} n)^r(\log^{d-1} n)^{\frac{1}{2}}, \quad p \ge 2, \quad \text{cf. [116]} \\ \lambda_n(S^r_{p}W(\mathbb{T}^d), L_p(\mathbb{T}^d)) &\asymp (n^{-1}\log^{d-1} n)^r, \quad 1 < p < \infty, \quad \text{cf. Theorem } D.24 \end{aligned}$$

do not coincide with the estimates for ϱ_n^{lin}, which are typically used to obtain lower bounds for ϱ_n^{lin}. A logarithmic gap appears. In fact, it is unknown whether linear approximation based on information generated by general linear functionals behaves better as linear approximation by sampling values. As a consequence of Chapter 5 (Theorem 5.14) we know that linear operators which sample functions on sparse grids behave worse compared to approximation with general linear information. For general point sets we have no indication concerning this phenomenon. Considering the limiting case $r = 2$ Bungartz, Griebel [8] proved for the Faber-Schauder sparse grid operator I_M defined in (5.1.1) the convergence rate

$$\begin{aligned} \|f - I_M f|L_2([0,1]^d\| &\lesssim M^{d-1}2^{-2M}\|f|S^2_pW([0,1]^d)\| \\ &\asymp (n^{-1}\log^{d-1} n)^2 \log^{d-1} n\|f|S^2_pW([0,1]^d)\|, \end{aligned}$$

see also Theorem 5.7. The method for the lower bound in Theorem 5.14 allows to prove for the sparse grid width

$$g^{SG}_n(S^2_2W([0,1]^d), L_2[(0,1)]^d) \gtrsim (n^{-1}\log^{d-1} n)^2 \log^{\frac{d-1}{2}} n.$$

In fact there is a gap of $\log^{\frac{d-1}{2}} n$ for the knowledge of the exact asymptotic approximation rate of I_M. It would be interesting to know whether the limited regularity of the hat functions causes a little worse approximation rate in the limiting case.

9.2 Higher smoothness in the non-periodic case

In the first (non-periodic) part of this thesis we are restricted to a maximal smoothness at around 2. This is caused by the limited smoothness of the Faber-Schauder system. To be more precise the Faber-Schauder hat functions belong to the spaces S^r_pW with $r < 1+\frac{1}{p}$. In fact, it is interesting that we can overcome this smoothness limitation up to a certain degree in our approximation results and benefit in the convergence rate from $r > 1+\frac{1}{p}$ up to a certain level. Nevertheless, these possibilities are limited. Dealing with significantly more smoothness would require us to switch to smoother basis functions. Triebel suggested in [120] so called Faber splines. They generalize the integration step going from Haar to Faber-Schauder bases. The idea is to start with a $(\ell-1)$-times continuous differentiable spline function $h^\ell(x)$. Representing the $(\ell+1)$-th derivative of f in terms of this system. Then integrating $(\ell+1)$-times gives an expansion of f by 2ℓ times continuous differentiable Faber splines $v^\ell(x)$ which allow a representation with coefficients generated by function evaluations of f. In [120] this theory was considered as an outlook. We do not know about further research in this direction. Another related approach are B-Splines introduced by I.J. Schoenberg (see also [16]). They are generated as an iterated convolution of characteristic functions. Dinh Dũng took up this concept and studied them successfully as a basis in $S^r_{p,\theta}B([0,1]^d)$, cf. [29, 30]. (B-)splines of higher order have the property that the supports of different translation are generally not disjoint as it is the case for the Faber-Schauder system. There is some overlap. Dealing with Sobolev spaces $S^r_pW(\mathbb{R}^d)$ the $L_p(\mathbb{R}^d)$ integration in the norm runs over all dilation levels. This makes a careful analysis much harder than in case of Besov spaces $S^r_{p,\theta}B(\mathbb{R}^d)$ and requires non-trivial tools from harmonic analysis. Based on approaches from an early preprint of the current thesis [31] was created. Here the author proves sampling representations using B-splines for periodic spaces $S^r_pW(\mathbb{T}^d)$. Nevertheless, the conditions stated there seem to be not sharp and can be improved with methods presented in this thesis. A carefully proven B-spline representation for $S^r_pW([0,1]^d)$ would allow to lift the results in Chapter 5 and 6 to any smoothness $r > 2$. A further approach for higher smoothness is the interpolation scheme by Deslauriers and Dubuc [24, 17, 23]. In [22] a result discretizing univariate Triebel-Lizorkin and Besov spaces using this interpolation scheme was stated. An extension to the case of dominating mixed smoothness would be interesting. As a last approach we mention higher order hierarchical basis introduced by Bungartz [7] which also use piecewise polynomials as basis functions.

9.3 Tractability and preasymptotics for standard information

In the present thesis convergence rates of type $C_d m^{-r}(\log^{d-1} m)^r$ appeared at several points. The constants C_d or at least their behavior for growing problem dimension d was mostly not calculated explicitly To approximate a function and estimate the number of information we have to spend to achieve a given accuracy the size of C_d will be very important to obtain useful estimates. The notion tractability from the mathematical area of information based complexity studies a quantity called information based complexity, which is defined as the minimal number of information required to approximate the compact embedding $id : X \to Y$ up to a certain given accuracy ε. One distinguishes in standard (samples) and general information (linear information)

$$\begin{aligned} n^{\mathrm{all}}(\varepsilon) &:= \inf\{n \in \mathbb{N} : c_n(B_X)_Y < \varepsilon\}, \\ n^{\mathrm{std}}(\varepsilon) &:= \inf\{n \in \mathbb{N} : \varrho_n(B_X)_Y < \varepsilon\}, \end{aligned}$$

(for the definition of Gelfand widths c_n see (1.4.3), sampling widths ϱ_n see (1.4.2)). Based on the behavior of $n^*(\varepsilon)$, $* \in \{\text{all,std}\}$ in d we assign the approximation problem to a tractability class. If $n^*(\varepsilon)$ increases exponentially in d then we speak about the curse of dimensionality see [84, 85, 86] and the references therein. For sampling approximation in spaces of dominating mixed smoothness nearly nothing is known in this direction. Explicit knowledge of the constants would allow us to translate our convergence rates into bounds for the information based complexity quantity. In [70, 71, 15] approximation with linear information in the sense of linear widths was considered. It turned out that tractability issues heavily depend on the explicit choice of the norm in the space of functions we consider, since different equivalent norms can essentially modify the unit ball of the respective norm with respect to d. Let us have a look on a closely related problem. Considering the function $f_d(t) = t^{-r}\log^{(d-1)r} t$ (related to our convergence rates) we recognize that this function is monotonically increasing for $t \in [1, e^{d-1}]$ and decreasing on $[e^{d-1}, \infty)$. Much later this function becomes smaller than 1. In fact, for $n < e^{d-1}$ samples the estimates make little sense, since they are increasing. In [70, 71, 69] the authors study so called preasymptotic rates. Convergence rates that are valid only for small degrees of freedom but that provide in this range decreasing with explicitly known constants. Concerning Monte-Carlo sampling approximation such an approach was considered in [68]. It would be of great interest to have similar results in the deterministic worst case setting which is considered in this thesis.

9.4 Sharp bounds for best m-term approximation

In Section 6.4 we studied best m-term approximation with respect to the Faber-Schauder system. The lower bounds

$$\sigma_m(S^r_{p,\theta}X, \mathbb{F}^d)_{L_q} \gtrsim m^{-r}, \quad X \in \{B, F\},$$

provided in Theorem 6.20 is our exclusive source to bound the corresponding best m-term quantity from below. The fooling argument in the proof is basically an univariate one, that is not able to generate d-dependent logarithms. Actually we saw in Theorem 6.21 that this provides sharp results for the small smoothness case, so we do not have to expect logarithms in general. For the large smoothness case (cf. Theorem 6.23) our upper bounds contain d-dependent logarithms

$$m^{-r} \lesssim \sigma_m(S^r_{p,\theta}X([0,1]^d), \mathbb{F}^d)_{L_q} \lesssim m^{-r}(\log m)^{(d-1)(r+1-\frac{1}{\theta})}, \; X \in \{B, F\}.$$

Comparing this to sharp results for Daubechies wavelets obtained by Hansen, Sickel [56], a logarithm seems to be required. In fact, we need an improved lower bound for the case of large smoothness. In case of Daubechies Wavelets vanishing moments allow to discretize L_q spaces into corresponding sequence spaces. Hence, lower bounds obtained in sequence spaces imply lower bounds for best m-term approximation. Our upper bounds for best m-term approximation with respect to the Faber-Schauder system use as a vehicle the discretization of the space $S^0_{q,1}B([0,1]^d)$ which is embedded into $L_q([0,1]^d)$. Hence, our upper bounds coincide with wavelet upper bounds that are calculated for fine index $\nu = 1$ in the target space. Littlewood-Paley (wavelet) theory shows that fine index $\nu = 2$ is the optimal one to discretize $L_q([0,1])$. For that reason the upper bonds in the large smoothness case we obtained behave by a factor $(\log^{d-1} m)^{\frac{1}{2}}$ worse compared to the large smoothness results for Daubechies Wavelets. It is not clear whether this is a technical difficulty or a serious deficiency comparing best m-term approximation for Daubechies Wavelets to best m-term for the Faber-Schauder dictionary.

9.5 Optimal sampling recovery in case $1 < p < 2 < q < \infty$

Groundbreaking innovations [46, 64] for the approximation of sequence spaces in the 1980s allowed Galeev [41, 42] to prove an interesting behavior of linear widths for the embedding $S^r_pW(\mathbb{T}^d) \to L_q(\mathbb{T}^d)$ in the parameter region $1 < p < 2 < q < \infty$.

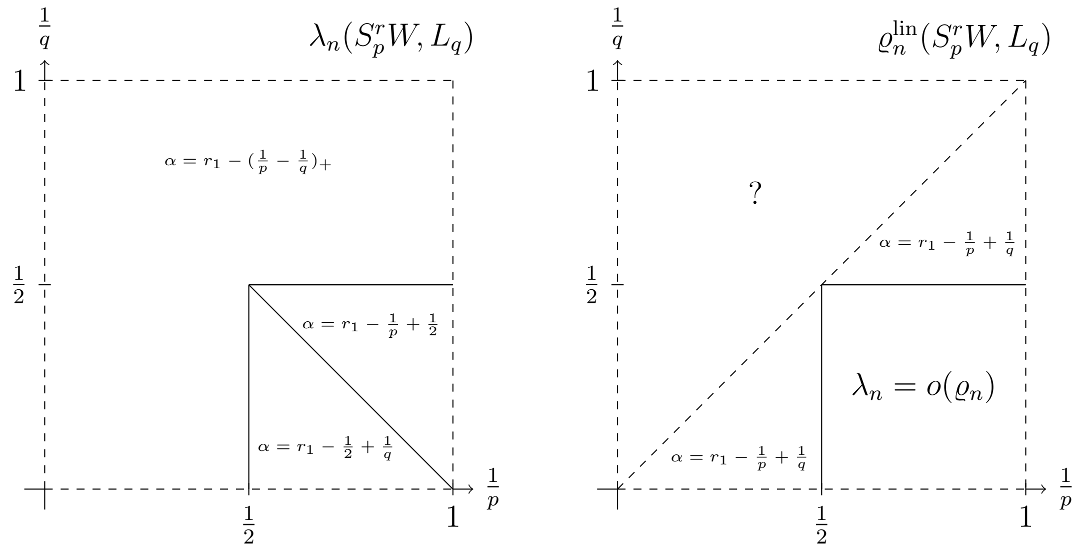


Figure 9.1: The parameter α refers to the sharp rate $(\frac{(\log n)^{d-1}}{n})^\alpha$.

In case $1 < p < q \le 2$ or $2 \le p < q < \infty$ the approximation rates are of type $(n^{-1}\log^{d-1} n)^\alpha$ with $\alpha = r - \frac{1}{p} + \frac{1}{q}$. Hence, they are the smoothness minus the difference of the integrabilities. In case $1 < p < 2 < q < \infty$ one integrability gets stuck at 2, cf. Theorem D.24

$$\lambda_n(S^r_pW(\mathbb{T}^d), L_q(\mathbb{T}^d)) \asymp \begin{cases} \left(\frac{(\log n)^{d-1}}{n}\right)^{r-1/p+1/2} & : \ 1/p + 1/q \ge 1, \ q > 2, \ r > 1/p, \\ \left(\frac{(\log n)^{d-1}}{n}\right)^{r-1/2+1/q} & : \ 1/p + 1/q \le 1, \ p < 2, \ r > 1 - 1/q. \end{cases}$$

This provides an improved rate. In case of linear sampling recovery (cf. Section 8.2) such an effect does not happen or at least not in the main rate. In Theorem 8.10 we show for $1 < p < 2 < q < \infty$ the relation

$$\lambda_n(S^r_pW(\mathbb{T}^d), L_q(\mathbb{T}^d)) \lesssim\!\!\!\!\!/\;\, n^{-(r-\frac{1}{p}+\frac{1}{q})} \lesssim \varrho^{\text{lin}}_n(S^r_pW(\mathbb{T}^d), L_q(\mathbb{T}^d)) \lesssim (n^{-1}\log^{d-1} n)^{r-\frac{1}{p}+\frac{1}{q}}.$$

The lower bound in the middle is based on a univariate fooling function argument by Novak, Triebel [83]. We conclude that approximation by linear information behaves significant better than sampling approximation. What remains unknown is the exact order of the logarithm for $\varrho^{\text{lin}}_n(S^r_pW(\mathbb{T}^d), L_q(\mathbb{T}^d))$.

Chapter 10

Appendix

A Quasi-Banach spaces

Definition A.1 (Quasi norm)**.** *Let X be a vector space. We call the mapping $\|\cdot\| : X \to \mathbb{R}$ quasi norm if and only if*

(i) $\|x\| = 0 \Longrightarrow x = 0$

(ii) $\|\lambda x\| = |\lambda|\|x\|$

(iii) $\exists C > 0 \forall x, y \in X :$

$$\|x + y\| \leq C(\|x\| + \|y\|).$$

Definition A.2 (p-norm)**.** *Let X be a vector space. We call the mapping $\|\cdot\| : X \to \mathbb{R}$ a p-norm if and only if*

(i) $\|x\| = 0 \Longrightarrow x = 0$

(ii) $\|\lambda x\| = |\lambda|\|x\|$

(iii) $\exists 0 < p \leq 1 \forall x, y \in X :$

$$\|x + y\|^p \leq \|x\|^p + \|y\|^p.$$

Definition A.3. *The tuple $(X, \|\cdot\|)$ where X is a vector space with either $\|\cdot\|$ is a quasi or p-norm is called quasi-Banach (p-Banach) space if and only if every Cauchy sequence $(x_j) \subset X$ converges (in the sense of $\|\cdot\|$) to an element $x \in X$.*

Theorem A.4 (Aoki-Rolewicz)**.** *For every quasi norm $\|\cdot\|$ exists an equivalent p-norm and the other way around.*

Proof. This result is due to Aoki [2] and Rolewicz [93]. □

B Basics from Fourier analysis

Fourier analysis on $\mathbb{R}^d$

Definition B.5. *Let $f \in L_1^{loc}(\mathbb{R}^d)$. Then we define the Hardy-Littlewood maximal operator as*

$$Mf(\boldsymbol{x}) := \sup_{Q \ni \boldsymbol{x}} \frac{1}{|Q|} \int_Q |f(\boldsymbol{x})| d\boldsymbol{x}$$

where the Q are axis parallel squares that are centered in $\boldsymbol{x}$.

Theorem B.6 (Fefferman-Stein maximal inequality)**.** *Let $1 < p < \infty$, $1 < \theta \le \infty$ and $(f_{\boldsymbol{k}})_{\boldsymbol{k}} \subset L_p(\ell_\theta, \mathbb{R}^d)$. Then we have*

$$\|Mf_{\boldsymbol{k}}|L_p(\ell_\theta, \mathbb{R}^d)\| \lesssim \|f_{\boldsymbol{k}}|L_p(\ell_\theta, \mathbb{R}^d)\|.$$

Proof. We refer to [39, Theorem 1]. □

As a main tool we introduce the following componentwise variant of the Hardy-Littlewood maximal operator, see [127, (1.14),(1.15)], [122, (10)]

Definition B.7. *Let $i \in [d]$ and $f \in L_1^{loc}(\mathbb{R}^d)$ then we define the Hardy-Littlewood maximal operator in the i-th direction as*

$$M_i f(\boldsymbol{x}) := \sup_{t>0} \frac{1}{2t} \int_{-t}^{t} |f(x_1, \dots, x_{i-1}, x_i + y, x_{i+1}, \dots, x_d)| dy. \tag{B.1}$$

Theorem B.8. *Let $1 < p, q < \infty$ and $(f_{\boldsymbol{k}})_{\boldsymbol{k}} \subset L_p(\mathbb{T}^d, \ell_\theta)$ and $i \in [d]$. Then we have*

$$\|M_i f_{\boldsymbol{k}}|L_p(\ell_\theta, \mathbb{R}^d)\| \lesssim \|f_{\boldsymbol{k}}|L_p(\ell_\theta, \mathbb{R}^d)\|.$$

Definition B.9 (Peetre maximal operator)**.** *Let $a > 0$ and $\boldsymbol{b} > 0$ then we define for $f \in C(\mathbb{T}^d)$*

$$P_{\boldsymbol{b},a} f(\boldsymbol{x}) := \sup_{\boldsymbol{y} \in \mathbb{R}^d} \frac{|f(\boldsymbol{x} + \boldsymbol{y})|}{(1 + b_1|y_1|)^a \dots (1 + b_d|y_d|)^a}.$$

Additionally we define for $e \subset [d]$ a component wise Peetre maximal operator by

$$P_{\boldsymbol{b},a|e} f(\boldsymbol{x}) := \sup_{\boldsymbol{y} \in \mathbb{R}^d(e)} \frac{|f(\boldsymbol{x} + \boldsymbol{y})|}{\prod_{i \in e}(1 + b_i|y_i|)^a}.$$

Lemma B.10. *Let $a, b > 0$ and $f \in L_1(\mathbb{R})$ with $\operatorname{supp} \mathcal{F} f \subset [-b, b]$. Then there exists a constant $C > 0$ such that*

$$|\Delta_h^m f(x)| \le C \min\{1, |bh|^m\} \max\{1, |bh|^a\} P_{b,a} f(x) \tag{B.2}$$

holds.

Proof. We refer to [122, Lemma 3.3.4]. □

Theorem B.11. *Let $0 < p < \infty$, $0 < \theta \le \infty$ and $(f_{\boldsymbol{j}})_{\boldsymbol{j}\in\mathbb{N}_0^d}$ be a sequence of bandlimited functions with*

$$\operatorname{supp} f_j \subset [-b, b]$$

and $a > \max\{\frac{1}{p}, \frac{1}{\theta}\}$. Then there is a constant $C > 0$ (independent of f and $\boldsymbol{b_j}$) such that

$$\|P_{\boldsymbol{b_j},a} f_{\boldsymbol{j}} | L_p(\ell_\theta, \mathbb{R}^d)\| \le C \|f_{\boldsymbol{j}} | L_p(\ell_\theta, \mathbb{R}^d)\|$$

holds.

Proof. We refer to [99, 1.6.4] and Theorem B.14. □

Lemma B.12. *Let $a > 0$, $b > 0$ and $f \in C(\mathbb{R})$.*

(i) If $|x - x_0| < \frac{1}{b}$ then

$$|f(x_0)| \le 2^a P_{b,a} f(x)$$

holds.

(ii) Furthermore let $b' > b > 0$. Then

$$P_{b,a} f(x) \le \left(\frac{b'}{b}\right)^a P_{b',a} f(x).$$

Proof. The following estimation yields (i)

$$|f(x_0)| \le \frac{|f(x_0)|}{(1 + |x - x_0|b)^a}(1 + |x - x_0|b)^a \le 2^a \sup_{x_0\in\mathbb{R}} \frac{|f(x_0)|}{(1 + |x - x_0|b)^a} = 2^a P_{b,a} f(x).$$

We prove (ii). The trivial estimation $\frac{p+1}{q+1} \le \frac{p}{q}$ for $p > q > 0$ yields

$$P_{b,a} f(x) = \sup_{y\in\mathbb{R}} \frac{|f(y)|}{(1 + b|x + y|)^a} \le \sup_{y\in\mathbb{R}} \frac{|f(y)|}{(1 + b'|x + y|)^a} \frac{(1 + b'|x + y|)^a}{(1 + b|x + y|)^a} \le \left(\frac{b'}{b}\right)^a P_{b',a} f(x).$$

□

Lemma B.13. *Let $0 < a \le 1$ and $R > \frac{1}{a}$. For any sequence $(\lambda_{\boldsymbol{j},\boldsymbol{k}})_{\boldsymbol{j}\in\mathbb{N}_0^d,\boldsymbol{k}\in\mathbb{Z}^d}$ of complex numbers and any $\boldsymbol{\ell} \in \mathbb{Z}^d$, $\boldsymbol{j} \in \mathbb{N}_0^d$ with $\boldsymbol{j} + \boldsymbol{\ell} \ge -1$ we have*

$$\sum_{\boldsymbol{k}\in\mathbb{Z}^d} |\lambda_{j+\ell,k}| \prod_{i=1}^d (1 + 2^{\min\{j_i, j_i+\ell_i\}} |x_i - x_{j_i+\ell_i,k_i}|)^{-R} \lesssim 2^{|\boldsymbol{\ell}_+|_1/a} \Big[M \Big| \sum_{\boldsymbol{k}\in\mathbb{Z}^d} \lambda_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}} \chi_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}} \Big|^a \Big]^{\frac{1}{a}} (x). \tag{B.3}$$

Here M denotes the Hardy-Littlewood maximal operator, cf. Definition B.5.

Proof. The proof is taken from [126, Lemma 4.3]. Which is a “hyperbolic” version of [65, Lem. 3, 7]. The lemma is originally due to Kyriazis [72, Lem. 7.1]. Let

$$\delta = R - \frac{1}{a}.$$

We introduce for $\mu \in \mathbb{N}$ the sets

$$\Omega_\mu(x) := \{k \in \mathbb{Z} : 2^{\mu-1} < 2^{\min\{j+\ell,j\}}|x - x_{j+\ell,k}| < 2^\mu\}$$

and

$$\Omega_0(x) := \{k \in \mathbb{Z} : 2^{\min\{j+\ell,j\}}|x - x_{j+\ell,k}| < 1\}.$$

For $\boldsymbol{\mu} \in \mathbb{N}_0^d$ and $\boldsymbol{x} \in \mathbb{R}^d$ we define

$$\Omega_{\boldsymbol{\mu}}(\boldsymbol{x}) = \Omega_{\mu_1}(x_1) \cdot \ldots \cdot = \Omega_{\mu_d}(x_d).$$

We start estimating

$$\begin{aligned} \sum_{\boldsymbol{k}\in\mathbb{Z}^d} |\lambda_{j+\ell,k}| \prod_{i=1}^d (1 + 2^{\min\{j_i,j_i+\ell_i\}}|x_i - x_{j_i+\ell_i,k_i}|)^{-R} &\lesssim \sum_{\mu\in\mathbb{N}_0^d} \sum_{\boldsymbol{k}\in\Omega_\mu(\boldsymbol{x})} |\lambda_{j+\ell,k}| 2^{-|\boldsymbol{\mu}|_1(\delta+\frac{1}{a})} \\ . &\lesssim \sup_{\boldsymbol{\mu}\in\mathbb{N}_0^d} \sum_{\boldsymbol{k}\in\Omega_\mu(\boldsymbol{x})} |\lambda_{j+\ell,k}| 2^{-\frac{|\boldsymbol{\mu}|_1}{a}} \\ &\lesssim \Big(\sup_{\boldsymbol{\mu}\in\mathbb{N}_0^d} \sum_{\boldsymbol{k}\in\Omega_\mu(\boldsymbol{x})} |\lambda_{j+\ell,k}|^a 2^{-|\boldsymbol{\mu}|_1} \Big)^{\frac{1}{a}} \end{aligned} \tag{B.4}$$

It is easy to verify that

$$\int_{\bigcup_{m\in\Omega_{\boldsymbol{\mu}}(\boldsymbol{x})} I_{\boldsymbol{j}+\boldsymbol{\ell},m}} \sum_{\boldsymbol{k}\in\Omega_\mu(\boldsymbol{x})} |\lambda_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}}|^a \chi_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}}(\boldsymbol{y}) d\boldsymbol{y} \asymp 2^{-|\boldsymbol{j}+\boldsymbol{\ell}|_1} \sum_{\boldsymbol{k}\in\Omega_\mu(\boldsymbol{x})} |\lambda_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}}|^a$$

holds. Inserting this into (B.4) gives

$$\sum_{\boldsymbol{k}\in\mathbb{Z}^d} |\lambda_{j+\ell,k}| \prod_{i=1}^d (1 + 2^{\min\{j_i,j_i+\ell_i\}}|x_i - x_{j_i+\ell_i,k_i}|)^{-R} \tag{B.5}$$

$$\lesssim \Big(2^{|\boldsymbol{j}+\boldsymbol{\ell}|_1} \sup_{\boldsymbol{\mu}\in\mathbb{N}_0^d} 2^{-|\boldsymbol{\mu}|_1} \int_{\bigcup_{m\in\Omega_{\boldsymbol{\mu}}(\boldsymbol{x})} I_{\boldsymbol{j}+\boldsymbol{\ell},m}} \sum_{\boldsymbol{k}\in\Omega_\mu(\boldsymbol{x})} |\lambda_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}}|^a \chi_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}}(\boldsymbol{y}) d\boldsymbol{y}\Big)^{\frac{1}{a}}$$

$$. \tag{B.6}$$

Defining

$$Q(x) = \bigcup_{m\in\Omega_{\boldsymbol{\mu}}(\boldsymbol{x})} I_{\boldsymbol{j}+\boldsymbol{\ell},m}$$

we observe

$$|Q(x)| \asymp 2^{|\boldsymbol{\mu}|_1} 2^{-|\boldsymbol{j}+\boldsymbol{\ell}_-|_1}.$$

Hence, inserting this into (B.6) allows to estimate

$$
\begin{aligned}
&\sum_{\boldsymbol{k}\in\mathbb{Z}^d} |\lambda_{j+\ell,k}| \prod_{i=1}^{d} (1+2^{\min\{j_i,j_i+\ell_i\}}|x_i - x_{j_i+\ell_i,k_i}|)^{-R} \\
\lesssim\ &\Big(2^{|\boldsymbol{\ell}_+|_1} \sup_{\boldsymbol{\mu}\in\mathbb{N}_0^d} \frac{1}{|Q(\boldsymbol{x})|} \int_{Q(\boldsymbol{x})} \sum_{\boldsymbol{k}\in\Omega_\mu(\boldsymbol{x})} |\lambda_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}}|^a \chi_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}}(\boldsymbol{y}) d\boldsymbol{y}\Big)^{\frac{1}{a}} \\
&\le 2^{\frac{|\boldsymbol{\ell}_+|_1}{a}} \Big(M\Big|\sum_{\boldsymbol{k}\in\Omega_\mu(\boldsymbol{x})} \lambda_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}}\chi_{\boldsymbol{j}+\boldsymbol{\ell},\boldsymbol{k}}\Big|^a\Big)^{\frac{1}{a}}
\end{aligned}
$$

□

Fourier analysis on $\mathbb{T}^d$

The Hardy-Littlewood maximal function and Peetre maximal function are defined as in the last subsection, by interpreting $f \in L_1(\mathbb{T}^d)$ as a 2π-periodic function on $\mathbb{R}^d$. There is a corresponding variant of the Fefferman-Stein theorem, see [122, Thm. 4.1.2] and the references therein.

Theorem B.14 (Fefferman-Stein maximal inequality)**.** *Let* $1 < p < \infty$, $1 < \theta \le \infty$ *and* $(f_{\boldsymbol{k}})_{\boldsymbol{k}} \subset L_p(\ell_\theta, \mathbb{T}^d)$. *Then we have*

$$\|Mf_{\boldsymbol{k}}|L_p(\ell_\theta\mathbb{T}^d)\| \lesssim \|f_{\boldsymbol{k}}|L_p(\ell_\theta, \mathbb{T}^d)\|.$$

Proof. We refer to [99, Proposition 3.2.4]. □

Similarly to the non-periodic case we have:

Theorem B.15. *Let* $1 < p, q < \infty$ *and* $(f_{\boldsymbol{k}})_{\boldsymbol{k}} \subset L_p(\mathbb{T}^d, \ell_\theta)$ *and* $i \in [d]$. *Then we have*

$$\|M_i f_{\boldsymbol{k}}|L_p(\ell_\theta, \mathbb{T}^d)\| \lesssim \|f_{\boldsymbol{k}}|L_p(\ell_\theta, \mathbb{T}^d)\|.$$

Theorem B.16. *Let* $0 < p \le \infty$ *and* f *be a trigonometric polynomial with*

$$f = \sum_{\substack{|k_i|\le b_i \\ i=1,\dots,d}} \hat{f}(\boldsymbol{k}) e^{i\boldsymbol{k}\cdot\boldsymbol{x}}$$

and $a > \frac{1}{p}$. *Then there is a constant* $C > 0$ *(independent of* f *and* $\boldsymbol{b}$*) such that*

$$\|P_{\boldsymbol{b},a} f|L_p(\mathbb{T}^d)\| \le C\|f|L_p(\mathbb{T}^d)\|$$

holds.

Proof. We refer to [99, 3.3.5] and Theorem B.15. □

Theorem B.17. *Let* $0 < p < \infty$, $0 < \theta \leq \infty$ *and* $(f_{\boldsymbol{j}})_{\boldsymbol{j} \in \mathbb{N}_0^d}$ *be a sequence of trigonometric polynomials with*

$$f_{\boldsymbol{j}} = \sum_{\substack{|k_i| \leq b_i^{\boldsymbol{j}} \\ i=1,\dots,d}} \hat{f}(\boldsymbol{k}) e^{i\boldsymbol{k}\cdot\boldsymbol{x}}$$

and $a > \max\{\frac{1}{p}, \frac{1}{\theta}\}$. *Then there is a constant* $C > 0$ *(independent of* f *and* $\boldsymbol{b_j}$*) such that*

$$\|P_{\boldsymbol{b_j},a} f_{\boldsymbol{j}} | L_p(\ell_\theta, \mathbb{T}^d)\| \leq C \|f_{\boldsymbol{j}} | L_p(\ell_\theta, \mathbb{T}^d)\|$$

holds.

Proof. We refer to [122, Thm. 4.1.3]. □

The next result is well known in harmonic analysis. We state it for completeness.

Lemma B.18. *Let* $f \in L_1(\mathbb{T})$ *with* $\sum_{\ell \in \mathbb{Z}} |\widehat{f}(\ell)| < \infty$. *Then*

$$f(\cdot) = \sum_{\ell \in \mathbb{Z}} \widehat{f}(\ell) e^{i\ell\cdot}$$

in $C(\mathbb{T})$.

Lemma B.19 (Poisson summation). *Let* $f \in L_1(\mathbb{R})$. *Then its periodization* $\sum_{k \in \mathbb{Z}} f(\cdot + 2\pi k)$ *converges absolutely in the norm of* $L_1([-\pi, \pi])$. *Furthermore its formal Fourier series is given by*

$$\sum_{k \in \mathbb{Z}} f(\cdot + 2\pi k) = \frac{1}{\sqrt{2\pi}} \sum_{\ell \in \mathbb{Z}} \mathcal{F}f(\ell) e^{i\ell\cdot}$$

Proof. We refer to [111, p. 252]. □

C Some multi-indexed geometric sums

Lemma C.20. *Let* $\boldsymbol{r}, \boldsymbol{\eta} \in \mathbb{R}^d$ *with* $0 < r_1 = \eta_1 = \ldots = r_\mu = \eta_\mu < r_{\mu+1} \leq \ldots \leq r_d$ *and* $r_1 < \eta_s < r_s$ *for* $s = \mu+1, \ldots, d$. *Then*

$$\sum_{\frac{1}{\eta_1}\boldsymbol{\eta}\cdot\boldsymbol{j} > m} 2^{-\boldsymbol{r}\cdot\boldsymbol{j}} \lesssim m^{\mu-1} 2^{-r_1 m}$$

holds for all $m \geq 1$.

Proof. We refer to [112, p. 9, Lemma B]. □

Lemma C.21. *Let* $\boldsymbol{r} \in \mathbb{R}^d$ *with*

$$0 < r_1 = \ldots = r_\mu < r_{\mu+1} \leq \ldots \leq r_d < \infty$$

and $\mu \leq d$. *Then*

$$\sum_{\frac{1}{r_1}\boldsymbol{r}\cdot\boldsymbol{j} \leq m} 2^{|\boldsymbol{j}|_1} \asymp m^{\mu-1} 2^m$$

holds for all $m \geq 1$.

Proof. We refer to [112, p. 10, Lemma D]. □

Lemma C.22. *Let* $\alpha > 0$, $\beta \geq 0$, *such that* $\alpha > \beta$. *Then*

$$\sum_{\boldsymbol{j} \in \Delta_{\alpha,\beta}(M)} 2^{|\boldsymbol{j}|_1} \asymp 2^{\frac{\xi}{\alpha-\beta}}$$

holds for all $M \geq \alpha - \beta$.

Proof. We refer to [9, Lemma 6.3]. □

Lemma C.23. *Let* $0 < \varepsilon < \gamma < \alpha$ *then*

$$\sum_{\boldsymbol{j} \notin \Delta_{\alpha-\varepsilon,\beta-\varepsilon}(M)} 2^{-2(\alpha|\boldsymbol{j}|_1 + \gamma|\boldsymbol{j}|_\infty)} \lesssim 2^{-M}$$

holds for all $M \in \mathbb{N}$.

Proof. We refer to [9, Theorem 4.1, 2nd step]. □

D Known results on linear and Kolmogorov-widths

Theorem D.24. *Let* $1 < p < \infty$, $1 \leq q < \infty$ *and* $\boldsymbol{r} > (1/p - 1/q)_+$ *with* (8.1.1). *Then we have*

$$\lambda_n(S^{\boldsymbol{r}}_p W(\mathbb{T}^d), L_q(\mathbb{T}^d)) \asymp \begin{cases} \left(\frac{(\log n)^{\mu-1}}{n}\right)^{r_1 - (1/p-1/q)_+} & : \ q \leq 2, \ \ or; p \geq 2, \\ \left(\frac{(\log n)^{\mu-1}}{n}\right)^{r_1 - 1/p + 1/2} & : \ 1/p + 1/q \geq 1, \ q > 2, \ \boldsymbol{r} > 1/p, \\ \left(\frac{(\log n)^{\mu-1}}{n}\right)^{r_1 - 1/2 + 1/q} & : \ 1/p + 1/q \leq 1, \ p < 2, \ \boldsymbol{r} > 1 - 1/q. \end{cases}$$

Proof. The case $1 < q < \infty$ was proven by Galeev[41, 42], see also [36, 37]. The case $q = 1$ by Romanyuk [96]. Additionally we refer to [33, Theorem 4.39] and the comments therein. □

Theorem D.25. *Let* $\boldsymbol{r}$ *as in* (8.3.2). *Let additionally* $1 < p < q \leq 2$ *and* $1 \leq \theta \leq \infty$ *or* $2 \leq p < q < \infty$ *and* $\theta \geq 2$. *Then we have*

$$\lambda_n(S^{\boldsymbol{r}}_{p,\theta} F(\mathbb{T}^d), L_q(\mathbb{T}^d)) \asymp \left(\frac{(\log n)^{\mu-1}}{n}\right)^{r_1 - 1/p - 1/q},$$

for all $n \in \mathbb{N}$.

Proof. The upper bound can be obtained for instance by sampling recovery, cf. Theorem 8.7. We focus on lower bounds. In case $\theta \geq 2$ the embedding $S^r_p W(\mathbb{R}^d) \hookrightarrow S^r_{p,\theta} F(\mathbb{R}^d)$ yields

$$\lambda_n(S^{\boldsymbol{r}}_{p,\theta} F(\mathbb{T}^d), L_q(\mathbb{T}^d)) \geq \lambda_n(S^{\boldsymbol{r}}_p W(\mathbb{T}^d), L_q(\mathbb{T}^d))$$

The results stated in Theorem D.24 provide the correct order. In case $\theta < p$ the embedding $S^r_{p,\theta}B(\mathbb{R}^d) \hookrightarrow S^r_{p,\theta}F(\mathbb{R}^d)$ yields

$$\lambda_n(S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d), L_q(\mathbb{T}^d)) \geq \lambda_n(S^{\boldsymbol{r}}_{p,\theta}B(\mathbb{T}^d), L_q(\mathbb{T}^d)).$$

This gives the right order in cases $1 < p < q \leq 2$ and $2 \leq q < p$, cf. [94]. Finally for $\theta \geq p$ we stress on the embedding

$$S^{\boldsymbol{r}}_{p,p}B(\mathbb{T}^d) \hookrightarrow S^r_{p,\theta}F(\mathbb{T}^d)$$

with

$$\lambda_n(S^{\boldsymbol{r}}_{p,\theta}F(\mathbb{T}^d), L_q(\mathbb{T}^d)) \geq \lambda_n(S^{\boldsymbol{r}}_{p,p}B(\mathbb{T}^d), L_q(\mathbb{T}^d)).$$

This provides the lower bound in case $1 < p < q \leq 2$. We refer again to [94]. □

The following is known for Kolmogorov widths in case of Sobolev spaces $S^{\boldsymbol{r}}_pW(\mathbb{R}^d)$ defined by

$$d_n(S^{\boldsymbol{r}}_pW(\mathbb{T}^d), L_q(\mathbb{T}^d)) = \inf_{\substack{A\subset L_q(\mathbb{T}^d)\\ \dim A\leq n}} \sup_{\|f|S^{\boldsymbol{r}}_pW(\mathbb{T}^d)\|\leq 1} \inf_{g\in A} \|f-g|L_q(\mathbb{T}^d)\|. \tag{D.1}$$

Theorem D.26. *Let* $1 < p, q < \infty$ *and*

$$\boldsymbol{r} > \begin{cases} (\frac{1}{p}-\frac{1}{q})_+ & : \quad 1 \leq p \leq q \leq 2 \ \text{or}\ 1 \leq q \leq p < \infty, \\ \max\{\frac{1}{2},\frac{1}{p}\} & : \quad \text{otherwise}, \end{cases}$$

as in (8.3.2). *Then*

$$d_n(S^{\boldsymbol{r}}_pW(\mathbb{T}^d), L_q(\mathbb{T}^d)) \asymp (m^{-1}\log^{\mu-1} m)^{r_1-(\frac{1}{p}-\max\{\frac{1}{2},\frac{1}{q}\})_+}.$$

Proof. The proof with every single case has a history of more than 20 years. For an overview we refer to [33, Section 4.3]. □

Theorem D.27. *Let* $1 < p \leq 2$ *and* $\boldsymbol{r} > 1$ *satisfying* (8.3.2). *Then*

$$\lambda_n(S^{\boldsymbol{r}}_pW(\mathbb{T}^d), L_\infty(\mathbb{T}^d)) \asymp n^{-(r_1-\frac{1}{2})}\log^{(\mu-1)r_1} n.$$

Proof. We refer to [78, Theorem 2.14] and the references therein. □

Bibliography

[1] M. Aigner and G. M. Ziegler. *Proofs from The Book*. Springer-Verlag, Berlin, second edition, 2001.

[2] T. Aoki. Locally bounded linear topological spaces. *Proc. Imp. Acad.*, 18(10):588–594, 1942.

[3] S. A. Balgimbayeva. Nonlinear approximation of function spaces of mixed smoothness. *Siberian Mathematical Journal*, 56(2):262–274, Mar 2015.

[4] S. A. Balgimbayeva and T. I. Smirnov. Estimates of the fourier widths of the classes of periodic functions with given majorant of the mixed modulus of smoothness. *Siberian Mathematical Journal*, 59(2):217–230, Mar 2018.

[5] D. B. Bazarkhanov. Nonlinear approximations of classes of periodic functions of many variables. *Proc. Steklov Inst. Math.*, 284(1):2–31, 2014.

[6] D. B. Bazarkhanov. Nonlinear trigonometric approximations of multivariate function classes. *Tr. Mat. Inst. Steklova*, 293:8–42, 2016.

[7] H.-J. Bungartz. Finite elements of higher order on sparse grids. *Habilitationsschrift, Institut für Informatik, TU München*, 1998.

[8] H.-J. Bungartz and M. Griebel. Sparse grids. *Acta Numerica*, 13:147–269, 2004.

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

[10] G. Byrenheid, L. Kämmerer, T. Ullrich, and T. Volkmer. Tight error bounds for rank-1 lattice sampling in spaces of hybrid mixed smoothness. *Num. Math.*, 136:993–1034, 2017.

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

[12] J. Céa. Approximation variationnelle des problèmes aux limites. *Ann. Inst. Fourier (Grenoble)*, 14(fasc. 2):345–444, 1964.

[13] A. Chkifa, A. Cohen, and C. Schwab. High-dimensional adaptive sparse polynomial interpolation and applications to parametric PDEs. *Found. Comput. Math.*, 14(4):601–633, 2014.

[14] M. Christ and A. Seeger. Necessary conditions for vector-valued operator inequalities in harmonic analysis. *Proc. London Math. Soc. (3)*, 93(2):447–473, 2006.

[15] F. Cobos, T. Kühn, and W. Sickel. Optimal approximation of multivariate periodic Sobolev functions in the sup-norm. *J. Funct. Anal.*, 270(11):4196–4212, 2016.

[16] C. de Boor. *A practical guide to splines*, volume 27 of *Applied Mathematical Sciences*. Springer-Verlag, New York-Berlin, 1978.

[17] G. Deslauriers and S. Dubuc. Symmetric iterative interpolation processes. *Constr. Approx.*, 5(1):49–68, 1989. Fractal approximation.

[18] R. A. DeVore. Nonlinear approximation. In *Acta numerica, 1998*, volume 7 of *Acta Numer.*, pages 51–150. Cambridge Univ. Press, Cambridge, 1998.

[19] R. A. DeVore, S. V. Konyagin, and V. N. Temlyakov. Hyperbolic wavelet approximation. *Constr. Approx.*, 14(1):1–26, 1998.

[20] R. A. DeVore and V. A. Popov. Interpolation of approximation spaces. In *Constructive theory of functions (Varna, 1987)*, pages 110–119. Publ. House Bulgar. Acad. Sci., Sofia, 1988.

[21] S. Dirksen and T. Ullrich. Gelfand numbers related to structured sparsity and besov space embeddings with small mixed smoothness. *Journal of Complexity*, 48:69 – 102, 2018.

[22] D. L. Donoho. Interpolating wavelet transform. Technical report, Department of Statistics, Stanford University, 1992.

[23] D. L. Donoho and T. P. Y. Yu. Deslauriers-Dubuc: ten years after. In *Spline functions and the theory of wavelets (Montreal, PQ, 1996)*, volume 18 of *CRM Proc. Lecture Notes*, pages 355–370. Amer. Math. Soc., Providence, RI, 1999.

[24] S. Dubuc. Interpolation through an iterative scheme. *J. Math. Anal. Appl.*, 114(1):185–204, 1986.

[25] D. Dũng. On optimal recovery of multivariate periodic functions. In *Harmonic analysis (Sendai, 1990)*, ICM-90 Satell. Conf. Proc., pages 96–105. Springer, Tokyo, 1991.

[26] D. Dũng. Optimal recovery of functions of a certain mixed smoothness. *Tạp chí Toán Học J. Math.*, 20(2):18–32, 1992.

[27] D. Dũng. Continuous algorithms in n-term approximation and non-linear widths. *J. Approx. Theory*, 102(2):217–242, 2000.

[28] D. Dũng. Asymptotic orders of optimal non-linear approximations. *East J. Approx.*, 7(1):55–76, 2001.

[29] D. Dũng. B-spline quasi-interpolant representations and sampling recovery of functions with mixed smoothness. *J. Complexity*, 27(6):541–567, 2011.

[30] D. Dũng. Sampling and cubature on sparse grids based on a B-spline quasi-interpolation. *Found. Comput. Math.*, 16(5):1193–1240, 2016.

[31] D. Dũng. B-spline quasi-interpolation sampling representation and sampling recovery in Sobolev spaces of mixed smoothness. *Acta Math. Vietnam.*, 43(1):83–110, 2018.

[32] D. Dũng and M. Griebel. Hyperbolic cross approximation in infinite dimensions. *Journal of Complexity*, 33:55–88, 2016. Also available as INS Preprint No. 1501.

[33] D. Dung, V. N. Temlyakov, and T. Ullrich. Hyperbolic cross approximation. *Advanced Courses in Mathematics. CRM Barcelona. Birkhäuser/Springer*, to appear.

[34] D. Dũng and T. Ullrich. N-widths and ε-dimensions for high-dimensional approximations. *Found. Comput. Math.*, 13:965–1003, 2013.

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

[36] D. E. Edmunds and H. Triebel. Entropy numbers and approximation numbers in function spaces. *Proc. London Math. Soc. (3)*, 58(1):137–152, 1989.

[37] D. E. Edmunds and H. Triebel. Entropy numbers and approximation numbers in function spaces. II. *Proc. London Math. Soc. (3)*, 64(1):153–169, 1992.

[38] G. Faber. Über stetige Funktionen. *Mathematische Annalen*, 66(1):81–94, 1908.

[39] C. Fefferman and E. M. Stein. Some maximal inequalities. *Amer. J. Math.*, 93:107–115, 1971.

[40] C. Feuersänger and M. Griebel. Principal manifold learning by sparse grids. *Computing*, 85(4):267–299, 2009.

[41] È. M. Galeev. On linear widths of classes of periodic functions of several variables. *Vestnik MGU, Ser.1 Mat.-Mekh.*, 4:13–16, 1987.

[42] È. M. Galeev. Linear widths of Hölder-Nikolskij classes of periodic functions of several variables. *Mat. Zametki*, 59(2):189–199, 317, 1996.

[43] J. Garcke and D. Pflüger, editors. *Sparse grids and applications—Stuttgart 2014*, volume 109 of *Lecture Notes in Computational Science and Engineering*. Springer, Cham, 2016. Selected papers from the Third Workshop (SGA 2014) held at University of Stuttgart, Stuttgart, September 1–5, 2014.

[44] T. Gerstner. *Adaptive hierarchical methods for landscape representation and analysis*, pages 75–92. Springer Berlin Heidelberg, Berlin, Heidelberg, 1999.

[45] T. Gerstner. Multiresolution compression and visualization of global topographic data. *GeoInformatica*, 7(1):7–32, 2003.

[46] E. D. Gluskin. Norms of random matrices and diameters of finite-dimensional sets. *Mat. Sb. (N.S.)*, 120(162)(2):180–189, 286, 1983.

[47] M. Griebel. Sparse grids and related approximation schemes for higher dimensional problems. In *Foundations of computational mathematics, Santander 2005*, volume 331 of *London Math. Soc. Lecture Note Ser.*, pages 106–161. Cambridge Univ. Press, Cambridge, 2006.

[48] M. Griebel and J. Hamaekers. Fast discrete Fourier transform on generalized sparse grids. In *Sparse grids and Applications*, volume 97 of *Lecture Notes in Computational Science and Engineering*, pages 75–108. Springer, 2014. INS Preprint No. 1305.

[49] M. Griebel and H. Harbrecht. A note on the construction of L-fold sparse tensor product spaces. *Constructive Approximation*, 38(2):235–251, 2013. Also available as INS Preprint No. 1205.

[50] M. Griebel and H. Harbrecht. On the construction of sparse tensor product spaces. *Mathematics of Computations*, 82(282):975–994, Apr. 2013. Also available as INS Preprint No. 1104, 2011.

[51] M. Griebel and M. Hegland. A finite element method for density estimation with Gaussian process priors. *SIAM J. Numer. Anal.*, 47(6):4759–4792, 2010.

[52] M. Griebel, S. Knapek, and G. Zumbusch. *Numerical simulation in molecular dynamics*, volume 5 of *Texts in Computational Science and Engineering*. Springer, Berlin, 2007. Numerics, algorithms, parallelization, applications.

[53] M. Hansen. *Nonlinear approximation and function spaces of dominating mixed smoothness*. Dissertation. Universität Jena, 2010.

[54] M. Hansen and W. Sickel. Best m-term approximation and tensor product of Sobolev and Besov spaces—the case of non-compact embeddings. *East J. Approx.*, 16(4):345–388, 2010.

[55] M. Hansen and W. Sickel. Best m-term approximation and Lizorkin-Triebel spaces. *J. Approx. Theory*, 163(8):923–954, 2011.

[56] M. Hansen and W. Sickel. Best m-term approximation and Sobolev-Besov spaces of dominating mixed smoothness—the case of compact embeddings. *Constr. Approx.*, 36(1):1–51, 2012.

[57] M. Hansen and J. Vybíral. The Jawerth-Franke embedding of spaces with dominating mixed smoothness. *Georgian Math. J.*, 16(4):667–682, 2009.

[58] T. Hastie, R. Tibshirani, and J. Friedman. *The elements of statistical learning.* Springer Series in Statistics. Springer, New York, second edition, 2009. Data mining, inference, and prediction.

[59] S. Heinrich. Randomized approximation of Sobolev embeddings. II. *J. Complexity*, 25(5):455–472, 2009.

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

[61] A. Hinrichs, E. Novak, M. Ullrich, and H. Woźniakowski. The curse of dimensionality for numerical integration of smooth functions ii. *Journal of Complexity*, 30(2):117 – 143, 2014. Dagstuhl 2012.

[62] E. Hlawka. Über die Diskrepanz mehrdimensionaler Folgen. *Math. Z.*, 77:273–284, 1961.

[63] B. S. Kashin. The widths of certain finite-dimensional sets and classes of smooth functions. *Izv. Akad. Nauk SSSR Ser. Mat.*, 41(2):334–351, 478, 1977.

[64] B. S. Kashin. Some properties of matrices of bounded operators from the space l_2^n into l_2^m. *Izv. Akad. Nauk Armyan. SSR Ser. Mat.*, 15(5):379–394, 417, 1980.

[65] H. Kempka. Atomic, molecular and wavelet decomposition of generalized 2-microlocal Besov spaces. *J. Funct. Spaces Appl.*, 8(2):129–165, 2010.

[66] V. Kien Nguyen, M. Ullrich, and T. Ullrich. Change of variable in spaces of mixed smoothness and numerical integration of multivariate functions on the unit cube. *ArXiv e-prints*, Nov. 2015.

[67] S. Knapek. *Approximation und Kompression mit Tensorprodukt-Multiskalenräumen.* Dissertation, Universität Bonn, April 2000.

[68] D. Krieg. Optimal Monte Carlo Methods for L^2-Approximation. *ArXiv e-prints*, May 2017.

[69] T. Kühn, S. Mayer, and T. Ullrich. Counting via entropy: new preasymptotics for the approximation numbers of Sobolev embeddings. *SIAM J. Numer. Anal.*, 54(6):3625–3647, 2016.

[70] T. Kühn, W. Sickel, and T. Ullrich. Approximation numbers of Sobolev embeddings—sharp constants and tractability. *J. Complexity*, 30(2):95–116, 2014.

[71] T. Kühn, W. Sickel, and T. Ullrich. Approximation of mixed order Sobolev functions on the d-torus: asymptotics, preasymptotics, and d-dependence. *Constr. Approx.*, 42(3):353–398, 2015.

[72] G. Kyriazis. Decomposition systems for function spaces. *Studia Math.*, 157(2):133–169, 2003.

[73] P. D. Lax and A. N. Milgram. Parabolic equations. In *Contributions to the theory of partial differential equations*, Annals of Mathematics Studies, no. 33, pages 167–190. Princeton University Press, Princeton, N. J., 1954.

[74] P. I. Lizorkin. Multipliers of Fourier integrals in the spaces $L_{p,\theta}$. *Trudy Mat. Inst. Steklov*, 89:231–248, 1967.

[75] V. E. Maĭorov. Discretization of the problem of diameters. *Uspehi Mat. Nauk*, 30(6(186)):179–180, 1975.

[76] Y. V. Malykhin and K. S. Ryutin. The product of octahedra is badly approximated in the $\ell_{2,1}$-metric. *Mat. Zametki*, 101(1):85–90, 2017.

[77] V. K. Nguyen. Bernstein numbers of embeddings of isotropic and dominating mixed Besov spaces. *Math. Nachr.*, 288(14-15):1694–1717, 2015.

[78] V. K. Nguyen. Gelfand numbers of embeddings of mixed Besov spaces. *preprint*, 2016.

[79] V. K. Nguyen. Weyl and Bernstein numbers of embeddings of Sobolev spaces with dominating mixed smoothness. *J. Complexity*, 36:46–73, 2016.

[80] V. K. Nguyen. Gelfand numbers of embeddings of mixed Besov spaces. *J. Complexity*, 41:35–57, 2017.

[81] V. K. Nguyen and W. Sickel. Weyl numbers of embeddings of tensor product Besov spaces. *J. Approx. Theory*, 200:170–220, 2015.

[82] F. Nobile, R. Tempone, and C. G. Webster. A sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM J. Numer. Anal.*, 46(5):2309–2345, 2008.

[83] E. Novak and H. Triebel. Function spaces in Lipschitz domains and optimal rates of convergence for sampling. *Constructive Approximation*, 23(3):325–350, 2005.

[84] E. Novak and H. Woźniakowski. *Tractability of multivariate problems. Vol. 1: Linear information*, volume 6 of *EMS Tracts in Mathematics*. European Mathematical Society (EMS), Zürich, 2008.

[85] E. Novak and H. Woźniakowski. *Tractability of multivariate problems. Volume II: Standard information for functionals*, volume 12 of *EMS Tracts in Mathematics*. European Mathematical Society (EMS), Zürich, 2010.

[86] E. Novak and H. Woźniakowski. *Tractability of multivariate problems. Volume III: Standard information for operators*, volume 18 of *EMS Tracts in Mathematics*. European Mathematical Society (EMS), Zürich, 2012.

[87] P. Oswald. *Multilevel finite element approximation*. Teubner Skripten zur Numerik. [Teubner Scripts on Numerical Mathematics]. B. G. Teubner, Stuttgart, 1994. Theory and applications.

[88] D. Pflüger. Spatially adaptive refinement. In *Sparse grids and applications*, volume 88 of *Lect. Notes Comput. Sci. Eng.*, pages 243–262. Springer, Heidelberg, 2013.

[89] A. Pietsch. Approximation spaces. *J. Approx. Theory*, 32(2):115–134, 1981.

[90] A. Pietsch. *Eigenvalues and S-numbers*. Cambridge University Press, New York, NY, USA, 1986.

[91] A. Pinkus. *N-widths in approximation theory*, volume 7 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*. Springer-Verlag, Berlin, 1985.

[92] S. Reich and C. J. Cotter. Ensemble filter techniques for intermittent data assimilation. In *Large scale inverse problems*, volume 13 of *Radon Ser. Comput. Appl. Math.*, pages 91–134. De Gruyter, Berlin, 2013.

[93] S. Rolewicz. On a certain class of linear metric spaces. *Bull. Acad. Polon. Sci. Cl. III.*, 5:471–473, XL, 1957.

[94] A. S. Romanyuk. Linear widths of Besov classes of periodic functions of several variables. II. *Ukraïn. Mat. Zh.*, 53(6):820–829, 2001.

[95] A. S. Romanyuk. Best M-term trigonometric approximations of Besov classes of periodic functions of several variables. *Izv. Ross. Akad. Nauk Ser. Mat.*, 67(2):61–100, 2003.

[96] A. S. Romanyuk. Best approximations and widths of classes of periodic functions of several variables. *Sbornik: Mathematics*, 199(2):253, 2008.

[97] V. A. Rvachëv. Compactly-supported solutions of functional-differential equations and their applications. *Uspekhi Mat. Nauk*, 45(1(271)):77–103, 222, 1990.

[98] H.-J. Schmeisser and W. Sickel. Sampling theory and function spaces. In *Applied mathematics reviews, Vol. 1*, pages 205–284. World Sci. Publ., River Edge, NJ, 2000.

[99] H.-J. Schmeisser and H. Triebel. *Topics in Fourier analysis and function spaces*, volume 42 of *Mathematik und ihre Anwendungen in Physik und Technik [Mathematics and its Applications in Physics and Technology]*. Akademische Verlagsgesellschaft Geest & Portig K.-G., Leipzig, 1987.

[100] A. Seeger and T. Ullrich. Haar projection numbers and failure of unconditional convergence in Sobolev spaces. *Math. Zeitschrift*, 285(1):91–119, 2017.

[101] A. Seeger and T. Ullrich. Lower bounds for Haar projections: deterministic examples. *Constr. Approx.*, DOI 10.1007/s00365-016-9353-3 (to appear).

[102] W. Sickel. Approximate recovery of functions and Besov spaces of dominating mixed smoothness. In *Constructive theory of functions*, pages 404–411. DARBA, Sofia, 2003.

[103] W. Sickel. Approximation from sparse grids and function spaces of dominating mixed smoothness. In *Approximation and probability*, volume 72 of *Banach Center Publ.*, pages 271–283. Polish Acad. Sci., Warsaw, 2006.

[104] W. Sickel and T. Ullrich. The Smolyak algorithm, sampling on sparse grids and function spaces of dominating mixed smoothness. *East Journal on Approximations*, 13(4):387–425, 2007.

[105] W. Sickel and T. Ullrich. Tensor products of Sobolev-Besov spaces and applications to approximation from the hyperbolic cross. *Journal of Approximation Theory*, 161:748–786, 2009.

[106] W. Sickel and T. Ullrich. Tensor products of Sobolev-Besov spaces and applications to approximation from the hyperbolic cross. *Journal of Approximation Theory*, 161:748–786, 2009.

[107] W. Sickel and T. Ullrich. Spline interpolation on sparse grids. *Appl. Anal.*, 90(3-4):337–383, 2011.

[108] S. Smolyak. Quadrature and interpolation formulas for tensor products of certain classes of functions. *Soviet Mathematics, Doklady*, 4:240–243, 1963.

[109] S. A. Stasyuk. Best m-term approximation of the classes $B^r_{\infty,\theta}$ of functions of many variables by polynomials in the Haar system. *Ukrainian Math. J.*, 63(4):638–645, 2011.

[110] S. A. Stasyuk. Approximations of the classes $MB^r_{p,\theta}$ of periodic functions of several variables by polynomials according to the Haar system. *J. Math. Sci. (N.Y.)*, 210(1):76–85, 2015. Translation of Ukr. Mat. Visn. **12** (2015), no. 1, 97–109.

[111] E. M. Stein and G. Weiss. *Introduction to Fourier analysis on Euclidean spaces.* Princeton University Press, Princeton, N.J., 1971. Princeton Mathematical Series, No. 32.

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

[113] V. Temlyakov. On approximate recovery of functions with bounded mixed derivative. *J. Complex.*, 9(1):41–59, Mar. 1993.

[114] V. Temlyakov. *Greedy approximation*, volume 20 of *Cambridge Monographs on Applied and Computational Mathematics*. Cambridge University Press, Cambridge, 2011.

[115] V. N. Temlyakov. Approximation of periodic functions of several variables by bilinear forms. *Izv. Akad. Nauk SSSR Ser. Mat.*, 50(1):137–155, 208, 1986.

[116] V. N. Temlyakov. Estimates for the asymptotic characteristics of classes of functions with bounded mixed derivative or difference. *Trudy Mat. Inst. Steklov.*, 189:138–168, 1989. Translated in Proc. Steklov Inst. Math. **1**990, no. 4, 161–197, A collection of papers from the All-Union School on the Theory of Functions (Russian) (Dushanbe, 1986).

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

[118] H. Triebel. Interpolation theory, function spaces, differential operators. page 528, 1978.

[119] H. Triebel. *Theory of function spaces*, volume 78 of *Monographs in Mathematics*. Birkhäuser Verlag, Basel, 1983.

[120] H. Triebel. *Bases in function spaces, sampling, discrepancy, numerical integration*, volume 11 of *EMS Tracts in Mathematics*. European Mathematical Society (EMS), Zürich, 2010.

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

[122] T. Ullrich. Function spaces with dominating mixed smoothness, characterization by differences. Technical report, Jenaer Schriften zur Math. und Inform., Math/Inf/05/06, 2006.

[123] T. Ullrich. Local mean characterization of Besov-Triebel-Lizorkin type spaces with dominating mixed smoothness on rectangular domains. *Preprint*, pages 1–26, 2008.

[124] T. Ullrich. Smolyak's algorithm, sampling on sparse grids and Sobolev spaces of dominating mixed smoothness. *East J. Approx.*, 14(1):1–38, 2008.

[125] T. Ullrich. Optimal cubature in Besov spaces with dominating mixed smoothness on the unit square. *J. Complexity*, 30:72–94, 2014.

[126] T. Ullrich and B. Vedel. Hyperbolic wavelet analysis of classical isotropic and anisotropic Besov-Sobolev spaces. *Preprint*, 2016.

[127] J. Vybiral. Function spaces with dominating mixed smoothness. *Dissertationes Math. (Rozprawy Mat.)*, 436:73, 2006.

[128] H. Yserentant. *Regularity and approximability of electronic wave functions*, volume 2000 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2010.

[129] C. Zenger. Sparse grids. In *Parallel algorithms for partial differential equations (Kiel, 1990)*, volume 31 of *Notes Numer. Fluid Mech.*, pages 241–251. Friedr. Vieweg, Braunschweig, 1991.