On N-term approximation by Haar functions in H^s-norms

Peter Oswald
Bell Labs, Lucent Technologies
600 Mountain Av., Rm. 2C403
Murray Hill, NJ 07974-0636, USA
e-mail: poswald@research.bell-labs.com

Dedicated to the anniversary of P. L. Ul'yanov

Abstract

In [8], we numerically observed that spaces spanned by linear combinations of a few tensor-product Haar functions lead to surprisingly good approximation rates when solving the single layer potential equation on a square. This effect has to do with the well-known property of dimension reduction for hyperbolic cross approximation, on the one hand, and the presence of strong edge singularities in the solution of such boundary integral equations, on the other. In this note we state results on hyperbolic cross and best N-term approximation of functions by linear combinations of Haar functions in H^s-norms ($-1 < s < 1/2$) which theoretically support our numerical findings. As far as we know, the case of negative smoothness $s < 0$ has not been considered before.

1 Introduction

Let us give the definition of the Haar systems under consideration. Let $\mathcal{D}_j$ be the system of dyadic intervals $\Delta = [(i-1)2^{-j}, i2^{-j})$, $i = 1, \ldots, 2^j$, of length $|\Delta| = 2^{-j}$, $j \geq 0$, of $I \equiv [0,1]$. The characteristic function of such an interval is denoted by χ_Δ. Any $\Delta \in \mathcal{D}_j$ uniquely splits into left (Δ^+) and right (Δ^-) half-intervals from $\mathcal{D}_{j+1}$. Set

$$\phi_\Delta = |\Delta|^{-1/2}\chi_\Delta , \qquad \psi_\Delta = |\Delta|^{-1/2}(\chi_{\Delta^+} - \chi_{\Delta^-}) , \quad \Delta \in \cup_{j\geq 0}\mathcal{D}_j ,$$

for the univariate L_2-normalized box functions and Haar functions, respectively. For notational convenience, we define $\mathcal{D}_{-1} = \{[0,2]\}$ and $\psi_{[0,2]} = \phi_I$. Now, set

$$\Psi_j = \{\psi_\Delta \,:\, \Delta \in \mathcal{D}_{j-1}\} , \quad \Phi_j = \{\phi_\Delta \,:\, \Delta \in \mathcal{D}_j\} , \qquad j \geq 0 .$$

The univariate Haar system

$$\mathbf{\Psi} = \cup_{j\geq 0}\Psi_j$$

is a complete orthonormal system in $L_2(I)$ (for short *CONS*). Its sections $\mathbf{\Psi}_j = \cup_{l=0}^{j}\Psi_l$ form orthonormal bases in the spaces $V_j = \operatorname{span}\Phi_j$ of piecewise constant functions with respect to $\mathcal{D}_j$, $j \geq 0$.

To define bivariate Haar systems we agree on the following notation: if ψ_1, ψ_2 are univariate functions then $\psi = \psi_1 \otimes \psi_2$ denotes the bivariate function given by $\psi(x) = \psi_1(x_1)\psi_2(x_2)$, $x \equiv (x_1, x_2)$. Correspondingly, given sets Ψ, Φ of univariate functions, we introduce $\Psi \otimes \Phi = \{\psi \otimes \phi \,:\, \psi \in \Psi\, ,\, \phi \in \Phi\}$. Define

$$\mathbf{\Psi}(I^2) = \cup_{j=0}^{\infty}\Psi_j(I^2) , \qquad \Psi_j(I^2) = \begin{cases} \Psi_0 \otimes \Psi_0 , & j = 0 , \\ (\Psi_j \otimes \Phi_{j-1}) \cup (\Phi_{j-1} \otimes \Psi_j) \cup (\Psi_j \otimes \Psi_j) , & j \geq 1 , \end{cases} \tag{1}$$

and

$$\mathbf{\Psi}^*(I^2) = \cup_{j_1,j_2=0}^{\infty}\Psi_{j_1,j_2} , \qquad \Psi_{j_1,j_2} = \Psi_{j_1} \otimes \Psi_{j_2} , \quad j_1, j_2 \geq 0 . \tag{2}$$

Both systems form CONS in $L_2(I^2)$. While $\mathbf{\Psi}^*(I^2)$ represents the standard tensor-product construction, from the viewpoint of applications $\mathbf{\Psi}(I^2)$ is more popular since the supports of its functions are associated with dyadic squares and are better localized. Roughly speaking, the system $\mathbf{\Psi}(I^2)$ corresponds to *isotropic refinement* while $\mathbf{\Psi}^*(I^2)$ covers *anisotropic behavior.*

In this note, we will study nonlinear approximation processes in the Sobolev spaces $H^s(I^2)$, $-1 < s < 1/2$, associated with the system $\mathbf{\Psi}^*(I^2)$. In particular, we will look for *best N-term approximation rates* with respect to $\mathbf{\Psi}^*(I^2)$, i.e., for estimates of the quantities

$$e^*_N(f)_s = \inf_{\Psi\subset\mathbf{\Psi}^*(I^2)\,:\,\#\Psi\le N} e_\Psi(f)_s\ , \quad N\to\infty \qquad (\ e_\Psi(f)_s = \inf_{g\in\operatorname{span}\Psi}\|f-g\|_{H^s}\ ). \tag{3}$$

under various assumptions on f. Occasionally, we will compare the behavior of $e^*_N(f)_s$ with the analogously defined best N-term approximations $e_N(f)_s$ with respect to the system $\mathbf{\Psi}(I^2)$. Since the Haar systems under consideration are CONS in $L_2(I^2)$ (and Riesz bases in $H^s(I^2)$ for $-1/2 < s < 1/2$, see below), this might seem a rather trivial question, and we need to provide some motivation.

First of all, this topic was triggered by our investigations [8] of the Galerkin method for the *single layer potential equation*

$$Tf \equiv \frac{1}{4\pi}\int_{I^2}\frac{f(y)}{|x-y|_2}\,dy = g(x)\ , \tag{4}$$

on I^2 which has applications in electrostatics. When cast in a variational setting, (4) leads to a symmetric positive definite variational problem in $H^{-1/2}(I^2)$. Due to the specifics of discretization methods for boundary integral equtions of this type (dense discretization matrices), one is strongly interested in discretization spaces of small dimension with good $H^{-1/2}$-approximation properties for the weak solutions f of (4). Recently, much research has been conducted on *adaptive wavelet methods*, see [2] for an overview. Theoretically, these methods are related to best N-term approximation from systems such as $\mathbf{\Psi}(I^2)$, see [3, 2]. Alternatively, so-called *hp-boundary element methods* based on piecewise polynomials with variable partition and variable polynomial degree have been tried and lead even to exponential convergence rates if g is sufficiently smooth [16].

In [8], we investigated the potential of *hyperbolic cross spaces*, i.e., we considered the choice $\mathbf{\Psi}^*_J(I^2) = \cup_{j_1+j_2\le J}\,\Psi_{j_1,j_2}$ in a Galerkin method for (4). In connection with finite element and boundary element methods, this type of approximation method has been introduced under the name *sparse grid method* [22]. Error rates are related to the best approximations $e_{\mathbf{\Psi}^*_J(I^2)}(f)_{-1/2}$, and can be found in [8]. Under certain additional regularity assumptions on mixed derivatives (see [17] for the corresponding theory in the periodic case), these rates, when formulated in dependence on the dimension N of the ansatz spaces, show the superiority of the sparse grid method over standard approximation schemes which typically use the much larger sets $\mathbf{\Psi}_J(I^2) = \cup_{j=0}^J\Psi_j(I^2)$ or $\mathbf{\Phi}_J(I^2) = \Phi_J\otimes\Phi_J$. Unfortunately, the solutions of (4) do not meet such regularity assumptions. E.g., if $g(x)\equiv 1$ in (4) (this is the *capacity problem*) then the solution f belongs to $H^{-\varepsilon}(I^2)$ for any $\varepsilon > 0$ but not to $L_2(I^2)$. This is due to singularity behavior of f near the vertices and edges of I^2 (see subsection 2.3). In our numerical tests in [8], we observed surprisingly good convergence rates when we tried to match this singularity behavior by constructing *adapted sparse grid spaces*

$$V_\Psi = \operatorname{span}\Psi\ , \quad \Psi\subset\mathbf{\Psi}^*(I^2)\ , \quad \#\Psi\le N\ , \tag{5}$$

by selecting functions $\psi\in\mathbf{\Psi}^*(I^2)$ with support along edges and near vertices with some preference. Altogether, this serves as a strong motivation to investigate the asymptotical behavior of the best N-term approximations (3), $N\to\infty$, especially for $s = -1/2$ and typical solutions of (4).

On the other hand, *efficient characterization theorems* for best N-term approximation with respect to tensor-product systems such as $\mathbf{\Psi}^*_J(I^2)$ are a widely open question (in contrast, the connection between the asymptotical behavior of the quantities $e_N(f)_s$ (and their generalizations to the L_p-setting) and smoothness properties of f is well-investigated [3]). Only a few papers deal with similar subjects [18, 5, 7]. Thus, we believe that investigating some special cases will add to a better understanding of this problem. In particular, extensions to $s < 0$ show some new effects which is worth knowing.

The material of this paper is organized as follows. In section 2 some preliminary information is collected. Since this material is essentially known and covered in [8, 13] (or relatively easy to derive from other references cited below), proofs are restricted to the absolute minimum. We give a brief account of the situation with H^s-approximation ($s < 0$) in the periodic case, to draw the attention of the reader to some differences to the case $s \geq 0$ considered in [17]. Next, the connection between Haar-Fourier coefficients and H^s-norms of functions is made precise. On the way, definitions of the Sobolev spaces $H^s(I^2)$ and $H^t_{\mathrm{mix}}(I^2)$ (with dominating mixed t-th derivative) used in this paper are included. Finally, we quote some regularity results for (4).

Section 3 contains the main results. These concern *upper estimates* for the quantities (3), $-1 < s < 1/2$, if $f \in H^t_{\mathrm{mix}}(I^2)$ or if $f \in L_1(I^2)$ possesses certain bounds on its derivatives (allowing for singularities along ∂I^2). Partially, these results are sharp. Applied to solutions of (4), with smooth right-hand sides g, these results imply

$$e^*_N(f)_{-1/2} \leq C_f N^{-5/4} , \quad N \to \infty , \tag{6}$$

which compares favorably with rates for the quantities $e_N(f)_{-1/2}$. The numerical verification and consequences of this result will not be discussed here.

Throughout the paper, $A \asymp B$ stands for a two-sided inequality between the expressions A and B, i.e., $cB \leq A \leq C \cdot B$, where $C, c > 0$ denote generic constants the value of which may change with each appearance. Dependencies of such constants c, C on parameters will not be indicated all the time but should be clear from the context. By $\#\mathcal{A}$ we denote the number of different elements of a finite set $\mathcal{A}$. For normed spaces X, Y, the notation $X \cong Y$ means that the spaces are identical as sets, and possess equivalent norms: $\| \cdot \|_X \asymp \| \cdot \|_Y$. Most of the spaces introduced below are Hilbert spaces which we define by indicating only the corresponding norm (the reader should be able to recover the associated scalar product).

2 Notation and auxiliary results

2.1 The periodic case

This section is included only to give the reader a certain orientation for the following exposition and to draw attention to some differences between H^s-approximation with $s < 0$ and $s \geq 0$. For simplicity, let us define periodic Sobolev spaces directly via formal Fourier series (periodic distributions) as follows. For $-\infty < s < \infty$, set

$$H^s(\mathbb{T}^d) = \{ f(\theta) \sim \sum_{\alpha \in \mathbb{Z}^d} c_\alpha e^{-i\alpha\theta} \; : \; \|f\|_{H^s} \equiv (\sum_{\alpha \in \mathbb{Z}^d} (1 + |\alpha|)^{2s} |c_\alpha|^2)^{1/2} < \infty \} ,$$

and

$$H^s_{\mathrm{mix}}(\mathbb{T}^d) = \{ f(\theta) \sim \sum_{\alpha \in \mathbb{Z}^d} c_\alpha e^{-i\alpha\theta} \; : \; \|f\|_{H^s_{\mathrm{mix}}} \equiv (\sum_{\alpha \in \mathbb{Z}^d} (\prod_{k=1}^{d} (1 + |\alpha_k|)^{2s}) |c_\alpha|^2)^{1/2} < \infty \} .$$

Throughout the paper, $d \geq 1$ is fixed (later we specialize to $d = 2$). Clearly, for $s \geq 0$ these are subspaces of $L_2(\mathbb{T}^d)$, and $H^{-s}(\mathbb{T}^d) \cong H^s(\mathbb{T}^d)'$ by definition of dual spaces. In particular, the Fourier coefficients of $f \in H^{-s}(\Omega)$ can be recovered as the value of the associated functional $f \in H^s(\mathbb{T}^d)$ evaluated at $e^{-i\alpha\theta} \in H^s(\mathbb{T}^d)$. Analogous interpretations are possible for $H^s_{\mathrm{mix}}(\mathbb{T}^d)$. Thus, as long as we are in the L_2-setting, Sobolev spaces are essentially reduced to the sequence spaces of Fourier coefficients (the L_p-case, $p \neq 2$, is much more elaborate, see [17]).

Define hyperbolic cross spaces $V^*_n(\mathbb{T}^d) \equiv V_{\mathbf{\Psi}^*_n(\mathbb{T}^d)}$ by setting

$$\mathbf{\Psi}^*_n(\mathbb{T}^d) = \{ e^{-i\alpha\theta} \; : \; \alpha \in \mathbb{Z}^d , \; \prod_{k=1}^{d} (1 + |\alpha_k|) \leq n \} , \quad n \geq 1 ,$$

Their counterpart for piecewise constant approximation is $\mathbf{\Psi}_J^*(I^2)$ (let $d=2$, $n=2^J$, and replace exponentials by Haar functions). Note that $\#\mathbf{\Psi}_n^*(\mathbb{T}^d) \asymp n(1+\log n)^{d-1}$. Concerning best approximations, we adopt the same notation as introduced in section 1, with the only difference that now all $\Psi \subset \mathbf{\Psi}^*(\mathbb{T}^d) \equiv \{e^{-i\alpha\theta} \,:\, \alpha \in \mathbb{Z}^d\}$. The next statement, although trivial, indicates a difference between $s<0$ and $s \geq 0$.

Proposition 1 *Let $f \in H^t_{\text{mix}}(\mathbb{T}^d)$.*
a) *The best H^s-approximations with respect to hyperbolic cross spaces $\mathbf{\Psi}_n^*(\mathbb{T}^d)$, $n \geq 1$, satisfy*

$$e_{\mathbf{\Psi}_n^*(\mathbb{T}^d)}(f)_s \leq C\|f\|_{H^t_{\text{mix}}} \begin{cases} n^{-(t-s)}\,, & 0 \leq s < t < \infty\,, \\ n^{-(t-s/d)}\,, & s<0\,,\ \ s/d < t < \infty\,. \end{cases} \tag{7}$$

b) *The best N-term approximations with respect to $\mathbf{\Psi}^*(\mathbb{T}^d)$ in $H^s(\mathbb{T}^d)$, $N \geq 1$, satisfy*

$$e_N^*(f)_s \leq C\|f\|_{H^t_{\text{mix}}} \begin{cases} N^{-(t-s)}\,, & 0 < s < t < \infty\,, \\ N^{-t}(1+\log N)^{(d-1)t}\,, & s = 0 < t < \infty\,, \\ N^{-(t-s/d)}\,, & s<0\,,\ \ s/d < t < \infty\,. \end{cases} \tag{8}$$

The estimates in (7), (8) cannot be improved on the class $H^t_{\text{mix}}(\mathbb{T}^d)$, neither with respect to the asymptotics for $n, N \to \infty$ nor with respect to the parameter range (s,t).

For the case $s=0$, this is covered in [17, Chapter III] (set $p=q=2, r=t$ there), where more general questions are considered. All upper estimates follow from the estimate

$$e_\Psi(f)_s^2 \leq \max_{\alpha \notin \Lambda_\Psi} \frac{(1+|\alpha|)^{2s}}{\prod_{k=1}^d (1+|\alpha_k|)^{2t}} \, \|f\|^2_{H^t_{\text{mix}}}\,, \quad \Psi \subset \mathbf{\Psi}^*(\mathbb{T}^d)\,,$$

where $\Lambda_\Psi = \{\alpha \in \mathbb{Z}^d \,:\, e^{-i\alpha\theta} \in \Psi\}$. The choices for Ψ are $\mathbf{\Psi}_n^*(\mathbb{T}^d)$ in **a)** and for $s=0$ in **b)** while for $s \neq 0$ in **b)** some modifications are required. We leave this and the construction of counterexamples as an exercise upon the reader (some hints can be found in section 3 where we derive similar results for the bivariate Haar system (2)).

Note that the setting of **b)** is not adequate to judge the potential of nonlinear N-term approximation. One should ask which properties of f would guarantee a certain rate of best N-term approximation, rather than perform a worst case analysis for an a priori chosen, large class of functions such as $H^t_{\text{mix}}(\mathbb{T}^d)$, compare [3]. Since $\mathbf{\Psi}^*(\mathbb{T}^d)$ becomes, after suitable scaling, a CONS in any of the Hilbert spaces $H^s(\mathbb{T}^d)$, one can give, following E. Schmidt and S. B. Stechkin, *necessary and sufficient conditions* for estimates of the type

$$e_N^*(f)_s \leq CN^{-r}\,, \quad N \to \infty\,,$$

in terms of the Fourier coefficients of f, see [6],[3, section 5] (this is an alternative route to take for deriving (8)).

We have formulated this result to show that H^s-approximation of functions from spaces with dominating mixed derivatives exhibits different qualitative behavior for $s<0$ and $s>0$. In particular, for $s<0$ estimates again become *dimension-dependent*. To beat the *curse of dimensionality* was one of the main reasons for considering hyperbolic cross approximation as a practical tool, and to look more closely at $H^t_{\text{mix}}(\mathbb{T}^d)$ instead of $H^t(\mathbb{T}^d)$. A simple explanation of why $-s$ should be replaced by $-s/d$ comes from the observation that the embedding relation

$$H^t_{\text{mix}}(\mathbb{T}^d) \subset H^s_{\text{mix}}(\mathbb{T}^d) \subset H^s(\mathbb{T}^d)\,, \qquad 0 \leq s < t < \infty\,,$$

does not hold for $s<0$, and can only be replaced by

$$H^t_{\text{mix}}(\mathbb{T}^d) \subset H^{s/d}_{\text{mix}}(\mathbb{T}^d) \subset H^s(\mathbb{T}^d)\,, \qquad s/d < t < \infty\,,\ s<0\,.$$

2.2 Haar coefficients and Sobolev norms

We give the definition of Sobolev spaces on I^d in a fashion that is convenient for the Hilbert space setting of this paper (for generalities on Besov-Sobolev spaces, also in connection with approximation methods see [1, 11, 12, 19]). Let first $d = 1$. Set $H^0(I) = L^2(I)$, and

$$H^m(I) = \{f \in L_2(I) \;:\; f^{(m)} \in L_2(I),\ \|f\|_{H^m} = (\|f\|_{L_2}^2 + \|f^{(m)}\|_{L_2}^2)^{1/2}\}$$

for integer $m \geq 1$. Finally, for the remaining $s > 0$, we use real interpolation to define

$$H^s(I) = [L_2(I), H^m(I)]_{s/m,2} \;, \quad 0 < s < m$$

(using different $m > s$ leads to the same space, with equivalent norms). Equivalently, $H^s(I)$ can be identified with Besov spaces

$$H^s(I) \cong B^{s;m}_{2,2}(I) = \{f \in L_2 \;:\; \|f\|_{B^{s;m}_{2,2}} = (\|f\|_{L_2}^2 + \sum_{l=1}^{\infty} 2^{2ls}\omega_m(2^{-l}, f)_{L_2}^2)^{1/2} < \infty\} \;, \quad 0 < s < m \;,$$

where $\omega_m(2^{-l}, f)_{L_2}$ denotes the L_2-moduli of smoothness of order m of a function $f \in L_2(I)$ (see [4, 19] for details). This opens the way to describe $H^s(I)$ by approximation methods (see below for the Haar case).

For $s < 0$, one uses duality. Set

$$\|f\|_{H^s}(I) = \sup_{0 \neq v \in H^{-s}(I)} \frac{(f, v)_{L_2}}{\|v\|_{H^{-s}}} \;, \quad f \in L_2(I) \;,\ s < 0 \;, \tag{9}$$

and define $H^s(I)$ as the closure of $L_2(I)$ under the norm (9). This is equivalent to setting $H^s(I) = H^{-s}(I)'$, $s < 0$, since the embedding $H^{-s}(I) \subset L_2(I)$ is dense. We prefer the above definition because it is the only way to explicitely work with H^s-norms for $s < 0$, and to avoid the introduction of distributions. All dual spaces below should be understood analogously.

For $d > 1$, we act by tensor-product techniques. Definitions of tensor products of Hilbert spaces and of tensor-product operators acting on them can be found in [21]. All results used here are elementary, and can be deduced from the basic information given in [21]. For the ease of exposition, we will concentrate on $d = 2$. Set

$$H^{s_1,s_2}_{\mathrm{mix}}(I^2) = H^{s_1}(I) \otimes H^{s_2}(I) \;, \quad -\infty < s_1, s_2 < \infty \;,$$

and define

$$H^s(I^2) = \begin{cases} H^{s,0}_{\mathrm{mix}}(I^2) \cap H^{0,s}_{\mathrm{mix}}(I^2) \;, & s \geq 0 \;, \\ H^{-s}(I^2)' \;, & s < 0 \;, \end{cases} \qquad H^s_{\mathrm{mix}}(I^2) = H^{s,s}_{\mathrm{mix}}(I^2) \;. \tag{10}$$

If X, Y are Hilbert spaces the norm for their intersection $X \cap Y$ is defined by $\|f\|_{X\cap Y}^2 = \|f\|_X^2 + \|f\|_Y^2$. $H^s_{\mathrm{mix}}(I^2)$ is called Sobolev space with dominating mixed derivative of order s. E.g., it is easy to prove that

$$\|f\|_{H^1_{\mathrm{mix}}}^2 \asymp \|f\|_{L_2}^2 + \|D^{(1,0)}f\|_{L_2}^2 + \|D^{(0,1)}f\|_{L_2}^2 + \|D^{(1,1)}f\|_{L_2}^2 \qquad \forall\, f \in H^1_{\mathrm{mix}}(I^2) \;.$$

Also, by definition of tensor products of Hilbert spaces, we have $(X \otimes Y)' \cong X' \otimes Y'$, and therefore $H^s_{\mathrm{mix}}(I^2) \cong H^{-s}_{\mathrm{mix}}(I^2)'$, $s < 0$.

What we are interested in is the connection between Sobolev norms and coefficient norms for Haar series. In contrast to subsection 2.1, restrictions should be expected due to the saturation and limited regularity properties of piecewise constant approximation. It is worthwhile to start with the one-dimensional case. Introduce Hilbert spaces $A^s(I)$ for $s \geq 0$ by setting

$$A^s(I) = \{f \in L_2(I) \;:\; \|f\|_{A^s} = |||f|||_s \equiv (\sum_{\Delta \in \mathcal{D}} |\Delta|^{-2s} c_\Delta(f)^2)^{1/2} < \infty\} \;, \tag{11}$$

where $c_\Delta(f) = (f, \psi_\Delta)_{L_2}$ denotes the Haar-Fourier coefficients of f, and

$$f = \sum_{\Delta \in \mathcal{D}} c_\Delta(f) \psi_\Delta \ , \qquad \mathcal{D} \equiv \cup_{j=-1}^{\infty} \mathcal{D}_j \ ,$$

represents the Haar-Fourier series (these definitions can be extended to arbitrary $f \in L_1(I)$, see [10] for a collection of basic facts on the univariate Haar system). The embedding $A^s(I) \subset L_2(I)$ is continuous and dense for $s > 0$; for $s = 0$ these spaces coincide. In the case $s < 0$, we define $A^s(I)$ as the closure of $L_2(I)$ under the norm $|||\cdot|||_s$ introduced in (11). Due to the CONS-property of $\mathbf{\Psi}$ in $L_2(I)$, we have

$$|||f|||_s = \sup_{0 \neq v \in A^{-s}(I)} \frac{(f, v)_{L_2}}{|||v|||_{-s}} \ , \quad f \in L_2(I) \ , \tag{12}$$

and thus $A^s(I) \cong A^{-s}(I)'$ for $s < 0$. However, the expression $|||f|||_s$ does not necessarily make sense for arbitrary $f \in A^s(I^d)$, $s < 0$, since the Haar-Fourier coefficients might not be defined properly.

Clearly, $H^0(I) = L_2(I) = A^0(I)$ for $s = 0$. Direct and inverse inequalities for the approximation by Haar polynomials in $L_p(I)$ have been obtained by Ul'yanov [20] and Golubov [9] (see also [4]), and lead, in conjunction with $H^s(I) \cong B^{s,1}_{2,2}(I)$, $0 < s < 1$, to

$$\begin{aligned} \|f\|_{A^s} &\leq C \|f\|_{H^s} \qquad \forall\, f \in H^s(I) \ , \quad 0 < s < 1 \ , \\ \|f\|_{H^s} &\leq C \|f\|_{A^s} \qquad \forall\, f \in A^s(I) \ , \quad 0 < s < 1/2 \ . \end{aligned}$$

By using these bounds together with (9) and (12), one derives complementary bounds for $s < 0$:

$$\begin{aligned} \|f\|_{A^s} &\leq C \|f\|_{H^s} \qquad \forall\, f \in H^s(I) \ , \quad -1/2 < s < 0 \ , \\ \|f\|_{H^s} &\leq C \|f\|_{A^s} \qquad \forall\, f \in A^s(I) \ , \quad -1 < s < 0 \ . \end{aligned}$$

Indeed, consider any $f \in L_2(I)$, and $-1/2 < s < 0$. Then $0 < -s < 1/2$, and we have

$$\|f\|_{A^s} \sup_{0 \neq v \in A^{-s}(I)} \frac{(f, v)_{L_2}}{\|v\|_{A^{-s}}} \leq C^{-1} \sup_{0 \neq v \in H^{-s}(I)} \frac{(f, v)_{L_2}}{\|v\|_{H^{-s}}} = C^{-1} \|f\|_{H^s} \ .$$

The usual density argument yields the first inequality for all $f \in H^s(I)$, $-1/2 < s < 0$. The argument for the second inequality is analogous. Consequently, we have

Proposition 2 *For the A^s-spaces associated with the univariate Haar system $\mathbf{\Psi}$, we have*

$$\begin{aligned} \|f\|_{A^s} &\leq C \|f\|_{H^s} \qquad \forall\, f \in H^s(I) \ , \quad -1/2 < s < 1 \ , \\ \|f\|_{H^s} &\leq C \|f\|_{A^s} \qquad \forall\, f \in A^s(I) \ , \quad -1 < s < 1/2 \ . \end{aligned}$$

In particular, $H^s(I) \cong A^s(I)$ if and only if $-1/2 < s < 1/2$. Equivalently, the scaled Haar system $\mathbf{\Psi}^s = \{\psi^s_\Delta \equiv |\Delta|^s \psi_\Delta \,:\, \Delta \in \cup_{j=-1}^{\infty} \mathcal{D}_j\}$ is a Riesz basis in $H^s(I)$ if and only if $-1/2 < s < 1/2$.

That the inequalities cannot hold for $s \geq 1$ resp. $s \geq 1/2$ follows from the saturation properties of piecewise constant approximation (consider, for instance, $f(x) = x$) resp. from the fact that Haar functions ψ_Δ, $\Delta \in \mathcal{D}_j$, $j \geq 0$, do not belong to $H^{1/2}(I)$. A counterexample that shows the failure of the first inequality for $s = -1/2$ is contained in [13, section 4].

Since the Haar system is a complete orthogonal basis in any of the Hilbert spaces $A^s(I)$ by definition, it is easy to identify the norm of the spaces

$$A^{s_1,s_2}(I^2) = A^{s_1}(I) \otimes A^{s_2}(I) \ , \quad -\infty < s_1, s_2 < \infty \ ,$$

as

$$\|f\|_{A^{s_1,s_2}} = \Big(\sum_{\Delta \in \mathcal{D}^*(I^2)} |\Delta'|^{-2s_1} |\Delta''|^{-2s_2} c_\Delta(f)^2 \Big)^{1/2} \ , \quad f \in A^{s_1,s_2}(I^2) \cap L_2(I^2) \ ,$$

where $c_\Delta(f) = \int_{I^2} f(x)\psi_\Delta(x)\,dx$ now denotes the Haar-Fourier coefficient of f with respect to $\psi_\Delta \in \mathbf{\Psi}^*(I^2)$. The following notation will be used throughout the paper:

$$\mathcal{D}^*(I^2) = \cup_{j_1,j_2=0}^{\infty}\mathcal{D}_{j_1,j_2}\ , \qquad \mathcal{D}_{j_1,j_2} = \{\Delta \equiv \Delta' \times \Delta'' \ :\ \Delta' \in \mathcal{D}_{j_1-1},\ \Delta' \in \mathcal{D}_{j_2-1}\},\ j_1,j_2 \geq 0\ ,$$

and $\psi_\Delta \equiv \psi_{\Delta'} \otimes \psi_{\Delta''}$ are the associated Haar functions. By $d(\Delta) = \min\{|\Delta'|,|\Delta''|\}$ we denote the maximum of the side-lengths of the rectangle Δ. Then, taking into account the definition of the Sobolev spaces $H^s(I^2)$ and $H^s_{\mathrm{mix}}(I^2)$ from $H^{s_1,s_2}(I^2)$, it is natural to introduce for the following spaces:

$$A^s(I^2) = \left\{ \begin{array}{ll} \{f \in L_2(I^2)\ :\ \|f\|_{A^s} = (\sum_{\Delta\in\mathcal{D}^*(I^2)} d(\Delta)^{-2s}c_\Delta(f)^2)^{1/2} < \infty\}\ , & s \geq 0\ , \\ A^{-s}(I^2)'\ , & s < 0\ , \end{array}\right. \tag{13}$$

and

$$A^s_{\mathrm{mix}}(I^2) = A^{s,s}_{\mathrm{mix}}(I^2)\ , \qquad A^{s_1,s_2}_{\mathrm{mix}}(I^2) = A^{s_1}(I) \otimes A^{s_2}(I)\ , \quad -\infty < s, s_1, s_2 < \infty\ . \tag{14}$$

Again, as long as $f \in A^s(I^2) \cap L_2(I^2)$, the above explicit expression for $\|f\|_{A^s}$ can be used also for $s < 0$. Analogously,

$$\|f\|_{A^{s_1,s_2}_{\mathrm{mix}}} = (\sum_{\Delta\in\mathcal{D}^*(I^2)} |\Delta'|^{-2s_1}|\Delta''|^{-2s_2}c_\Delta(f)^2)^{1/2}\ , \qquad f \in A^{s_1,s_2}_{\mathrm{mix}}(I^2) \cap L_2(I^2)\ . \tag{15}$$

Note that $A^s(I)^2 \cong A^{s,0}_{\mathrm{mix}}(I^2) \cap A^{0,s}_{\mathrm{mix}}(I^2)$ for $s \geq 0$ by definition of $d(\Delta)$.

After these preparations, the following corollary to Proposition 2 is obvious.

Proposition 3 *For the $A^s(I^2)$-norms (13) associated with the bivariate Haar system $\mathbf{\Psi}^*(I^2)$, we have*

$$\begin{array}{lll} \|f\|_{A^s} \leq C\|f\|_{H^s} & \forall\, f \in H^s(I^2)\ , & \\ \|f\|_{A^s_{\mathrm{mix}}} \leq C\|f\|_{H^s_{\mathrm{mix}}} & \forall\, f \in H^s_{\mathrm{mix}}(I^2)\ , & -1/2 < s < 1\ , \end{array} \tag{16}$$

$$\begin{array}{lll} \|f\|_{H^s} \leq C\|f\|_{A^s} & \forall\, f \in A^s(I^2)\ , & \\ \|f\|_{H^s_{\mathrm{mix}}} \leq C\|f\|_{A^s_{\mathrm{mix}}} & \forall\, f \in A^s_{\mathrm{mix}}(I^2)\ , & -1 < s < 1/2\ . \end{array} \tag{17}$$

In particular, $H^s(I^2) \cong A^s(I^2)$ resp. $H^s_{\mathrm{mix}}(I^2) \cong A^s_{\mathrm{mix}}(I^2)$ if and only if $-1/2 < s < 1/2$ (equivalently, this can be stated as the Riesz basis property of certain scaled bivariate Haar systems).

To see how Proposition 3 follows from Proposition 2, just apply the definitions and the fact that if the Hilbert spaces X_k, Y_k satisfy $X_k \subset Y_k$ (continuous embedding), $k = 1,2$, then $X_1 \otimes X_2 \subset Y_1 \otimes Y_2$, with continuous embedding operator. An analogous statement can be formulated for the Haar system $\mathbf{\Psi}(I^2)$ defined by (1) and the spaces $H^s(I^2)$ (but not for $H^s_{\mathrm{mix}}(I^2)$), compare [13].

In section 3, we will use the first inequality in (17) and the second inequality in (16) to reduce the study of H^s-approximation properties of functions from H^t_{mix}-spaces to estimates between the corresponding spaces $A^s(I^2)$ and $A^t_{\mathrm{mix}}(I^2)$. Two slight extensions of Proposition 3 will be needed, too. First, we need to estimate H^s-norms for certain functions $f \in L_1(I^2)$ for which (17) extends as follows. Assume that $f \in L_1(I^2)$ satisfies

$$|||f|||_s^2 = \sum_{\Delta\in\mathcal{D}^*(I^2)} d(\Delta)^{-2s}c_\Delta(f)^2 < \infty \tag{18}$$

for some $-1 < s < 1/2$. If $s \geq 0$ this implies $f \in L_2(I^2)$ and $f \in A^s(I^2)$ with $\|f\|_{A^s} = |||f|||_s$. Therefore, if $0 \leq s < 1/2$ then by (17) we also have $f \in H^s(I^2)$ and $\|f\|_{H^s} \leq C|||f|||_s$. For $s < 0$, we consider the partial sums (with respect to squares) of the Haar-Fourier series of f:

$$v_j(f) = \sum_{\Delta\in\mathcal{D}^*:\, d(\Delta)\geq 2^{-j}} c_\Delta(f)\psi_\Delta\ , \qquad j \geq 0\ .$$

From the assumptions it immediately follows that $\{v_j(f)\} \subset L_2(I^2)$ is also a Cauchy sequence in $A^s(I^2)$, and defines an element $\tilde{f} \in A^s(I^2)$ with $\|\tilde{f}\|_{A^s} = |||f|||_s$. Moreover, if $-1 < s < 0$ then by (17) it follows that $\tilde{f} \in H^s(I^2)$ with $\|\tilde{f}\|_{H^s} \le C|||f|||_s$. Thus, under the assumption (18) we have

$$\|\tilde{f}\|_{H^s} \le C|||f|||_s \ , \quad -1 < s < 1/2 \ , \ \ f \in L_1(I^2) \ , \tag{19}$$

where $\tilde{f} = \lim_{j\to\infty} v_j(f)$ is an element of $H^s(I^2) \subset A^s(I^2)$. Since also $v_j(f) \to f$ in $L_1(I^d)$, we can identify f and $\tilde{f}$.

The other extension concerns the case $s = 1$ for which we mention the following weaker version of the second inequality in (16):

$$2^{2(j_1+j_2)} \sum_{\Delta\in\mathcal{D}_{j_1,j_2}} |c_\Delta(f)|^2 \le C\|f\|^2_{H^1_{\mathrm{mix}}} \quad j_1, j_2 \ge 0 \ , \ \ f \in H^1_{\mathrm{mix}}(I^2) \ . \tag{20}$$

This inequality follows by a tensor-product argument (see [8]) from the univariate Jackson-type estimate

$$\sum_{\Delta\in\mathcal{D}_{j-1}} |c_\Delta(f)|^2 \le \|f - \sum_{\Delta\in\cup_{l=-1}^{j-2}\mathcal{D}_l} c_\Delta(f)\psi_\Delta\|^2_{L_2} \le C2^{-j}\|f\|^2_{H^1} \ , \quad j \ge 0 \ , \ \ f \in H^1(I) \ .$$

2.3 Singularity functions

Finally, let us introduce a simple class of singularity functions. We consider $f \in L_1(I^2)$ which have continuous derivatives $D^{(k,l)}f$ of order $k, l \le m$ in the *interior* of I^2. Given $0 \le \alpha, \beta < 1$, we call f an *edge singularity function of type* $(m; \alpha, \beta)$ associated with the vertex $(0, 0)$ if its partial derivatives up to order m in each direction can be majorized by

$$|D^{(k,l)}f(x_1, x_2)| \le Cx_1^{-(\alpha+k)}x_2^{-(\beta+l)} \ , \quad x_1, x_2 \in (0, 1) \ , \ \ 0 \le k, l \le m \ . \tag{21}$$

For $0 \le \alpha < 2$, we call f an *point singularity function of type* $(m; \alpha)$ associated with the vertex $(0, 0)$ if its partial derivatives up to order m in each direction can be majorized by

$$|D^{(k,l)}f(x_1, x_2)| \le Cr^{-(\alpha+k+l)} \ , \qquad r \equiv \sqrt{x_1^2 + x_2^2} \ , \ \ x_1, x_2 \in (0, 1) \ , \ \ 0 \le k, l \le m \ . \tag{22}$$

Analogous definitions are introduced for the other vertices of I^2. These definitions are designed to cover strong edge and vertex singularities. Notice that we are essentially working with *majorants*, the actual singularity functions might be more general. E.g., straightforward differentiation reveals that the function

$$f(x, y) = r^{-\gamma}x_1^{-\alpha}x_2^{-\beta} \ , \qquad r = \sqrt{x_1^2 + x_2^2} \ ,$$

is an edge singularity function of type $(m; \alpha', \beta')$ where $\alpha' = \max(\alpha+\gamma, \alpha)$, $\beta' = \max(\beta+\gamma, \beta)$. From the arguments to follow it will be clear that weaker assumptions might suffice, e.g., local L_1-bounds for the function g near the boundary of the square and local L_1-estimates for the highest m-th order derivatives would be enough in some cases. Also, a more complete consideration would require logarithmical terms to be incorporated. We will not go into all these refinements here but rather proceed with the above definition of singularity functions. For the approximation schemes associated with the Haar systems of subsection 2.2, the case $m = 1$ suffices.

Here is some information for the example of specific interest to us. From the results of [14, 15], one obtains that for $g \in H^3(I^2)$ the solution f of (4) possesses a decomposition

$$f = f^{reg} + f^{sing} \ , \qquad f^{reg} \in H^2(I^2) \subset H^1_{\mathrm{mix}}(I^2) \ , \tag{23}$$

where the singular part f^{sing} is a linear combination of edge singularity functions of type $(m; 1-\gamma, 1-\gamma)$ $(\gamma = 0.2966...)$ associated with the four vertices of I^2. In particular, (23) holds for the solution of the capacitance problem, where $g(x) = 1$. This rough information is, as we will see, sufficient for our purposes. A closer look into [15] shows that f^{sing} could also be represented as a linear combination of edge singularity functions of type $(m; 1/2, 1/2)$ and *point singularity functions* of type $(m; 1-\gamma)$ with respect to the vertices. No better parameters than $\alpha = \beta = 1/2$ resp. $\alpha = 1-\gamma$ can be obtained. In all these statements, $m \geq 1$ can be chosen arbitrarily.

3 Best N-term approximation from $\mathbf{\Psi}^*(I^2)$

We first prove an analog of Proposition 1 for the Haar case.

Theorem 4 *Let $-1 < s < 1/2$, $-1/2 < tle1$, and consider arbitrary $f \in H^t_{\mathrm{mix}}(I^2)$.*
a) *The best H^s-approximations with respect to hyperbolic cross spaces $V^*_J(I^2) \equiv V_{\mathbf{\Psi}^*_J(I^2)}$, $J \geq 0$, satisfy*

$$e_{\mathbf{\Psi}^*_J(I^2)}(f)_s \leq C\|f\|_{H^t_{\mathrm{mix}}} \begin{cases} 2^{-J(t-s)} , & 0 \leq s < t , \\ 2^{-J(t-s/2)} , & s < 0 , \;\; s/2 < t . \end{cases} \tag{24}$$

b) *The best N-term approximations with respect to $\mathbf{\Psi}^*(I^2)$ in $H^s(I^2)$, $N \geq 1$, satisfy*

$$e^*_N(f)_s \leq C\|f\|_{H^t_{\mathrm{mix}}} \begin{cases} N^{-(t-s)} , & 0 < s < 1/2 , \;\; s < t , \\ N^{-t}(1+\log N)^t , & s = 0 < t < 1 , \\ N^{-1}(1+\log N)^{3/2} , & s = 0 , \; t = 1 , \\ N^{-(t-s/2)} , & -1 < s < 0 , \;\; s/2 < t . \end{cases} \tag{25}$$

Proof. Let $-1 < s < 1/2$, $-1/2 < t < 1$ (the limit case $t = 1$ will be dealt with later). For any finite subset $\Psi \subset \mathbf{\Psi}^*(I^2)$ with support $\Lambda_\Psi = \{\Delta \in \mathcal{D}^*(I^2) \, : \, \psi_\Delta \in \Psi\}$ and any $f \in H^t_{\mathrm{mix}}(I^2)$, by using (17) and (16) we have

$$\begin{aligned} e_\Psi(f)^2_s &\leq C \inf_{v \in V_\Psi} \|f - v\|^2_{A^s} = C \sum_{\Delta \notin \Lambda_\Psi} d(\Delta)^{-2s} |c_\Delta(f)|^2 \\ &\leq C \max_{\Delta \notin \Lambda_\Psi} \frac{d(\Delta)^{-2s}}{(|\Delta'||\Delta''|)^{-2t}} \sum_{\Delta \notin \Lambda_\Psi} (|\Delta'||\Delta''|)^{-2t} |c_\Delta(f)|^2 \\ &\leq C \max_{\Delta \notin \Lambda_\Psi} \frac{2^{2s\max(j_1,j_2)}}{2^{2t(j_1+j_2)}} \, \|f\|^2_{H^t_{\mathrm{mix}}} \, . \end{aligned}$$

Choosing the appropriate Ψ, the upper estimates of Theorem 4 follow. E.g., if Ψ coincides with the hyperbolic cross space $\mathbf{\Psi}^*_J(I^2)$ then

$$\max_{\Delta \notin \Lambda_{\mathbf{\Psi}^*_J(I^2)}} \frac{2^{2s\max(j_1,j_2)}}{2^{2t(j_1+j_2)}} = \max_{j_1+j_2 > J} \frac{2^{2s\max(j_1,j_2)}}{2^{2t(j_1+j_2)}} \asymp \begin{cases} 2^{-2(t-s)J} , & s \geq 0 , \\ 2^{-2(t-s/2)J} , & s < 0 . \end{cases}$$

This gives (24), and the case $s = 0$ of (25) since $\#\mathbf{\Psi}^*_J(I^2) \asymp J2^J$, $J \to \infty$.

For the remaining cases in (25), set

$$\Psi^{s,*}_J \equiv \begin{cases} \cup_{j_1,j_2 \geq 0 \, ; \, j_1+j_2-\mu\max(j_1,j_2) \leq (1-\mu)J} \Psi_{j_1,j_2} , & s > 0 , \\ \cup_{j_1,j_2 \geq 0 \, ; \, j_1+j_2+2\mu\max(j_1,j_2) \leq (1+\mu)J} \Psi_{j_1,j_2} , & s < 0 , \end{cases}$$

where $\mu > 0$ will be fixed below. Note that for any fixed $\mu > 0$, in both cases $s > 0$ and $s < 0$, we have $\#\Psi^{s,*}_J \asymp 2^J$. Let $t > s > 0$. By symmetry, we have

$$\max_{\Delta \notin \Lambda_{\Psi^{s,*}_J}} \frac{2^{2s\max(j_1,j_2)}}{2^{2t(j_1+j_2)}} = \max_{j_1 \geq j_2 \geq 0 \, : \, j_1+j_2-\mu j_1 > (1-\mu)J} 2^{2sj_1 - 2t(j_1+j_2)} \, .$$

If we now require $0 < \mu < s/t$ we immediately see that, asymptotically, this maximal value is attained near $j_2 = 0, j_1 = J+1$, and satisfies $\asymp 2^{2J(s-t)}$. Since, $\#\Psi_J^{s,*} \asymp 2^J$, this gives (25) for $0 < s < t < 1$ and $N \asymp 2^J$.

Similarly, for $s < 0$, $s/2 < t < 1$, we have

$$\max_{\Delta \notin \Lambda_{\Psi_J^{s,*}}} \frac{2^{2s\max(j_1,j_2)}}{2^{2t(j_1+j_2)}} = \max_{j_1 \geq j_2 \geq 0\,:\, j_1+j_2+2\mu j_1 > (1+\mu)J} 2^{2sj_1 - 2t(j_1+j_2)} .$$

Choosing now $0 < \mu < -s/(2t)$ if $t > 0$ (if $s/2 < t \leq 0$ then any $\mu > 0$ will do), it can be shown that the maximal value satisfies $\asymp 2^{2J(s/2-t)}$ and is attained near $j_1 = j_2 = [J/2]+1$. This concludes the proof of Theorem 4 in the case $t < 1$.

The changes for $t = 1$ are simple. We shall rely on (20) as a substitute for (16). Then, as long as the considered Ψ are of the form $\Psi = \cup_{(j_1,j_2)\in\mathcal{J}} \Phi^*_{j_1,j_2}$ for some finite index set $\mathcal{J}$, we obtain

$$e_\Psi(f)_s^2 \leq C\Big(\sum_{(j_1,j_2)\notin\mathcal{J}} \frac{2^{2s\max(j_1,j_2)}}{2^{2(j_1+j_2)}}\Big)\|f\|^2_{H^1_{\mathrm{mix}}} .$$

After this, computing the occcuring sum for the same choices of Ψ as considered before (instead of evaluating the above maxima), we arrive at the correct result. In particular, for $s = 0$, $\Psi = \boldsymbol{\Psi}^*_J(I^2)$, we get

$$e^*_{\#\boldsymbol{\Psi}^*_J(I^2)}(f)_0 \leq C\Big(\sum_{j_1+j_2>J} 2^{-2(j_1+j_2)}\Big)\|f\|^2_{H^1_{\mathrm{mix}}} \leq CJ2^{-2J}\|f\|^2_{H^1_{\mathrm{mix}}} ,$$

which, due to $\#\boldsymbol{\Psi}^*_J(I^2) \asymp J2^J$, leads to the exponent $3/2$ for the logarithmical term if $t = 1$.

We will provide several examples which show that the estimates in Theorem 4 are sharp, at least for certain parameter cases. Clearly, we have some difficulties if $s \leq -1/2$ resp. $t \geq 1/2$ since for these parameter ranges the used inequalities (16), (17) cannot be reversed. Let us start with considering the C^∞-function $f(x_1,x_2) = (1+x_1)(1+x_2)$, $x = (x_1,x_2) \in I^2$ which covers the case $t = 1$, $-1/2 < s < 1/2$. For this f, the Haar-Fourier coefficients can be computed as

$$|c_\Delta(f)| = \sqrt{2}\cdot 2^{-3/2(j_1+j_2)} \qquad \forall \Delta \in \mathcal{D}_{j_1,j_2}\ ,\ \ j_1,j_2 \geq 1 .$$

When replacing equality by $\asymp$, this relationship also holds if $j_1 = 0$ or $j_2 = 0$. Thus, for large J, we have

$$N_J(f) \equiv \#\{\Delta\,:\, 2^{-3J} \leq |c_\Delta(f)|^2 \leq 32^{-3J}\} \asymp J2^J .$$

Thus, for $s = 0$ we conclude that for $N = [N_J(f)/2] \asymp J2^J$ we get

$$e^*_N(f)_0^2 \geq N_J(f)2^{-3J-1} \geq cJ2^{-2J} \geq cN^{-2}(\log N)^3 .$$

In the first step of the above estimation we have used (16) for $s = 0$, together with the fact that, whatever N-term approximation v is selected, there are still $\geq N_J(f) - N \geq N_J(f)/2$ coefficients of $f - v$ left such that

$$|c_\Delta(f-v)|^2 = |c_\Delta(f)|^2 \geq 2^{-3J} .$$

For $-1/2 < s < 0$, we take analogously $N = \#\Psi_{J,J}/2 \asymp 2^{2J}$, and use (16):

$$e^*_N(f)_s^2 \geq cN2^{2Js}2^{-6J} \geq c2^{-4(1-s/2)J} \geq cN^{-2(1-s/2)} .$$

To see the first step, observe that at least N coefficients associated with $\Psi_{J,J}$ remain unchanged if f is replaced by $f - v$ where v is any N-term combination of Haar functions from $\boldsymbol{\Psi}^*(I^2)$. Finally, if $0 < s < 1/2$, consider the sequence $N = \#\Psi_{J,0}/2 \asymp 2^J$. Then, repeating the argument, we get

$$e^*_N(f)_s^2 \geq cN2^{2Js}2^{-3J} \geq c2^{-2(1-s)J} \geq cN^{-2(1-s)} .$$

Thus, the example $f(x) = (1 + x_1)(1 + x_2)$ confirms that the estimate (25) correctly reflects the asymptotic behavior of the best N-term approximations of functions from $H^1_{\mathrm{mix}}(I^2)$ in H^s-norms $-1/2 < s < 1/2$. In order to measure the behavior of best N-term approximation in the H^s-norm on a class X, we introduce the quantities

$$e^*_N(X)_s = \sup_{f \in X \,:\, \|f\|_X \le 1} e^*_N(f)_s \; .$$

Proposition 5 *We have*

$$e^*_N(H^1_{\mathrm{mix}})_s \asymp \begin{cases} N^{-(1-s)} \, , & 0 < s < 1/2 \, , \\ N^{-1}(\log N)^{3/2} \, , & s = 0 \, , \\ N^{-(1-s/2)} \, , & -1/2 < s < 0 \, , \end{cases} \qquad N \to \infty \; . \tag{26}$$

Due to the fact that the counterexample for the lower estimates is from $C^\infty(I^2)$, this gives also the correct saturation order *of N-term approximation with respect to the Haar system $\mathbf{\Psi}^*(I^2)$. The case $-1 < s \le 1/2$ remains open.*

Another case which can be dealt with easily is that of $1/2 < s \le \max(s, s/2) < t < 1/2$, where the counterexamples are simply given by Haar polynomials

$$f = f^s_J \equiv \begin{cases} v_{J,0} \, , & 0 < s < 1/2 \, , \\ \sum_{j=0}^J v_{J-j,j} \, , & s = 0 \, , \\ v_{J,J} \, , & -1/2 < s < 0 \, , \end{cases} \qquad v_{j_1,j_2} \equiv \sum_{\Delta \in \mathcal{D}_{j_1,j_2}} \psi_\Delta \; .$$

Using (17), (16) we see that

$$\|f^s_J\|^2_{H^t_{\mathrm{mix}}} \asymp \begin{cases} 2^{J(1+2t)} \, , & 0 < t < 1/2 \, , \\ J2^{J(1+2t)} \, , & t = 0 \, , \\ 2^{2J(1+2t)} \, , & -1/2 < t < 0 \, , \end{cases}$$

for any $-1/2 < t < 1/2$. As above, estimating with $N = [N^s_J/2]$, where

$$N^s_J \equiv \begin{cases} \#\mathbf{\Psi}_{J,0} \asymp 2^J \, , & 0 < s < 1/2 \, , \\ \#\mathbf{\Psi}_{J,0} + \ldots + \#\mathbf{\Psi}_{0,J} \asymp J2^J \, , & s = 0 \, , \\ \#\mathbf{\Psi}_{J,J} \asymp 2^{2J} \, , & -1/2 < s < 0 \, , \end{cases}$$

we obtain

$$e^*_N(f^s_J)^2_s \asymp \begin{cases} 2^{J(1+2s)} \asymp N^{-2(t-s)} \|f^s_J\|^2_{H^t_{\mathrm{mix}}} \, , & 0 < s < 1/2 \, , \\ J2^J \asymp N^{-2t}(\log N)^{2t} \|f^s_J\|^2_{H^t_{\mathrm{mix}}} \, , & s = 0 \, , \\ 2^{2J(1+s)} \asymp N^{-(t-s/2)} \|f^s_J\|^2_{H^t_{\mathrm{mix}}} \, , & -1/2 < s < 0 \, , \end{cases} \qquad (\, N = [N^s_J/2] \,) \; .$$

Altogether we have

Proposition 6 *We have*

$$e^*_N(H^t_{\mathrm{mix}})_s \asymp \begin{cases} N^{-(t-s)} \, , & 0 < s < t < 1/2 \, , \\ N^{-t}(\log N)^t \, , & s = 0 < t < 1/2 \, , \\ N^{-(t-s/2)} \, , & -1/2 < s < s/2 < t < 1/2 \, . \end{cases} \qquad N \to \infty \; . \tag{27}$$

We believe that Proposition 6 extends to the case $1/2 \le t < 1$ which would close the gap between (27) valid for $t < 1/2$, and (26) from Proposition 5 corresponding to $t = 1$.

The next result deals with best N-term approximation rates of singularity functions in H^s-norms with respect to the tensor-product Haar system $\mathbf{\Psi}^*(I^2)$. It shows that, despite their low regularity, such

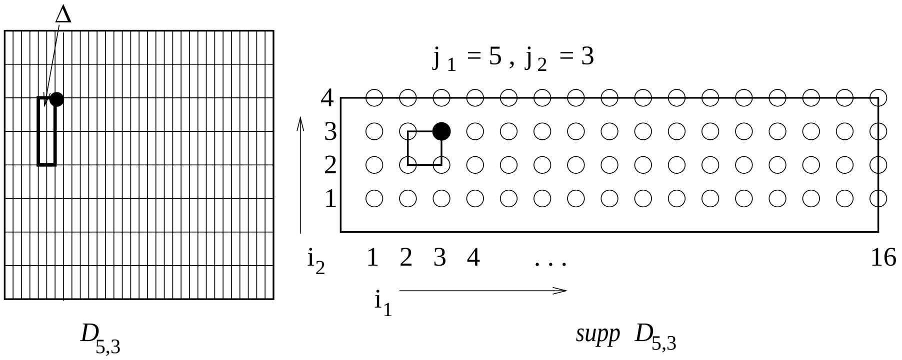

Figure 1: Notation: $\mathcal{D}_{j_1,j_2}$ and $\operatorname{supp}\mathcal{D}_{j_1,j_2}$

singularity functions can be approximated at the saturation rate, i.e., as well as very smooth f (compare Theorem 4 **b)** resp. (26) for $t = 1$). Thus, they are a particular instance of functions for which best N-term approximation procedures pay off. From now on, $m = 1$ is fixed. To make it easier to follow the elementary but rather technical details, we only consider the case of edge singularity functions of type $(1;\alpha,\alpha)$ with $0 \le \alpha < 1$. An analogous statement can be given for the case of point singularity functions (22) by similar reasoning.

Theorem 7 *Let f be an edge singularity function of type $(1;\alpha,\alpha)$ with respect to $(0,0)$, see (21). Let $-1 < s < 1/2$ be fixed. Then $f \in H^s(I^2)$ if $0 \le \alpha < \min(1/2 - s, 1/2 - s/2)$, and*

$$e_N^*(f)_s \le C_{\alpha,s} \begin{cases} N^{-(1-s)} , & 0 < s < 1/2 , \\ N^{-1}(\log N)^{3/2} , & s = 0 , \\ N^{-(1-s)} , & -1 < s < 0 , \end{cases} \qquad N \to \infty . \tag{28}$$

Proof. *Step 1.* Since $\alpha < 1$, we have $f \in L_1(I^2)$ (use the bound (21) for $k = l = 0$). Thus, the Haar-Fourier coefficients $c_\Delta(f)$ are properly defined, and satisfy the trivial estimate

$$|c_\Delta(f)| \le C 2^{-(j_1+j_2)(1/2-\alpha)}(i_1 i_2)^{-\alpha} , \tag{29}$$

where

$$\Delta \equiv [(i_1-1)2^{-(j_1-1)}, i_1 2^{-(j_1-1)}] \times [(i_2-1)2^{-(j_2-1)}, i_2 2^{-(j_2-1)}] \in \mathcal{D}_{j_1,j_2}; ,$$

for $i_l = 1,\ldots,2^{j_l-1}$, $j_l \ge 0$ $l = 1,2$ (if $j_l = 0$ then $i_l = 1$ is required). The notation connects a dyadic rectangle $\Delta \in \mathcal{D}^*(I^2)$ with the integers i_l, j_l and will be used below without further explanation. Note that $(i_1 2^{-(j_1-1)}, i_2 2^{-(j_2-1)})$ is the upper-right vertex of the support of the tensor-product Haar function $\psi_\Delta \in \Psi_{j_1,j_2}$ (with obvious modifications if $j_1 = 0$ or $j_2 = 0$). Below, the set of index pairs (i_1, i_2) corresponding to $\mathcal{D}_{j_1,j_2}$, will be denoted by $\operatorname{supp}\mathcal{D}_{j_1,j_2}$. Figure 1 illustrates this notation.

If $i_1 > 1$ resp. $i_2 > 1$ then better estimates than (29) can be obtained. E.g., if $i_1, i_2 > 1$ (which implicitely ensures $j_1, j_2 > 1$) then by definition of the Haar function ψ_Δ we have

$$\begin{aligned} & |c_\Delta(f)| \\ & \le (h_1h_2)^{-1/2} \int_{\Delta'} |f(x_1,x_2) - f(x_1+h_1/2,x_2) - f(x_1,x_2+h_2/2) + f(x_1+h_1/2,x_2+h_2/2)|\, dx_1 dx_2 \\ & \le \tfrac{1}{16}(h_1h_2)^{3/2}\|D^{(1,1)}f\|_{L_\infty(\Delta')} \le C h_1^{1/2-\alpha} h_2^{1/2-\beta} i_1^{-\alpha-1} i_2^{-\beta-1} . \end{aligned}$$

Here, $\Delta' = [(i_1-1)h_1, (i_1-1/2)h_1] \times [(i_2-1)h_2, (i_2-1/2)h_2]$ and $h_l = 2^{-(j_l-1)}$, $l = 1,2$. If, e.g., $i_1 > 1$ but $i_2 = 1$ then an analogous estimation using the first-order differences $|f(x_1,x_2) - f(x_1+h_1/2,x_2)|$

and the bound for $D^{(1,0)}f$ from (21) gives the same result. If $i_1 = i_2 = 1$ then (29) will do. This completes the reasoning for

Lemma 8 *For any edge singularity function f of type $(1;\alpha,\alpha)$, and any dyadic rectangle $\Delta \in \mathcal{D}^*(I^2)$, the Haar-Fourier coefficient $c_\Delta(f)$ can be estimated by*

$$|c_\Delta(f)| \le C 2^{-(1/2-\alpha)(j_1+j_2)}(i_1 i_2)^{-(\alpha+1)} \ , \tag{30}$$

where i_l, j_l, $l = 1,2$, are the indices associated with Δ as explained above.

Step 2. From (17) in Proposition 3 and its modification (19) (without further mentioning, we will not differentiate between f and $\tilde{f}$ for $-1 < s < 0$) and the definition of $e_N^*(f)_s$ we obtain

$$e_N^*(f)_s^2 \le C \inf_{\Lambda\subset\mathcal{D}^*(I^2)\,:\,\#\Lambda\le N} \sum_{\Delta\in\mathcal{D}^*(I^2)\setminus\Lambda} d(\Delta)^{-2s}|c_\Delta(f)|^2 \ , \quad -1 < s < 1/2 \ . \tag{31}$$

The idea is to replace $|c_\Delta(f)|$ by the upper bound (30) from Lemma 8, and to carry out the minimization with respect to Λ afterwards (see step 3). We first consider $N = 0$ which is equivalent to estimating the H^s-norm of g (in this case, $\Lambda = \emptyset$ and the inf-expression is superflous). The corresponding bound in the right-hand side can be split into separate summations corresponding to $j_1 \ge j_2$ and to $j_1 < j_2$. Let us consider the first case. Then $d(\Delta) \approx 2^{-j_1}$. Carrying out the summation (first with respect to i_1, i_2 for each fixed index pair (j_1, j_2), then with respect to $j_1 \ge j_2$, and finally with respect to $j_2 \ge 0$) one obtains a finite bound if $1 - 2(\alpha + s) > 0$ and $1 - 2\alpha - s > 0$ at the same time. The consideration for the case $j_1 < j_2$ is analogous. Thus, for an edge singularity function f of type $(1;\alpha,\beta)$, the condition $0 \le \alpha < \min(1/2 - s, 1/2 - s/2)$ implies $f \in H^s(I^2)$ for the given range of s which gives the first statement of Theorem 7.

Step 3. To determine suitable Λ for optimally estimating $e_N^*(f)_s^2$ via (30), (31), the best one can do is to apply *thresholding* to the sequence $\{d(\Delta)^{-s}2^{-(1/2-\alpha)j_1}2^{-(1/2-\beta)j_2}i_1^{-(\alpha+1)}i_2^{-(\beta+1)}\}$. Given any $\delta > 0$, we will define

$$\Lambda^\delta = \sum_{j_1,j_2\ge 0} \Lambda^\delta_{j_1,j_2} \ , \tag{32}$$

and

$$E^\delta \equiv \sum_{j_1,j_2\ge 0} \underbrace{\Big(\sum_{\Delta\in\mathcal{D}_{j_1,j_2}\setminus\Lambda^\delta_{j_1,j_2}} d(\Delta)^{-2s}2^{-(1-2\alpha)(j_1+j_2)}(i_1 i_2)^{-2(\alpha+1)} \Big)}_{\equiv E^\delta_{j_1,j_2}} \ , \tag{33}$$

where

$$\Lambda^\delta_{j_1,j_2} = \{\Delta \in \mathcal{D}_{j_1,j_2} \,:\, d(\Delta)^{-s}2^{-(1/2-\alpha)(j_1+j_2)}(i_1 i_2)^{-(\alpha+1)} \ge \delta\} \ , \quad j_1, j_2 \ge 0 \ . \tag{34}$$

These definitions will allow us to give upper bounds for the number $N^\delta = \#\Lambda^\delta$ of different dyadic rectangles in Λ^δ and for the H^s-error indicator E^δ in terms of δ. In conjunction with (31), this will finally lead to the desired N-term approximation rates.

Figure 2 shows schematically the partition of the (j_1, j_2)-domain into 5 subregions for which the condition in (34) leads to qualitatively the same set

$$\operatorname{supp}\Lambda^\delta_{j_1,j_2} = \{(i_1,i_2) \,:\, \Delta \in \Lambda^\delta_{j_1,j_2}\} \ .$$

Figure 3 illustrates the size of these index sets. The letter N indicates the part of $\operatorname{supp}\mathcal{D}_{j_1,j_2}$ that corresponds to $\operatorname{supp}\Lambda^\delta_{j_1,j_2}$ and contributes to the number of selected terms $N^\delta_{j_1,j_2} = \#\Lambda^\delta_{j_1,j_2}$, while the complement set denoted by E is responsible for the error part $E^\delta_{j_1,j_2}$. Although the picture in Figure 2 corresponds to $s = -1/2$, $\alpha = 1/2$, it qualitatively remains the same for other parameter choices. The axes in Figure 2 are scaled by a factor $\log 1/\delta$.

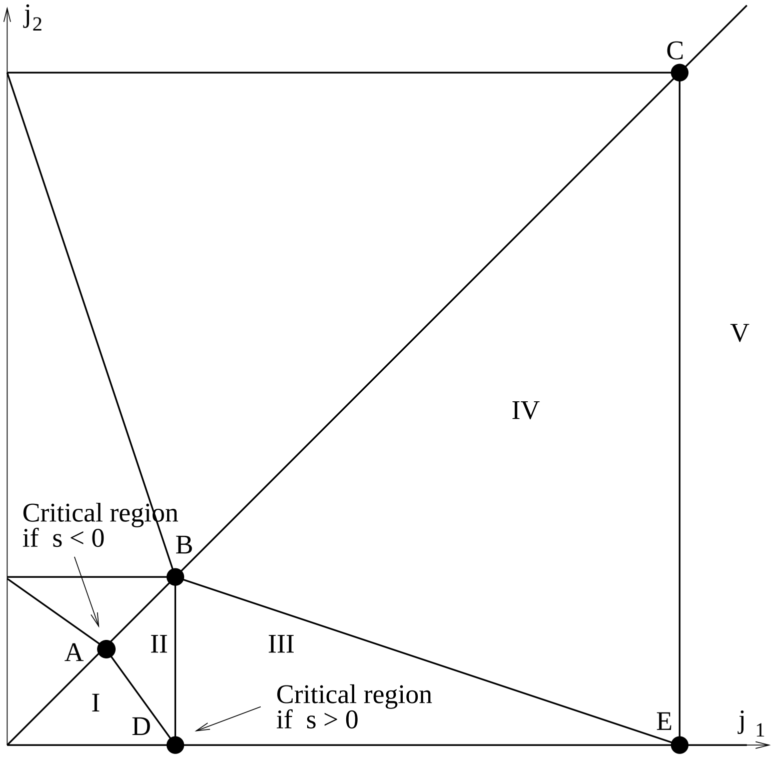

Figure 2: Partitioning of (j_1, j_2) domain

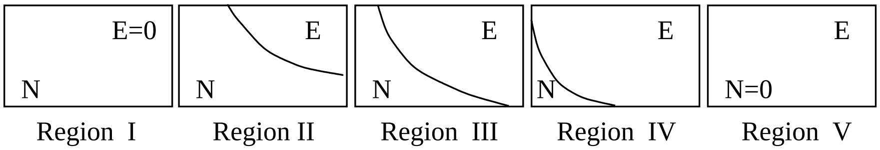

Figure 3: Qualitative form of the index sets supp $\Lambda^{\delta}_{j_1,j_2}$

It is clear from these pictures that the estimation procedure for the quantities $E^\delta_{j_1,j_2}$ and $N^\delta_{j_1,j_2} = \#\Lambda^\delta_{j_1,j_2}$ is the same if (j_1, j_2) belong to the same region, and that due to symmetry only the case $j_1 \geq j_2$ needs to be investigated. For future reference, let us give a list of the values for (j_1, j_2) corresponding to the vertex points of the partition in Figure 2:

$$\begin{array}{lcl} A & : & j_1 = j_2 \approx (3-s)^{-1} \log_2 1/\delta \ , \\ B & : & j_1 = j_2 \approx (2-s-\alpha)^{-1} \log_2 1/\delta \ , \\ C & : & j_1 = j_2 \approx (1-s-2\alpha)^{-1} \log_2 1/\delta \ , \\ D & : & j_1 \approx (3/2-s)^{-1} \log_2 1/\delta \ , \ \ j_2 = 0 \ , \\ E & : & j_1 \approx (1/2-s-\alpha)^{-1} \log_2 1/\delta \ , \ \ j_2 = 0 \ . \end{array}$$

For obvious reasons, we expect the *extremal values* of $E^\delta_{j_1,j_2}$ and $N^\delta_{j_1,j_2}$ to occur near one of these points, it will turn out that this will be true near A if $s < 0$ and D if $s > 0$. The L_2-case ($s = 0$) is in some sense exceptional, and more delicate to handle.

We start with the easy cases. Consider region I. Here $E^\delta_{j_1,j_2} = 0$ and $N^\delta_{j_1,j_2} \approx 2^{j_1+j_2}$. Summing with respect to all (j_1, j_2) in region I, we get $E^\delta_{\mathrm{I}} = 0$ and

$$N^\delta_{\mathrm{I}} \approx \left\{ \begin{array}{lll} N^\delta_A & , & -1 < s < 0 \\ N^\delta_A \log_2 1/\delta & , & s = 0 \\ N^\delta_D & , & 0 < s < 1/2 \end{array} \right\} \approx \left\{ \begin{array}{lll} \delta^{-2/(3-s)} & , & -1 < s < 0 \\ \delta^{-2/3} \log_2 1/\delta & , & s = 0 \\ \delta^{-2/(3-2s)} & , & 0 < s < 1/2 \end{array} \right\} . \tag{35}$$

Analogously, in region V we have $N^\delta_{j_1,j_2} = 0$ and

$$E^\delta_{j_1,j_2} \leq C 2^{2sj_1} 2^{-(1-2\alpha)(j_1+j_2)} \sum_{i_1=1}^{2^{j_1-1}} \sum_{i_2=1}^{2^{j_2-1}} (i_1 i_2)^{-2(\alpha+1)} \leq C 2^{2sj_1} 2^{-(1-2\alpha)(j_1+j_2)} \ .$$

Again, carrying out the summation, we obtain

$$N^\delta_{\mathrm{V}} = 0 \ , \qquad E^\delta_{\mathrm{V}} \approx E^\delta_C \log_2 1/\delta \approx \delta^2 \log_2 1/\delta \ . \tag{36}$$

The latter estimate holds whenever the condition $\alpha < \min(1/2-s, 1/2-s/2)$ (which was sufficient for $f \in H^s(I^2)$ to hold, see step 2) is met.

The estimations are a bit more delicate for the intermediate regions II-IV. We will carry out the details for regions IV and III, and state the result for region II (since the considerations are similar in all three cases, the reader will be able to recover the details). Let us start with region II. Recall from (34) that $(i_1, i_2) \in \operatorname{supp} \Lambda^\delta_{j_1,j_2}$ is equivalent to

$$1 \leq i_1 \leq i_2^{-1} \left(\delta^{-1} 2^{-j_1(1/2-(\alpha+s))} 2^{-j_2(1/2-\alpha)} \right)^{\frac{1}{1+\alpha}} \equiv \frac{\kappa}{i_2} \ ,$$

where $1 \leq i_2 \leq \kappa$. Since we are in region IV, we should have $\kappa \in [1, 2^{j_2-1}]$. Thus,

$$N^\delta_{j_1,j_2} = \sum_{i_2=1}^{[\kappa]} \frac{\kappa}{i_2} \leq C \kappa \log_2(\kappa) \ ,$$

and

$$\begin{array}{rcl} E^\delta_{j_1,j_2} & \leq & C 2^{-2j_1(1/2-(s+\alpha))-2j_2(1/2-\alpha)} \left(\displaystyle\sum_{i_2=1}^{[\kappa]} \sum_{i_1=[\frac{\kappa}{i_2}]+1}^{2^{j_1-1}} + \sum_{i_2=[\kappa]+1}^{2^{j_2-1}} \sum_{i_1=1}^{2^{j_1-1}} \right) (i_1 i_2)^{-2(1+\alpha)} \\ & \leq & C 2^{-2j_1(1/2-(s+\alpha))-2j_2(1/2-\alpha)} \left(\displaystyle\sum_{i_2=1}^{[\kappa]} \frac{\kappa^{-2(1+\alpha)+1}}{i_2} + \kappa^{-2(1+\alpha)+1} \right) \\ & \leq & C \delta^2 \displaystyle\sum_{i_2=1}^{[\kappa]} \frac{\kappa^{-2(1+\alpha)+1}}{i_2} = C \delta^2 N^\delta_{j_1,j_2} \end{array}$$

for all (j_1, j_2) in region IV. The relation

$$E^{\delta}_{j_1,j_2} \le C\delta^2 N^{\delta}_{j_1,j_2} \tag{37}$$

can be proved in exactly the same way for regions II and III.

Now we need to carry out the summation with respect to (j_1, j_2) in the triangular region IV. Since $1/2 - (s+\alpha) > 0$ by assumption, we will first sum with respect to j_1 for any fixed j_2. This gives

$$N^{\delta}_{\mathrm{IV}} \le C \sum_{(j_1,j_2)\in\overline{EB}\cup\overline{BC}} \kappa \log_2(\kappa)\ .$$

Since $2^{-j_1(1/2-(s+\alpha))} \approx \delta 2^{3j_2/2}$ for $(j_1, j_2) \in \overline{EB}$ (by this we mean the integer points closest to the right from EB), and $2^{-j_1} = 2^{-j_2}$ for $(j_1, j_2) \in \overline{BC}$, we have

$$\kappa^{1+\alpha} = \delta^{-1}(\delta 2^{3j_2/2})2^{-j_2(1/2-\alpha)} = 2^{j_2(1+\alpha)}(\delta 2^{j_2 s})^{-(1+\alpha)/(3/2-s)}$$

for $(j_1, j_2) \in \overline{EB}$, and

$$\kappa^{1+\alpha} = \delta^{-1} 2^{-j_2(1-s-2\alpha)}\ , \qquad (j_1, j_2) \in \overline{BC}\ .$$

By assumption, $1 - s - 2\alpha > 0$, and summation with respect to $\overline{BC}$ yields

$$\sum_{(j_1,j_2)\in\overline{BC}} \kappa \log_2(\kappa) \le C(\kappa \log_2(\kappa))|_B \le C(j_2 2^{j_2})|_B\ .$$

Analogously,

$$\sum_{(j_1,j_2)\in\overline{EB}} \kappa \log_2(\kappa) = \sum_{(j_1,j_2)\in\overline{EB}} j_2 2^{j_2} \le C(j_2 2^{j_2})|_B\ .$$

But at B we have $2^{j_2} \approx \delta^{-1/(2-(s+\alpha))}$ which leads to

$$N^{\delta}_{\mathrm{IV}} \le C\delta^{-1/(2-s-\alpha)}|\log_2 \delta|\ . \tag{38}$$

The estimate for $E^{\delta}_{j_1,j_2}$ follows from (37) and (38).

Next, consider region III. Using the same notation for κ, we now have $1 \le i_1 \le \kappa/i_2$, $i_2 = 1, \ldots, 2^{j_2-1}$, as description for $\operatorname{supp} \Lambda^{\delta}_{j_1,j_2}$ (by definition of region III, we have $1 \le \kappa/i_2 \le 2^{j_1-1}$ for all those i_2). This gives

$$N^{\delta}_{j_1,j_2} = \sum_{i_2=1}^{2^{j_2-1}} \frac{\kappa}{i_2} \le C\kappa j_2\ ,$$

and, as before,

$$N^{\delta}_{\mathrm{III}} \le C \sum_{(j_1,j_2)\in\overline{CB}} \kappa j_2\ .$$

Along $\overline{CB}$ we have

$$2^{-j_1(1/2-(s+\alpha))} \approx (\delta 2^{j_2(1/2-\alpha)})^{(1/2-(s+\alpha))/(3/2-s)}\ ,$$

which results in

$$\kappa^{1+\alpha} \approx (\delta 2^{j_2(1/2-\alpha)})^{(1/2-(s+\alpha))/(3/2-s)-1} = (\delta 2^{j_2(1/2-\alpha)})^{-(1+\alpha)/(3/2-s)}$$

for $(j_1, j_2) \in \overline{CB}$. Summation leads to

$$N^{\delta}_{\mathrm{III}} \le C \begin{cases} \delta^{-1/(3/2-s)} & ,\quad \alpha < 1/2\ , \\ \delta^{-2/3}|\log_2 \delta|^2 & ,\quad \alpha = 1/2\ , \\ \delta^{-1/(2-(s+\alpha))} & ,\quad \alpha > 1/2\ . \end{cases} \tag{39}$$

Here, it was used that $2^{-j_2} \approx \delta^{-1/(2-(s+\alpha))}$ at B. The case $\alpha \geq 1/2$ is only possible for $-1 < s < 0$.

Finally, in region II the set $\operatorname{supp} \Lambda^\delta_{j_1,j_2}$ is given by $1 \leq i_1 \leq 2^{j_1-1}$ if $1 \leq i_2 \leq i_2^a st$ and $1 \leq i_1 \leq \kappa/i_2$ if $j_2^a st < i_2 \leq 2^{j_2-1}$, where $i_2^* \approx \kappa 2^{-j_1}$. Thus,

$$N^\delta_{j_1,j_2} \leq C(\kappa + \sum_{i_2=i_2^*+1}^{2^{j_2-1}} \kappa/i_2) \leq C\kappa(1+\log_2(2^{j_2+j_1}/\kappa)) \ .$$

On the other hand,

$$E^\delta_{j_1,j_2} \leq C\delta^2 \sum_{i_2=i_2^*+1}^{2^{j_2-1}} \kappa/i_2 \leq C\delta^2\kappa \log_2(2^{j_2+j_1}/\kappa) \ .$$

Summation with respect to all (j_1,j_2) from region II yields

$$N^\delta_{\mathrm{II}} \leq C \sum_{(j_1,j_2)\in\overline{DA}\cup\overline{AB}} \kappa(1+\log_2(2^{j_2+j_1}/\kappa)) \ .$$

Since $2^{-j_1} \approx (\delta 2^{3j_2/2})^{1/(3/2-s)}$ along DA, and $j_1 = j_2$ along AB, we obtain

$$\kappa^{1+\alpha} \approx \delta^{-1}(\delta 2^{3j_2/2})^{(1/2-(s+\alpha))/(3/2-s)} 2^{-j_2(1/2-\alpha)} = (\delta 2^{j_2 s})^{-(1+\alpha)/(3/2-s)}$$

for $(j_1,j_2) \in \overline{DA}$, and

$$\kappa^{\alpha+1} = \delta^{-1} 2^{-j_2(1-s-2\alpha)} \ , \quad (j_1,j_2) \in \overline{AB} \ .$$

As for region IV, the latter relation allows us to conclude that the dominant part is the summation with respect to $\overline{DA}$. Examining the cases $s > 0$, $s = 0$, and $s < 0$ separately, we arrive at

$$N^\delta_{\mathrm{II}} \leq C \left\{ \begin{array}{lcl} \delta^{-1/(3/2-s)} & , & s > 0 \ , \\ \delta^{-2/3}|\log_2\delta| & , & s = 0 \ , \\ \delta^{-2/(3-s)} & , & s < 0 \ . \end{array} \right. \tag{40}$$

The estimate for E^δ_{II} leads to the same bounds multiplied by δ^2:

$$E^\delta_{\mathrm{II}} \leq C\delta^2 \left\{ \begin{array}{lcl} \delta^{-1/(3/2-s)} & , & s > 0 \ , \\ \delta^{-2/3}|\log_2\delta| & , & s = 0 \ , \\ \delta^{-2/(3-s)} & , & s < 0 \ . \end{array} \right. \tag{41}$$

In conclusion, we see from (35)-(41) that

$$N^\delta \leq C \left\{ \begin{array}{lcl} \delta^{-1/(3/2-s)} & , & s > 0 \ , \\ \delta^{-2/3}|\log_2\delta| & , & s = 0 \ , \\ \delta^{-2/(3-s)} & , & s < 0 \ , \end{array} \right. \qquad E^\delta \leq C\delta^2 \left\{ \begin{array}{lcl} \delta^{-1/(3/2-s)} & , & s > 0 \ , \\ \delta^{-2/3}|\log_2\delta| & , & s = 0 \ , \\ \delta^{-2/(3-s)} & , & s < 0 \ . \end{array} \right.$$

Eliminating $\delta > 0$ from these relationships and using (33) in connection with the definition of $e^*_N(g)_s$, we obtain

$$\begin{array}{rcl} e^*_{N^\delta}(g)_s & \leq & (E^\delta)^{1/2} \leq (E^\delta_{\mathrm{II}})^{1/2} \leq C \left\{ \begin{array}{lcl} \delta^{1-1/(3-2s)} & , & 0 < s < 1/2 \ , \\ \delta^{2/3}|\log_2\delta|^{1/2} & , & s = 0 \ , \\ \delta^{1-1/(3-s)} & , & -1 < s < 0 \ , \end{array} \right. \\ & \leq & C \left\{ \begin{array}{lcl} (N^\delta)^{-(1-s)} & , & 0 < s < 1/2 \ , \\ (N^\delta)^{-1}(\log N^\delta)^{3/2} & , & s = 0 \ , \\ (N^\delta)^{-(1-s/2)} & , & -1 < s < 0 \ . \end{array} \right. \end{array}$$

Finally, for given N, we can choose δ such that $N^\delta \approx N$ which establishes (28). Theorem 7 is proved.

We will now specialize to $s = -1/2$ and to solutions f of (4) for which the decomposition (23) holds. Since f^{sing} decomposes into edge singularities of type $(1; 1-\gamma, 1-\gamma)$ and $1-\gamma = 0.7034... < 3/4 = \min(1/2 - s, 1/2 - s/2)$, Theorem 7 is applicable and yields

$$e^*_N(f^{sing})_{-1/2} \le CN^{-5/4}\ , \qquad N \to \infty\ .$$

The regular part $f^{reg} \in H^1_{\text{mix}}(I^2)$ is covered by Theorem 4:

$$e^*_N(f^{reg})_{-1/2} \le CN^{-5/4}\ , \qquad N \to \infty\ .$$

Altogether, since $e^*_{m+n}(f_1 + f_2)_s \le e^*_m(f_1)_s + e^*_n(f_2)_s$, we arrive at

$$e^*_N(f)_{-1/2} \le CN^{-5/4}\ , \qquad N \to \infty\ . \tag{42}$$

for f satisfying (23). As was mentioned in section 2.3, this assumption (and therefore the estimate (42)) holds if the right-hand side g in (4) is sufficiently smooth (e.g., $g \in H^3(I^2)$). In particular, it applies to the solution of the capacitance problem, where $g(x) \equiv 1$. The method of proof also allows to find explicit subsets $\Psi \subset \mathbf{\Psi}^*(I^2)$ which realize the asymptotic rate (42). This is important for the numerical applications but needs careful verification in practice.

More generally, Theorem 4 and 7 imply the following result which might be of practical interest not only for $s = -1/2$ but also for $s = 0$.

Corollary 9 *Let $-1 < s < 1/2$, and $f = f^{reg} + f^{sing} \in H^s(I^2)$ be such that $f^{reg} \in H^1_{\text{mix}}(I^2)$ and f^{sing} is a linear combination of edge singularity functions of type $(1; \alpha, \alpha)$ with respect to the vertices of I^2 for some $0 \le \alpha < \min(1/2 - s, 1/2 - s/2)$. Then*

$$e^*_N(f)_s \le C \begin{cases} N^{-(1-s)}\ , & 0 < s < 1/2\ , \\ N^{-1}(\log N)^{3/2}\ , & s = 0\ , \\ N^{-(1-s)}\ , & -1 < s < 0\ , \end{cases} \qquad N \to \infty\ . \tag{43}$$

We conclude with some remarks.

- The result of Theorem 7 should be compared with the best N-term approximation rates obtainable from the system $\Psi(I^2)$ defined in (1). This Haar system is the prototype of a *wavelet system*. Estimates of the best N-term approximations for such systems are covered by a well-developed theory and describe the potential of *adaptive wavelet methods* for elliptic operator equations (see [3, 2]). Let us look at a simple example. Consider the function $f(x) = x_1^{-\alpha}\hat{f}(x_2)\,,\ (x_1, x_2) \in I^2$, with $0 < \alpha < 1$ and $\hat{f} \in C^1(I)$. E.g., to see that the estimates below are sharp, set $\hat{f}(x_2) = 1 + x_2$. Obviously, such an f is an edge singularity function of type $(1; \alpha, \alpha)$. By repeating the elementary steps that led to Lemma 8, the Haar-Fourier coefficients of f associated with the Haar functions from $\Psi_j(I^2)$ can be estimated by

  $$|c_\Delta(f)| \le C 2^{-j(1-\alpha)} i^{-(1+\alpha)}\ , \qquad \Delta \in \mathcal{D}_{j,j}\ ,\ j \ge 1\ .$$

  Here $c_\Delta(f)$ stands for the coefficient $c_\psi(f)$ with respect to any of the three functions ψ from $\Psi_j(I^2)$ with support on the same dyadic square $\Delta = \Delta' \times \Delta''$ ($\Delta', \Delta'' \in \mathcal{D}_{j-1}$), where $\Delta' = [(i-1)2^{j-1}, i2^{j-1}]$ determines i. The estimate is sharp for the functions ψ from $\Psi_j \otimes \Phi_{j-1}$.

  As in section 2.2, a bound

  $$\|f\|^2_{H^s} \le C \sum_{j=0}^{\infty} \sum_{\psi \in \Psi_j(I^2)} 2^{2js} |c_\psi(f)|^2\ , \qquad -1 < s < 1/2\ ,$$

  similar to (19) can be established for any $f \in L_1(I^2)$ for which the right-hand side of the last inequality is finite. This allows us to apply the same thresholding procedure and to find upper estimates of $e_N(f)_s$. Without repeating the details, we give the result for a particular parameter range: If $\max(0,-s/2) < \alpha < \min(1/2-s/2, 1/2-s)$, $-1 < s < 1/2$, then

$$e_N(f)_s \le C_f N^{-(1/2-(\alpha+s))} , \quad N \to \infty . \tag{44}$$

This estimate is again sharp, at least for $-1/2 < s < 1/2$. The comparison of (44) and (43) reveals that for functions with *strong edge singularities* best N-term approximation with respect to $\mathbf{\Psi}^*$ should be preferred over best N-term approximation with respect to $\mathbf{\Psi}(I^2)$. E.g., if we take the parameters $s = -1/2, \alpha = 1/2$, applicable in the case of the singular parts f^{sing} of solutions to (4), see section 2.3, then the estimate leads only to $e_N(f)_{-1/2} = \mathrm{O}(N^{-1/2})$ instead of $e_N^*(f)_{-1/2} = \mathrm{O}(N^{-5/4})$ as covered by (42).

- Generalizations of the above results to spline systems of arbitrary order m and to similar wavelet systems on I^d, $d \ge 2$, are only a technical matter. What is easy to do is to use tensor-product systems obtained from univariate *biorthogonal wavelet systems.* Then the estimation of wavelet coefficients

$$c_\Delta(f) = (f, \tilde{\psi}_\Delta)_{L_2}$$

can be based on the local support and zero-moment properties of the dual system $\{\tilde{\psi}_\Delta\}$ instead of using the special nature of the orthonormal Haar system. The (s,t)-region and s-intervals for which analogs of Theorem 4 and 7 hold will depend on the regularity and order of polynomial reproduction of the underlying biorthogonal multiresolution analysis (see [2] for a more complete treatment of all these questions). If one wishes to work with semi-orthogonal spline wavelets (as was done in [8] for $m = 2$), one needs to go through more technical details in connection with the global support of the associated $\tilde{\psi}_\Delta$.

- This note leaves some principal questions open. E.g., it remains an open problem to find a nontrivial description of *classes of functions characterized by a certain rate of best N-term approximation* with respect to systems such as $\mathbf{\Psi}^*(I^2)$ and to extend the results to the L_p-setting ($p \ne 2$).

References

[1] O. V. Besov, V. P. Il'yin, S. M. Nikolskij, Integral Representations of Functions and Embedding Theorems. I, II, Wiley, New York, 1979 (Russian edition: Nauka, Moscow, 1975).

[2] W. Dahmen, Wavelet and multiscale methods for operator equations, Acta Numerica 6, Cambridge Univ. Press, 1997, 55–228.

[3] R. A. DeVore, Nonlinear approximation, Acta Numerica 7, Cambridge Univ. Press, 1998, 51–150.

[4] R. A. DeVore, G. G. Lorentz, Constructive Approximation, Springer, New York, 1993.

[5] R. A. DeVore, S. V. Konyagin, V. N. Temlyakov, Hyperbolic wavelet approximation, Constr. Approx. 14 (1998), 1–26.

[6] R. A. DeVore, V. N. Temlyakov, Some remarks on greedy algorithms, Adv. in Comput. Math. (to appear).

[7] M. Griebel, S. Knapek, Optimized approximation spaces for operator equations, Preprint SFB 256, Universität Bonn, 1998.

[8] M. Griebel, P. Oswald, T. Schiekofer, Sparse grids for boundary integral equations, Numer. Math. (1998) (to appear).

[9] B. I. Golubov, Series in the Haar system, in Itogi Nauki i Tekniki: Mat. Anal. 1970, VINITI, Moscow, 1971, pp. 109–146 (English translation: J. Soviet Math. 1 (1973), no. 6).

[10] B. S. Kashin, A. A. Saakyan, Orthogonal Series, Transl. Math. Monogr., vol. 75, AMS, Providence, 1989 (Russian edition: Nauka, Moscow, 1984).

[11] J.-L. Lions, E. Magenes, Non-Homogeneous Boundary Value Problems and Applications, vol. 1, Springer, Berlin, 1972.

[12] P. Oswald, Multilevel Finite Element Approximation: Theory and Applications, Teubner, Stuttgart, 1994.

[13] P. Oswald, Multilevel norms for $H^{-1/2}$, Computing (1998) (to appear).

[14] T. von Petersdorff, E. P. Stephan, Decomposition in edge and corner singularities for the solution of the Dirichlet problem of the Laplacian in a polyhedron, Math. Nachr. 149 (1990), 71–104.

[15] T. von Petersdorff, E. P. Stephan, Regularity of mixed boundary value problems in $\mathbb{R}^3$ and boundary element methods on graded meshes, Math. Meth. Appl. Sci. 12 (1990), 229–249.

[16] E. P. Stephan, The h-p boundary element method for solving 2- and 3-dimensional problems, Preprint, University Hannover, 1995.

[17] V. N. Teml'yakov, Approximation of Periodic Functions, Nova Science Publ., New York, 1993.

[18] V. N. Teml'yakov, Nonlinear m-term approximation with regard to the multivariate Haar system, East J. Approx. 4 (1998), 87–106.

[19] H. Triebel, Interpolation Theory, Function Spaces, Differential Operators, Dt. Verlag Wiss., Berlin, 1978; North-Holland, Amsterdam, 1978 (Russian translation: Mir, Moscow, 1980).

[20] P. L. Ul'yanov, On Haar series, Mat. Sbornik 63(105) (1964), 356–391 (in Russian).

[21] J. Weidmann, Linear Operators in Hilbert Spaces, Springer, New York, 1980.

[22] C. Zenger, Sparse grids, in Parallel Algorithms for Partial Differential Equations, Proc. 6th GAMM Seminar, Kiel (W. Hackbusch, ed.), Vieweg, Braunschweig, 1991, pp. 241–251.