Which Spaces can be Embedded in Reproducing Kernel Hilbert Spaces?

Max Schölpple and Ingo Steinwart
University of Stuttgart
Faculty 8: Mathematics and Physics
Institute for Stochastics and Applications
Stuttgart, Germany
max.schoelpple@mathematik.uni-stuttgart.de
ingo.steinwart@mathematik.uni-stuttgart.de

June 26, 2025

Abstract

Given a Banach space E consisting of functions, we ask whether there exists a reproducing kernel Hilbert space H with bounded kernel such that $E \subset H$. More generally, we consider the question, whether for a given Banach space consisting of functions F with $E \subset F$, there exists an intermediate reproducing kernel Hilbert space $E \subset H \subset F$. We provide both sufficient and necessary conditions for this to hold. Moreover, we show that for typical classes of function spaces described by smoothness there is a strong dependence on the underlying dimension: the smoothness s required for the space E needs to grow *proportional* to the dimension d in order to allow for an intermediate reproducing kernel Hilbert space H.

Mathematical Subject Classification (2010). Primary 46E22; Secondary 46B20, 46E15, 46E35, 47A68, 68T05.

Key Words. Reproducing kernel Hilbert spaces, Inclusion maps, Spaces of functions, Factorization of Operators, Type and Cotype, Besov- and Triebel-Lizorkin spaces, Sobolev spaces with mixed smoothness, Neural networks.

1 Introduction

Reproducing kernel Hilbert spaces (RKHSs) are a powerful concept in various branches of mathematics and applications such as numerical analysis [42, 10], stochastics [7], signal processing [17] analysis [30, 2, 18, 3], and machine learning [31, 9, 39, 26]. In many of these applications, understanding RKHSs is central task and the choice of suitable RKHSs is essential in numerous algorithms.

Given an RKHS H on some underlying space X, properties of its kernel k or its elements $f \in H$ can often be described in terms of a surrounding space F, that is, by an inclusion $H \subset F$. For example, it is well-known that the kernel k is bounded, if and only if $H \subset \ell_\infty(X)$ holds, where $\ell_\infty(X)$ denotes the Banach space of all bounded functions $f : X \to \mathbb{R}$ equipped with the usual supremum norm. Bounded kernels do not only play a central role in the statistical analysis of kernel-based learning algorithms, e.g. [39], but they are also key for kernel mean embeddings, see e.g. [36, 40]. Moreover, the kernel is bounded and continuous, if and only if $H \subset C_b(X)$, where $C_b(X)$ denotes the space of all bounded, continuous functions $f : X \to \mathbb{R}$, and generalizations of such inclusions to higher notion of smoothness are, of course, possible.

In machine learning applications one is usually interested in sufficiently large RKHSs. Here, one way to describe the "size" of an RKHS H is by means of denseness in a surrounding space F. For example, a continuous kernel on some compact metric space X is said to be universal [37], if $H \subset C(X)$ is dense, where $C(X)$ denotes the space of all continuous functions $f : X \to \mathbb{R}$. Several papers have investigated universal kernels, see e.g. [22, 35, 32, 40] and the various references mentioned therein. In addition, recall that basic learning guarantees, such as universal consistency, can be established for kernel-based learning algorithms if universal kernels are used, see e.g. [39]. Moreover, universal kernels are very closely related to characteristic kernels, which play a central role for identifying probability distributions [36, 35], e.g. with the help of kernel-based non-parametric two-sample tests [16, 15].

Of course, the size of an RKHS H could also be described by specifying a "lower bound" on H, that is, a set, or vector space, of functions E it is supposed to contain. This is the focus of this paper. Namely, we consider the question:

Question. *Given two Banach spaces $E \subset F$ of functions $X \to \mathbb{R}$, does there exist an RKHS H on X with*

$$E \subset H \subset F\,? \tag{1}$$

Here, the Banach space of functions (BSF) F encodes additional properties an RKHS H with $E \subset H$ is supposed to satisfy. If no additional properties are needed, negative answers can thus be formulated by saying that (1) is impossible for all BSFs F with $E \subset F$. For example, if X is an uncountable, compact metric space such as $X = [0,1]^d$, then [38] has recently shown that for the BSF $E := C(X)$ Question (1) has a negative answer for all BSFs F. Besides [38], however, Question (1) has, to the best of our knowledge, not been considered in the literature, yet. Nonetheless, having positive or negative answers to this question may have various applications as we will discuss later in this introduction.

Let us now briefly summarize the results we obtain in this paper. Here, it seems fair to mention that for the greatest part of this paper we investigate Question (1) under the additional assumption that the point evaluations on E and F are continuous. This additional assumption is, however, at best a mild restriction as it is satisfied for practically all BSFs, and by definition, for all RKHSs. Moreover, it turns out that under this additional assumption the inclusion maps $\mathrm{id} : E \to H$ and $\mathrm{id} : H \to F$ are automatically continuous, see Lemma 23, which in turn opens the door for tools from functional analysis. Accordingly, we can freely switch between (1) and $E \hookrightarrow H \hookrightarrow F$, and in particular we will do so in this introduction.

Now, in the abstract setting of generic BSFs E and F, we show, among other results:

- Question (1) has a positive answer if and only if the inclusion map $\mathrm{id} : E \to F$ is 2-factorable, that is there exist an (abstract) Hilbert space H and bounded linear operators $U : E \to H$ and $V : H \to V$ with $\mathrm{id} = V \circ U$, such that we have

$$\begin{tikzcd} E \arrow[rr, hook, "\mathrm{id}"] \arrow[dr, "U"'] & & F \\ & H \arrow[ur, "V"'] & \end{tikzcd},$$

see Theorem 6. Of course, the inclusion (1) gives such a factorization, but the converse requires an explicit construction of a suitable RKHS.

- Question (1) has a positive answer if $\mathrm{id} : E \to F$ is 2-summing, see e.g. [11, Ch. 2] for a definition and key properties, or if E is of type 2 and F is of cotype 2, see Theorem 7. Here we note that these sufficient conditions can be derived from general results on 2-factorable operators, see e.g. [28] and [11].
- If Question (1) has a positive answer, then $\mathrm{id} : E \to F$ has both type 2 and cotype 2, see Corollary 9.

For concrete families of spaces, we provide negative and often also positive results in the following cases:

- For the spaces $E = \mathrm{Höl}_d^\alpha(X)$ of bounded, α-Hölder continuous functions on some metric space, where $\alpha \in (0, 1]$, we show in Theorem 12 that a positive answer to Question (1) with $F = \ell_\infty(X)$ implies a strong upper bound on the covering, or packing numbers of X. As a consequence, for the open, k-dimensional cube $X := (0, 1)^k$ and $\alpha < k/2$ no positive answer is possible, while for $\alpha \in (k/2, 1)$ we have a positive answer, see Theorem 22. In particular, for $k \geq 3$, none of the spaces $\mathrm{Höl}^\alpha((0, 1)^k)$ can be embedded into an RKHS, while for $k = 1$ such an embedding is possible for all $\alpha \in (1/2, 1)$ using a fractional Sobolev space $H = H_2^u((0, 1))$ for $u \in (1/2, \alpha)$. Moreover, in this case the inclusion $\mathrm{Höl}^1((0, 1)) \subset \mathrm{Höl}^{1-\varepsilon}((0, 1))$, which holds for all $0 < \varepsilon < 1/2$, shows that $\mathrm{Höl}^1((0, 1))$ can also be embedded into the RKHS $H_2^u((0, 1))$ with bounded kernel for any $u \in (1/2, 1)$.
- For Sobolev spaces E and F of integer (mixed) smoothness on some bounded domain $X \subset \mathbb{R}^d$ we show that if Question (1) has a positive answer, then the difference of the involved smoothness parameters needs to be sufficiently large, see Theorem 14 and Theorem 15. In the case of Sobolev spaces this difference needs to grow linearly with the dimension d, while for Sobolev spaces with mixed smoothness, this is not the case.
- For Sobolev-Slobodeckij spaces E and F on some bounded domain $X \subset \mathbb{R}^d$ we provide positive and negative results that match to each other modulo some limit cases. In a nutshell, positive results are possible if and only if the difference of the involved smoothness parameters is sufficiently large compared to d, see Theorem 20 for details. The same is true if we consider $F = \ell_\infty(X)$, instead.
- The results for Sobolev-Slobodeckij spaces can be extended to spaces E and F from the Besov or Triebel-Lizorkin family of spaces, see Theorem 21 and Theorem 22.

Here we note that our negative answers to Question (1) are always derived by showing that the inclusion map $\mathrm{id} : E \to F$ fails to have type 2 or cotype 2. On the other hand, we always construct positive results by employing well-known embedding theorems for the involved spaces. Moreover, in some cases our positive and negative results complement each other (modulo some limit cases), and therefore our type 2/cotype 2 technique is at least sometimes sharp. In these cases, we further see (modulo limit cases) that (1) has a positive answer, if and only if one already knows an RKHS H from the literature.

Let us finally illustrate the potential impact of positive and negative answers to Question (1). Here, our first example considers kernel-based learning algorithms, see e.g. [9, 39]. For such learning algorithms, standard (and fixed) RKHSs are sometimes viewed to be too small. To be specific, while the Gaussian kernel with width γ is known to be universal for every single value of γ, it is also known that from a quantitative perspective it only slowly approximates non-C^∞-functions, see e.g. [34]. For this reason one either lets γ depend on the sample (size), see e.g. [13], or considers e.g. so-called hyper-RKHS, see [25, 20]. Rather than employing such approaches, one could, however, also look for e.g. tailored RKHSs: Namely, if we have a space E of target functions for which we like to construct a kernel-based algorithm with "good" learning behaviour, an RKHS H with $E \subset H \subset F$ can be desirable. Indeed, $E \subset H$ ensures a good approximation error during the analysis of the resulting algorithm, while $H \subset F$ can provide a strong eigenvalue or entropy number decay, see e.g. [39, Lem. 7.21 and Thm. 6.26] for the cases $H \subset \ell_\infty(X)$ or $H \subset C^\alpha(X)$, which in turn leads to a small estimation error.

Here we note that recently, BSFs have attracted interest in the machine learning community since neural networks define particular forms of such spaces, see e.g. [6] and the references mentioned therein. In addition, [5, Sec. 2.3] compares the differences between neural network specific BSFs E and related (but unfortunately smaller) RKHSs H in some detail. We provide sufficient requirements for the corresponding integral reproducing kernel Banach spaces E to be included in an RKHS H in Section 4.4.

Another possible application are so-called integral probability metrics, see e.g. [23] and the references mentioned therein. Given a measurable space $(X, \mathcal{A})$ and a set $\mathcal{F}$ of measurable functions $X \to \mathbb{R}$, these metrics are defined as

$$\gamma_{\mathcal{F}}(P, Q) := \sup_{f \in \mathcal{F}} \Big| \int_X f \, dP - \int_X f \, dQ \Big|$$

for all probability measures P and Q on $(X, \mathcal{A})$ for which the integrals do exist. For example, if $\mathcal{F} = B_{\mathcal{L}_\infty(X)}$ is the unit ball in the space $\mathcal{L}_\infty(X)$ of all bounded measurable functions $X \to \mathbb{R}$ equipped with the supremum norm, then $\gamma_{\mathcal{F}}$ equals the metric obtained from the total variation norm. In addition, if (X, d) is a separable metric space equipped with the Borel σ-algebra and $\mathcal{F}$ is the unit ball in the space $\mathrm{H\ddot{o}l}_d^1(X)$ of all *bounded* Lipschitz functions, then $\gamma_{\mathcal{F}}$ is known as the *Dudley metric*, which metrizes the weak convergence of probability measures, see e.g. [12, Theorem 11.3.3]. Furthermore, if $\mathcal{F}$ is the set of all Lipschitz continuous functions with Lipschitz constant ≤ 1 and (X, d) is a bounded metric space, then the famous Kantorovich–Rubinstein theorem shows that $\gamma_{\mathcal{F}}$ equals the Wasserstein 1-distance, see e.g. [12, Theorem 11.8.2]. Finally, if $\mathcal{F}$ is the unit ball B_H of an RKHS H with bounded, measurable kernel, then $\gamma_{\mathcal{F}}$ is called H-maximum mean discrepancy (MMD), see [16]. Moreover, it can be shown that having a bounded and measurable kernel is also necessary for the MMD to be defined for all probability measures on $(X, \mathcal{A})$, see e.g. [36, Prop. 2] in combination with [39, Lem. 4.23]. Finally, recall that, unlike many other integral probability metrics, MMDs can be both expressed in closed form and estimated from data, see e.g. [15].

This raises the question of how powerful MMDs are compared to the general class of integral probability metrics. To discuss this question, let us fix an RKHS H with bounded measurable kernel. Then $\gamma_{B_H}(P, Q)$ exists for all probability measures P and Q on X and we have

$$\gamma_{B_H}(P, Q) \leq \| \mathrm{id} : H \to \mathcal{L}_\infty(X) \| \cdot \|P - Q\|_{\mathrm{TV}} \, .$$

Consequently, if an empirical estimate ensures that $\gamma_{B_H}(P, Q)$ is "large" with high probability, then P and Q are "rather distinct" in the sense of the total variation norm. However, if $\gamma_{B_H}(P, Q) > 0$ is "small", then we cannot guarantee that $\|P - Q\|_{\mathrm{TV}}$ is sufficiently small by the inequality above, see also [40] for a rigorous negative result in this direction. To overcome this issue, the current MMD literature focuses on *characteristic* kernels, that is, on kernels for which $\gamma_{B_H}(P, Q) = 0$ implies $P = Q$. To obtain a more quantitative result, assume that we have a BSF E with unit ball B_E and $E \hookrightarrow H$. Then the inequality above can be extended to

$$\| \mathrm{id} : E \to H \|^{-1} \cdot \gamma_{B_E}(P, Q) \leq \gamma_{B_H}(P, Q) \leq \| \mathrm{id} : H \to \mathcal{L}_\infty(X) \| \cdot \|P - Q\|_{\mathrm{TV}} \, .$$

Consequently, $\gamma_{B_H}(P, Q) > 0$ is "small", then P and Q are "similar" in the sense of the integral probability metric γ_{B_E}. Depending on the space E, the integral probability metric γ_{B_E} might have a clear and intuitive meaning, but it might be difficult or even infeasible to estimate γ_{B_E} from data. In this case, MMDs may still be helpful. Indeed, if we have RKHSs H_1 and H_2 with $H_1 \hookrightarrow E \hookrightarrow H_2 \hookrightarrow \mathcal{L}_\infty(X)$, then

$$\| \mathrm{id} : H_1 \to E \|^{-1} \cdot \gamma_{B_{H_1}}(P, Q) \leq \gamma_{B_E}(P, Q) \leq \| \mathrm{id} : E \to H_2 \| \cdot \gamma_{B_{H_2}}(P, Q)$$

holds for all probability measures P and Q on X. Consequently, if we know from data that $\gamma_{B_{H_1}}(P,Q)$ is large with high probability, then, with the same probability, P and Q are rather distinct in the sense of γ_{B_E}. Conversely, if we know from data that $\gamma_{B_{H_2}}(P,Q)$ is small with high probability, then with the same probability, P and Q are rather similar in the sense of γ_{B_E}. In other words, using H_1 and H_2 we may gain a two-sided *quantitative* control over γ_{B_E}.

In view of this discussion let us now illustrate the consequences of our findings for MMDs. To this end, let $X := (0,1)^k$. If $k = 1$, we have already noted above that we have the inclusions

$$\mathrm{H\ddot{o}l}^1((0,1)) \subset H_2^u((0,1)) \subset \mathcal{L}_\infty((0,1))$$

for all $u \in (1/2, 1)$. Consequently, we have MMDs that provide upper bounds for the Dudley metric, while RKHS H with sufficiently smooth kernel satisfy $H \subset \mathrm{H\ddot{o}l}^1((0,1))$, that is, they provide lower bounds for the Dudley metric. On the other hand, in the case $k \geq 3$, the well-established inclusions between Besov spaces do not provide a fractional Sobolev space between $\mathrm{H\ddot{o}l}^1((0,1)^k)$ and $\mathcal{L}_\infty((0,1)^k)$, and our results now show that there is actually no RKHS H with $\mathrm{H\ddot{o}l}^1((0,1)^k) \subset H \subset \mathcal{L}_\infty((0,1)^k)$. As a consequence, we do not obtain an MMD-based upper bound of the Dudley metric as soon as the dimension satisfies $k \geq 3$, while a lower bound is still possible by using sufficiently smooth kernels.

The examples on learning algorithms and integral probability metrics can be generalized. Namely, if we wish to design an algorithm that deals with functions $f \in E$ via their point evaluations, then having an RKHS H with $E \subset H$ could have significant algorithmic advantages: Indeed, the "kernel trick" [31] can simplify the computation of inner products in H, which in turn makes it possible to use inner products in the algorithm design by interpreting a function $f \in E$ as an element of H. Similarly, if no such H exists, then we know that no such algorithmic short-cut is possible.

The rest of this paper is organized as follows: In Section 2 we recall some notion and technical tools used in this paper. In addition, most of the notation is fixed. Section 3 presents all our results in the abstract setting and Section 4 contains all results related to concrete spaces. In addition, we discuss our general approach for deriving negative results for two simple families of spaces. Finally, all proofs can be found in Section 5.

2 Preliminaries

Given a number $s \in \mathbb{R}$, we write $\lfloor s \rfloor := \max\{k \in \mathbb{Z} \mid k \leq s\}$ and $(s)_+ := \max\{0, s\}, (s)_- := \max\{0, -s\}$.

In the following, E and F denote Banach spaces with norms $\|\cdot\|_E$ and $\|\cdot\|_F$ and closed unit balls B_E and B_F, respectively. Moreover, E' and F' denote their dual spaces. In the case $E \subset F$, we write $E \hookrightarrow F$, if the corresponding inclusion map is continuous.

In this work, we are mostly interested in Banach spaces consisting of functions. The following definition introduces these spaces formally.

Definition 1. *Let $X \neq \emptyset$. Then a Banach space E is called a Banach space of functions (BSF) on X, if all its elements are functions mapping from X to $\mathbb{R}$. Moreover, we say that E is a proper BSF, if for all $x \in X$, the evaluation functional*

$$\begin{aligned} \delta_x : E &\to \mathbb{R} \\ f &\mapsto f(x) \end{aligned}$$

is continuous. A proper BSF with a Hilbert space norm is called reproducing kernel Hilbert space (RKHS).

In the literature, Banach spaces of functions are also called *reproducing kernel Banach spaces*. These spaces have been recently gained interest for the analysis of neural networks, see e.g. [6].

The usual sequence spaces c_0 and ℓ_p for $p \in [1, \infty]$ are all proper BSFs. Moreover, the space of $\mathbb{R}$-valued, bounded continuous functions $C^0(X)$ on some metric space X is also a proper BSF, while the usual Lebesgue spaces on e.g. the unit interval, that is, $L_p([0, 1])$, fail to be a BSF. Finally, the space $\ell_\infty(X)$ of all bounded functions $f : X \to \mathbb{R}$ equipped with the supremum norm is also a proper BSF.

Note that in all proper BSFs E on X norm-convergence implies pointwise convergence, that is, if $(f_n) \subset E$ is a sequence converging to some $f \in E$ in the sense of $\|f_n - f\|_E \to 0$, then $f_n(x) \to f(x)$ for all $x \in X$.

RKHSs have been studied in great detail in the literature. For basic information we refer to e.g. [39, Chapter 4]. For reproducing kernel Hilbert spaces H on X, the corresponding (reproducing) kernel $k : X \times X \to \mathbb{R}$ can be defined by

$$k(x, x') \coloneqq \langle \delta_x, \delta_{x'} \rangle_{H'}\,, \qquad\qquad x, x' \in X.$$

It is well-known that various properties of the functions in H can be characterized by properties of k. For example, all functions $f \in H$ are bounded if and only if the kernel k is bounded, see e.g. [39, Lem. 4.23]. Consequently, if we have a proper BSF E and we seek a surrounding RKHS H with bounded kernel, we can express this equivalently as

$$E \subset H \subset \ell_\infty(X)\,. \tag{2}$$

Note that in this case, the inclusion maps are automatically continuous, see Lemma 23.

Many results of this paper rely on the type and cotype of operators and spaces. For this reason, let us quickly recall these notions. To this end, we fix a Banach space E and some $n \geq 1$. Then we define a norm on the n-fold product space E^n by

$$\|(x_1, \dots, x_n)\|_{\ell_2^n(E)} \coloneqq \Big(\sum_{i=1}^n \|x_i\|_E^2 \Big)^{1/2}\,, \qquad\qquad x_1, \dots, x_n \in E,$$

which we sometimes call the *sequence norm*. Moreover, let $\varepsilon = (\varepsilon_i)$ be a *Rademacher sequence*, that is, a sequence of i.i.d. random variables fulfilling $\mathbb{P}(\varepsilon_1 = 1) = \mathbb{P}(\varepsilon_1 = -1) = 1/2$. Then

$$\|(x_1, \dots, x_n)\|_{\mathrm{Rad}^n(E)} \coloneqq \mathbb{E}\Big\| \sum_{i=1}^n \varepsilon_i x_i \Big\|_E\,, \qquad\qquad x_1, \dots, x_n \in E \tag{3}$$

defines another norm on E^n, which we sometimes call the *Rademacher norm*. Here we note that in the literature one also considers p-norms on the right hand side, but due to Kahane's inequality, see e.g. [11, Thm. 11.1], the resulting norms are equivalent to (3) with constants independent of n.

Now, the basic idea of type and cotype is to compare sequence norms and Rademacher norms uniformly in n. Namely, for an operator $A \in \mathcal{L}(E, F)$ we define

$$\|A\|_{\mathrm{type}_2} \coloneqq \sup_{\substack{n \in \mathbb{N} \\ (x_1, \dots, x_n) \in E^n \setminus \{0\}}} \frac{\|(Ax_1, \dots, Ax_n)\|_{\mathrm{Rad}^n(F)}}{\|(x_1, \dots, x_n)\|_{\ell_2^n(E)}}. \tag{4}$$

and we say that A is of *type 2* if $\|A\|_{\mathrm{type}_2} < \infty$. Analogously, we write

$$\|A\|_{\mathrm{cotype}_2} \coloneqq \sup_{\substack{n \in \mathbb{N} \\ (x_1, \dots, x_n) \in E^n \setminus \{0\}}} \frac{\|(Ax_1, \dots, Ax_n)\|_{\ell_2^n(F)}}{\|(x_1, \dots, x_n)\|_{\mathrm{Rad}^n(E)}} \tag{5}$$

and we say that A is of *cotype 2*, if $\|A\|_{\mathrm{cotype}_2} < \infty$. It is straightforward to verify that $\|\cdot\|_{\mathrm{type}_2}$ defines a norm on the space of all type 2 operators $A \in \mathcal{L}(E, F)$ and an analogous statement is true for $\|\cdot\|_{\mathrm{cotype}_2}$ on the space of all cotype 2 operators $A \in \mathcal{L}(E, F)$. Moreover, if if we have Banach spaces E_0 and F_0, as well as bounded linear operator $S : E_0 \to E$ and $T : F \to F_0$, then some simple calculations show the following two inequalities, which describe an ideal property of type and cotype 2 operators:

$$\|TAS\|_{\mathrm{type}_2} \le \|T\|\|A\|_{\mathrm{type}_2}\|S\|\,, \tag{6}$$

$$\|TAS\|_{\mathrm{cotype}_2} \le \|T\|\|A\|_{\mathrm{cotype}_2}\|S\|\,. \tag{7}$$

Finally, we say that a Banach space E is of type 2 or of cotype 2, if the identity map $\mathrm{id}_E : E \to E$ is of type 2, respectively cotype 2. In this respect we also need to recall that Hilbert spaces H are of both type 2 and cotype 2. For more information on type and cotype we refer to [11, Chapter 11].

Let (X, d) be a metric space. Then, we denote the closed ball of radius δ with center $x \in X$ by $B(x, \delta)$. Moreover, given a $\delta > 0$, a sequence $x_1, \ldots, x_n \in X$ is called a δ-packing in X, if $d(x_i, x_j) \ge \delta$ holds for any $i \ne j$. The *packing numbers* of X are defined as

$$N^{\mathrm{pack}}(\delta, X) := N^{\mathrm{pack}}_d(\delta, X) := \sup\{n \in \mathbb{N} \mid \text{ there exists a } \delta\text{-packing } x_1, \ldots, x_n \in X\}$$

for all $\delta > 0$. It is well known that for bounded $X \subset \mathbb{R}^d$ with non-empty interior there exist constants $0 < c \le C$ such that

$$c\delta^{-d} \le N^{\mathrm{pack}}_d(X, \delta) \le C\delta^{-d}\,, \qquad\qquad \delta \in (0, 1]. \tag{8}$$

3 General Results

In this section we present several results investigating the situation $E \subset H \subset F$, where E and F are proper BSFs on X and H is an RKHS on X. In particular, we derive both sufficient and necessary conditions for the existence of such an RKHS H. We begin with a simple characterization in the case that E is a *closed subspace* of F. In a nutshell it shows that in this case, no work for finding an RKHS H is required.

Proposition 2. *Let F be a proper BSF on X and E be a closed subspace of F. Then E is a proper BSF and the following statements are equivalent:*

i) There exists an RKHS H with $E \subset H \subset F$.

ii) E is isomorphic to a Hilbert space.

Moreover, in this case we have $E \hookrightarrow H \hookrightarrow F$.

Since neither c_0 nor ℓ_∞ are isomorphic to a Hilbert space we immediately conclude by (2) and Proposition 2 with $F := \ell_\infty$ that there exists no RKHS H with bounded kernel such that $c_0 \subset H$ or $\ell_\infty \subset H$. Similarly, if (X, d) is a compact metric space such that $E := C(X)$ is infinite dimensional, then this space is not isomorphic to a Hilbert space, and by considering $F := \ell_\infty(X)$ in Proposition 2, we conclude that there is no RKHS H with bounded kernel and $C(X) \subset H$.

Given a proper BSF E on X and a non-empty subset $Y \subset X$, we can consider the restricted functions $f_{|Y} : Y \to \mathbb{R}$ for all $f \in E$. The set $E_{|Y}$ of such restrictions is again a BSF, see Lemma 24 for a formal statement. Moreover, we note that the restriction $H_{|Y}$ of an RKHS H on X with kernel k is again an RKHS.

Its kernel is given by restricting k to $Y \times Y$, that is by $k_{|Y\times Y}$. This is a direct consequence of the fundamental theorem of RKHS, see [39, Thm. 4.21] and [39, Ex. 4.4].

The following proposition shows that inclusions $E \subset F$ are preserved when applying the same restriction operator to both spaces E and F.

Proposition 3. *Let E and F be proper BSFs on X with $E \subset F$. Then for all non-empty $Y \subset X$ we have*

$$E_{|Y} \hookrightarrow F_{|Y}$$

with $\|\mathrm{id} : E_{|Y} \to F_{|Y}\| \leq \|\mathrm{id} : E \to F\|$.

To illustrate Proposition 3, we assume that we have two proper BSF E and F on X with $E \subset F$ and we seek an RKHS H with $E \subset H \subset F$. If there exists such an H, then Proposition 3 ensures

$$E_{|Y} \hookrightarrow H_{|Y} \hookrightarrow F_{|Y}\,, \tag{9}$$

and in addition, $H_{|Y}$ is an RKHS. In other words, if we find a non-empty $Y \subset X$, for which there is no RKHS $\tilde{H}$ on Y with $E_{|Y} \subset \tilde{H} \subset F_{|Y}$, then there is also no RKHS H on X with $E \subset H \subset F$. As a consequence, we only need to consider our Question (1) on subsets Y that we can suitably control.

The following definition is crucial for deriving both positive and negative answers to our Question (1).

Definition 4. *Let E, F be Banach spaces and $A : E \to F$ be a bounded linear operator. We say that A is* 2-factorable *if there exist a Hilbert space H and bounded linear operators $U : E \to H$ and $V : H \to F$ such the following* 2-factorization *of A holds:*

$$\begin{tikzcd} E \arrow[rr, "A"] \arrow[dr, "U"'] & & F \\ & H \arrow[ur, "V"'] & \end{tikzcd} . \tag{10}$$

Moreover, we define $\|A\|_{\gamma_2} := \inf \|U\|\|V\|$, where the infimum runs over all 2-factorizations of A.

It can be shown that $\|\cdot\|_{\gamma_2}$ defines an operator ideal norm, see [11, Thm. 7.1 in combination with p. 155] for details. Namely, if we have Banach spaces E_0 and F_0 and bounded linear operators $S : E_0 \to E$ and $T : F \to F_0$, and a 2-factorable operator $A : E \to F$, then $TAS : E_0 \to F_0$ is also 2-factorable and we have

$$\|TAS\|_{\gamma_2} \leq \|T\|\|A\|_{\gamma_2}\|S\|\,.$$

With this observation, we almost immediately obtain the following lemma, which will play a central role in our analysis.

Lemma 5. *Let E, F be Banach spaces and $A : E \to F$ be a bounded linear operator. If A is 2-factorable, then it is both of type 2 and cotype 2.*

Note that if we have BSFs E and F and an RKHS H with $E \hookrightarrow H \hookrightarrow F$, then the inclusion map $\mathrm{id} : E \to F$ is obviously 2-factorable. The following theorem shows that the converse implication is also true.

Theorem 6. *Let E and F be BFSs on some set X with $E \hookrightarrow F$. Then this inclusion map is 2-factorable if and only if there exists an RKHS H over X such that*

$$E \subset H \subset F\,.$$

In this case we also have $E \hookrightarrow H \hookrightarrow F$.

In the following, we will derive both sufficient and necessary conditions that are in many cases easier to check than an abstract 2-factorization. In particular, our sufficient conditions make it possible to avoid an explicit construction of a 2-factorization, while our necessary conditions can be used to show that such a construction is impossible. We begin with two sufficient conditions.

Theorem 7. *Let E and F be BFSs on some set X with $E \subset F$. Then there exists an RKHS H over X such that*

$$E \hookrightarrow H \hookrightarrow F$$

if one of the following two conditions are satisfied:

i) The inclusion map $\mathrm{id} : E \to F$ *is 2-summing.*

ii) E is of type 2 and F is of cotype 2.

In view of *i)* recall that 2-summing, or more generally, p-summing operators have been extensively investigated in the literature, see e.g. [11]. In particular, is is known that for certain pairs of Banach spaces E and F, every bounded linear operator $A : E \to F$ is 2-summing. Indeed, if, for example E is a $\mathcal{L}_1$-space and F is a $\mathcal{L}_2$-space, see e.g. [11, p. 60] for a definition, then one can show with the help of Grothendieck's inequality that every bounded linear operator $A : E \to F$ is 1-summing, see e.g. [11, Thm. 3.1], and therefore also 2-summing, see e.g. [11, Thm. 2.8]. Moreover, every Hilbert space is an $\mathcal{L}_2$-space, and since the 2-summing operators form an operator ideal, we directly obtain the following result, which shows that *i)* of Theorem 7 is sharp in some cases.

Lemma 8. *Let E and F be BFSs on some set X with $E \subset F$. If E is an $\mathcal{L}_1$-space and there exists an RKHS H over X with $E \subset H \subset F$, then the inclusion map* $\mathrm{id} : E \to F$ *is 2-summing.*

We note that a similar result holds, if E is an $\mathcal{L}_\infty$-space thanks to another application of Grothendieck's inequality, see e.g. [11, Thm. 3.7]. In addition, if E is a subspace of an $\mathcal{L}_p$-space for some $p \in [1, 2]$, then every bounded linear $A : E \to F$ that is q-summing for some $q > 2$ is also 2-summing. For such E, we can thus replace *i)* of Theorem 7 by the q-summability of $\mathrm{id} : E \to F$. Similarly, if E is an $\mathcal{L}_p$-space for some $p \in [2, \infty]$ and $1/q \geq 1/2 - 1/p$, then every q-summing operator $A : E \to F$ is 2-factorizing, see e.g. [11, p. 168].

The assumption in *ii)* of Theorem 7 can be slightly relaxed. Indeed, if both E' and F have cotype 2 and one of these spaces has the so-called approximation property, then [27], see also [28, Thm. 4.1] has shown that every bounded linear operator $A : E \to F$ is 2-factorable. Here we also refer to [28, Thm. 8.17] for another relaxation. Moreover, [28, Thm. 3.4] shows that every operator A that has a factorization $A = VU$, where U is of type 2 and V is of cotype 2, is 2-factorable, and since Hilbert spaces are of type 2 and cotype 2, the converse implication is obviously also true, using the ideal property of type 2 and cotype 2. In this direction we also note that [19] has found examples of Banach spaces E and F such that neither E' nor F have cotype 2, but every bounded linear operator $A : E \to F$ is 2-factorable. However, these spaces are not BSFs. In addition, the embeddings $\ell_1 \hookrightarrow \ell_2 \hookrightarrow \ell_\infty$ show that these (co)type 2 assumptions on E and F are in general *not* necessary for a positive answer to our Question (1). Moreover, our results on Hölder spaces, see Theorem 12 and Theorem 22, show that in the absence of (co)type 2 assumptions on E and F both positive and negative results are possible even for natural embeddings within the same scale of spaces. In contrast, the following straightforward corollary shows that the type 2 and cotype 2 of $\mathrm{id} : E \to F$ is a necessary condition for $E \subset H \subset F$.

Corollary 9. *Let E and F be BFSs on some set X with $E \subset F$. If there exists an RKHS H over X with $E \subset H \subset F$, then the inclusion map* $\mathrm{id} : E \to F$ *is of both type 2 and cotype 2.*

4 Applications to Various Concrete Spaces

We have seen in Theorem 6 that our Question (1) is characterized by the 2-factorability of the embedding $E \hookrightarrow F$. In this section, we thus investigate the 2-factorability of embeddings $E_\theta \hookrightarrow F_{\theta'}$ in various parameterized families of spaces of "functions", establishing both *positive* and *negative results*. The positive results explicitly state 2-factorizations given sufficiently well-behaved parameters θ, θ', which we straightforwardly derive from known embedding theorems. Negative results are derived from investigating type 2 and cotype 2, which are required for 2-factorability as stated in Lemma 5.

4.1 Warm-up: ℓ_p-spaces and L_p-spaces

In this subsection, we quickly investigate inclusions between ℓ_p-spaces, respectively L_p-spaces. Here, our focus lies on explaining the common strategy for deriving negative results rather than on deriving particularly interesting new results. This common strategy will then applied to more interesting scenarios in the subsequent subsections.

We begin with the most simple example of a family of spaces for which we investigate 2-factorability, namely ℓ_p-spaces. Recall that these spaces are proper BSFs, and hence the following lemma also precisely answers our Question (1) for this family of BSFs.

Lemma 10. *Let $1 \leq p \leq q \leq \infty$. Then the embedding*

$$\ell_p \hookrightarrow \ell_q$$

is 2-factorable if and only if $p \leq 2 \leq q$. In this case it 2-factorizes over the RKHS ℓ_2 by $\ell_p \hookrightarrow \ell_2 \hookrightarrow \ell_q$.

Proof. Clearly, we only need to show that $p \leq 2 \leq q$ is a necessary requirement for 2-factorability. To this end, let $e_1, e_2, \ldots \in \ell_p$ be the sequence of unit vectors. For $n \geq 1$ we then have

$$\|(e_1, \ldots, e_n)\|_{\mathrm{Rad}^n(\ell_p)} = \mathbb{E}_{\varepsilon \sim \mathrm{Rad}} \Big\| \sum_{i=1}^{n} \varepsilon_i e_i \Big\|_{\ell_p} = n^{1/p}$$

and

$$\|(e_1, \ldots, e_n)\|_{\ell_2^n(\ell_q)} = \Big(\sum_{i=1}^{n} \|e_i\|_{\ell_q}^2 \Big)^{1/2} = n^{1/2}.$$

Now, if $\ell_p \hookrightarrow \ell_q$ is 2-factorable, then it is of type 2 by Lemma 5, and hence we obtain

$$\|\ell_p \hookrightarrow \ell_q\|_{\mathrm{type}_2} \geq \sup_{(x_1,\ldots,x_n) \in \ell_p^n \setminus \{0\}} \frac{\|(x_1, \ldots, x_n)\|_{\ell_2^n(\ell_q)}}{\|(x_1, \ldots, x_n)\|_{\mathrm{Rad}^n(\ell_p)}} \geq \frac{\|(e_1, \ldots, e_n)\|_{\ell_2^n(\ell_q)}}{\|(e_1, \ldots, e_n)\|_{\mathrm{Rad}^n(\ell_p)}} = n^{1/2-1/p}$$

for all $n \geq 1$. Since $\|\ell_p \hookrightarrow \ell_q\|_{\mathrm{type}_2} < \infty$, this implies $p \leq 2$. Using $\|\ell_p \hookrightarrow \ell_q\|_{\mathrm{cotype}_2} < \infty$ we analogously obtain $q \geq 2$. $\square$

The next lemma investigates L_p-spaces, where for the sake of simplicity we only consider the domain $[0, 1]^d$ and the Lebesgue measure. Here we note that these spaces are, of course, not BSFs, but the common strategy becomes more visible than in the previous example.

Lemma 11. *Let $\Omega := [0,1]^d$ and $1 \leq q \leq p \leq \infty$. Then, the embedding $L_p(\Omega) \hookrightarrow L_q(\Omega)$ is 2-factorable if and only if $q \leq 2 \leq p$. In this case we have $L_p(\Omega) \hookrightarrow L_2(\Omega) \hookrightarrow L_q(\Omega)$.*

Proof. Again, we only need to show that 2-factorability implies $p \geq 2 \geq q$. To this end, let $n \in \mathbb{N}$ and $m := n^d$. Moreover, let $A_1, \dots, A_m$ be the partition of $[0,1]^d$ into m cubes of side length $1/n$. Then, we obtain

$$\|(\mathbb{1}_{A_1}, \dots, \mathbb{1}_{A_m})\|_{\mathrm{Rad}^m(L_p(\Omega))} = \mathbb{E}_{\varepsilon\sim\mathrm{Rad}} \Big\| \sum_{i=1}^m \varepsilon_i \mathbb{1}_{A_i} \Big\|_{L_p(\Omega)} = 1$$

and

$$\|(\mathbb{1}_{A_1}, \dots, \mathbb{1}_{A_m})\|_{\ell_2^m(L_q(\Omega))} = \left(\sum_{i=1}^m \|\mathbb{1}_{A_i}\|_{L_q(\Omega)}^2 \right)^{1/2} = n^{d(1/2-1/q)}.$$

Now, if $L_p(\Omega) \hookrightarrow L_q(\Omega)$ is 2-factorable, then it is of type 2 by Lemma 5, and hence we obtain

$$\|L_p(\Omega) \hookrightarrow L_q(\Omega)\|_{\mathrm{type}_2} \geq \frac{\|(\mathbb{1}_{A_1}, \dots, \mathbb{1}_{A_m})\|_{\ell_2^m(L_q(\Omega))}}{\|(\mathbb{1}_{A_1}, \dots, \mathbb{1}_{A_m})\|_{\mathrm{Rad}^m(L_p(\Omega))}} = n^{d(1/2-1/q)}.$$

for all $n \geq 1$. Since $\|L_p(\Omega) \hookrightarrow L_q(\Omega)\|_{\mathrm{type}_2} < \infty$, this implies $q \leq 2$. Analogously, we obtain $p \geq 2$ using $\|L_p(\Omega) \hookrightarrow L_q(\Omega)\|_{\mathrm{cotype}_2} < \infty$. $\square$

It is easy to generalize Lemma 11 to atom-free finite measure spaces (Ω, μ). Since Lemma 11 is mostly stated to illustrate our common proof strategy, we omit the details.

4.2 Spaces of Hölder Continuous Functions

Let (X, d) be a bounded metric space. In this section, we establish necessary requirements for the existence of an RKHS H on X that can be squeezed in between two fixed Hölder spaces over (X, d), that is

$$\mathrm{Höl}_d^\alpha(X) \hookrightarrow H \hookrightarrow \mathrm{Höl}_d^\beta(X)\,,$$

where $0 < \beta \leq \alpha \leq 1$. Here, our result will show that the existence of such an RKHS H implies an upper asymptotic bound on the growth of the packing numbers $N_d^{\mathrm{pack}}(X, \delta)$ of (X, d). Moreover, we obtain a similar result if we replace $\mathrm{Höl}_d^\beta(X)$ by $\ell_\infty(X)$.

Let us begin by recalling the definition of these spaces. To this end, we denote the α-Hölder norm of a function $f : X \to \mathbb{R}$ by

$$\|f\|_{\mathrm{Höl}_d^\alpha} := \max\left\{ \sup_{x\neq y\in X} \frac{|f(x) - f(y)|}{d^\alpha(x,y)}, \sup_{x\in X} |f(x)| \right\}, \tag{11}$$

where $\alpha \in (0, 1]$. Clearly, we have $\|f\|_{\mathrm{Höl}_d^\alpha} < \infty$ if and only if f is both α-Hölder continuous and bounded. Moreover, we define the Hölder spaces by

$$\mathrm{Höl}_d^\alpha(X) := \{ f : X \to \mathbb{R} \mid \|f\|_{\mathrm{Höl}_d^\alpha} < \infty \}\,.$$

It is well-known and easy to verify that $(\mathrm{Höl}_d^\alpha(X), \|\cdot\|_{\mathrm{Höl}_d^\alpha})$ is a Banach space. In addition, since Hölder spaces consist of bounded functions, we can quickly check that $\mathrm{Höl}_d^\alpha(X) \hookrightarrow \mathrm{Höl}_d^\beta(X)$ holds for all $0 < \beta \leq$

$\alpha \leq 1$. Moreover, another routine check shows that it is also a proper BSF. By Theorem 6, the existence of an intermediate RKHS H is therefore equivalent to the 2-factorability of an embedding $\mathrm{H\ddot{o}l}_d^\alpha(X) \hookrightarrow \mathrm{H\ddot{o}l}_d^\beta(X)$.

Finally, let us recall that the diameter of a metric space (X, d) is $\operatorname{diam} X := \sup_{s,t \in X} d(s,t) > 0$. With these preparations our result reads as follows.

Theorem 12. *Let (X, d) be a metric space and $0 < \beta \leq \alpha \leq 1$. If X is connected and there exists an RKHS H over X with*

$$\mathrm{H\ddot{o}l}_d^\alpha(X) \hookrightarrow H \hookrightarrow \mathrm{H\ddot{o}l}_d^\beta(X)\,, \tag{12}$$

then there exists a constant $C > 0$ such that for all $0 < \delta < \min\{1, \operatorname{diam} X\}$ we have

$$N_d^{\mathrm{pack}}(X, \delta) \leq C \delta^{-2(\alpha-\beta)}\,. \tag{13}$$

Moreover, if there exists an RKHS H over X such that

$$\mathrm{H\ddot{o}l}_d^\alpha(X) \hookrightarrow H \hookrightarrow \ell_\infty(X)\,, \tag{14}$$

then there exists a constant $C > 0$, such that (13) *holds for $\beta := 0$ and all $0 < \delta < \min\{1, \operatorname{diam} X\}$.*

Let us consider a bounded $X \subset \mathbb{R}^d$ with non-empty interior. By Theorem 12 and the packing number bound (8) we conclude that there is no RKHS H satisfying (12) if $\alpha - \beta < d/2$. Moreover, there is no RKHS H satisfying (14) if $\alpha < d/2$. Hence, (1) is infeasible for $d > 2$ within the family of Hölder spaces. We observe that the difference of the smoothness between the Hölder spaces in (12) needs to grow proportional to the underlying dimension in order to enable (1).

In Section 4.5, (26) will generalize Hölder spaces over X, allowing arbitrarily large non-integer smoothness. In Theorem 21 we will show that if $\alpha - \beta > d/2$, then there exists an intermediate RKHS H such that we have

$$\mathrm{H\ddot{o}l}^\alpha(X) \hookrightarrow H \hookrightarrow \mathrm{H\ddot{o}l}^\beta(X)$$

and in Theorem 22 we will show that for $\alpha > d/2$ there exists an intermediate RKHS H such that

$$\mathrm{H\ddot{o}l}^\alpha(X) \hookrightarrow H \hookrightarrow C^0(X)$$

holds. Hence, the bounds derived in Theorem 12 are tight up to the border cases $\alpha - \beta = d/2$, respectively $\alpha = d/2$. For general metric spaces (X, d) however, we do not know any condition that implies the existence of an RKHS H satisfying (12) or (14).

The requirement that X is a *bounded* subset of $\mathbb{R}^d$ is crucial to obtain positive answers to (1), as the following theorem shows.

Theorem 13. *Let $\Omega \subset \mathbb{R}^d$ be open and unbounded. Then there exists no RKHS H on Ω with bounded kernel such that $C^\infty(\Omega) \subset H$.*

By $C^\infty(\Omega)$ we denote the space of smooth functions, given by the norm $\|f\|_{C^\infty(\Omega)} = \sup_{\alpha \in \mathbb{N}_0^d} \|\partial_\alpha f\|_\infty$.

4.3 Spaces of (Generalized) Mixed Smoothness

In this section we investigate our Question 1 for a scale of spaces that generalize both *classical Sobolev spaces* and *Sobolev spaces of mixed smoothness*, see e.g. [1], respectively [33]. Since some of these spaces are not BSFs, we will focus on the 2-factorability of embeddings between such spaces, where we recall that by Theorem 6 2-factorability is equivalent to our initial Question 1 if the involved spaces are BSFs.

To introduce these spaces, we fix a finite, non-empty set A of multi-indices, i.e. $A \subset \mathbb{N}_0^d$. Then we say that A is coherent, if for all $\alpha \in A$ and $\beta \in \mathbb{N}_0^d$ with $\beta \leq \alpha$ we have $\beta \in A$, where $\beta \leq \alpha$ denotes the usual element-wise partial order. Moreover, we write $|A|_1 := \max\{|\alpha|_1 : \alpha \in A\}$, where $|\alpha|_1 := \alpha_1 + \cdots + \alpha_d$ denotes the sum of the entries of $\alpha \in \mathbb{N}_0^d$.

Given a bounded domain $\Omega \subset \mathbb{R}^d$ and a coherent $A \subset \mathbb{N}_0^d$, we now define the set of A-times weakly differentiable functions by

$$W^A_{\text{aux}}(\Omega) := \{f \in \mathcal{L}_0(\Omega) \mid \text{the weak derivative } \partial_\alpha f \text{ exists for all } \alpha \in A\}\,, \tag{15}$$

where $\mathcal{L}_0(\Omega)$ denotes the space of all measurable $f : \Omega \to \mathbb{R}$. Moreover, we define the *Sobolev space of generalized mixed smoothness* A as

$$W^A_p(\Omega) := \{f \in W^A_{\text{aux}}(\Omega) : \|f\|_{W^A_p(\Omega)} < \infty\},$$

where the *mixed Sobolev norm* of f is given by

$$\|f\|_{W^A_p(\Omega)} := \sup_{\alpha\in A} \|\partial_\alpha f\|_{L_p(\Omega)}. \tag{16}$$

Note that for $s \in \mathbb{N}_0$ and $A := \{\alpha \in \mathbb{N}_0^d : |\alpha|_1 \leq s\}$ the spaces $W^A_p(\Omega)$ equal the classical Sobolev spaces $H^s_p(\Omega)$ of integer smoothness s modulo an equivalent norm. Here we note that both our initial Question 1 and 2-factorability are invariant with respect to equivalent norms, and therefore these small differences between $W^A_p(\Omega)$ and $H^s_p(\Omega)$ do not affect our results. Finally note that we obviously have $|A|_1 = s$.

If, for a fixed $s \in \mathbb{N}_0$, we define $A := \{\alpha \in \mathbb{N}_0^d : \alpha_1, \ldots, \alpha_d \leq s\}$, then $W^{s,\text{mix}}_p(\Omega) := W^A_p(\Omega)$ equals a Sobolev space of mixed smoothness s, where again the latter spaces may have an equivalent norm. Note that for such A we have $|A|_1 = sd$.

In the following we provide necessary parameter requirements for 2-factorability of embeddings between Sobolev spaces of generalized mixed smoothness.

Theorem 14. *Let $\Omega \subset \mathbb{R}^d$ be a bounded domain, $A, B \subset \mathbb{N}_0^d$ be coherent sets, and $1 \leq p_1, p_2 < \infty$ such that the embedding $W^A_{p_1}(\Omega) \hookrightarrow W^B_{p_2}(\Omega)$ exists. We write $s := |A|_1$ and $t := |B|_1$, and further assume that $s - t \geq d(1/p_1 - 1/p_2)$ holds. If the embedding $W^A_{p_1}(\Omega) \hookrightarrow W^B_{p_2}(\Omega)$ is 2-factorable, then we have*

$$s - t \geq (d/p_1 - d/2)_+ + (d/2 - d/p_2)_+\,. \tag{17}$$

Recall that for classical Sobolev spaces $H^s_{p_1}(\Omega)$ and $H^t_{p_2}(\Omega)$ of integer smoothness with $s \geq t$, it is asserted in [29, p. 82, Remark], see also the discussion in [1, p. 108ff], that the embedding $H^s_{p_1}(\Omega) \hookrightarrow H^t_{p_2}(\Omega)$ exists if and only if $s - t \geq d/p_1 - d/p_2$. Moreover, recall that for these spaces we have $H^s_p(\Omega) = W^A_p(\Omega)$, where $A := \{\alpha \in \mathbb{N}_0^d : |\alpha|_1 \leq s\}$. Since this gives $|A|_1 = s$, Theorem 14 shows that if $H^s_{p_1}(\Omega) \hookrightarrow H^t_{p_2}(\Omega)$ is 2-factorable then (17) holds.

In contrast, if we consider Sobolev spaces of mixed smoothness s respectively t as introduced above, then the 2-factorability of the embedding $W^{s,\text{mix}}_{p_1}(\Omega) \hookrightarrow W^{t,\text{mix}}_{p_2}(\Omega)$ implies

$$s - t \geq (1/p_1 - 1/2)_+ + (1/2 - 1/p_2)_+$$

since the considered coherent sets A and B are of size $|A|_1 = sd$ and $|B|_1 = td$. Note that compared to classical Sobolev spaces, the dimension no longer appears in the necessary condition.

The next theorem investigates the case in which $W^B_{p_2}(\Omega)$ is replaced by $\ell_\infty(\Omega)$, that is we consider $W^A_p(\Omega) \hookrightarrow \ell_\infty(\Omega)$, where the embedding is understood in the usual sense of Sobolev's embedding theorem. Informally speaking this complements with the case Theorem 14 with $B = \{0\}$ and $p_2 = \infty$.

Theorem 15. *Let $\Omega \subset \mathbb{R}^d$ be a bounded domain, $A \subset \mathbb{N}_0^d$ be a coherent set, and $1 \leq p < \infty$ such that the embedding $W^A_p(\Omega) \hookrightarrow \ell_\infty(\Omega)$ exists. We define $s := |A|_1$ and assume $s \geq d/p$. If there exists an RKHS H with $W^A_p(\Omega) \subset H \subset \ell_\infty(\Omega)$, then we have*

$$s \geq (d/p - d/2)_+ + d/2\,.$$

We can generalize Theorem 14 to *mixed fractional smoothness* in the style of *Slobodeckij spaces*, which we will use in Section 5.2.4. A space $W^A_p(\Omega)$ of mixed fractional smoothness is given by an integration index $p \geq 1$ and a *coherent set of mixed smoothness* A. The elements (α, θ) of A consist of a multi-index $\alpha \in \mathbb{N}_0^d$ which ensures the coherent structure of A, and a parameter $\theta \in [0, 1)$ that refines the smoothness of the derivative ∂_α to a fractional smoothness as in (39). Defining $|A|_1 := \max\{|\alpha|_1 + \theta \mid (\alpha, \theta) \in A\}$ we obtain analogously to Theorem 14 that for an embedding $W^A_{p_1}(\Omega) \hookrightarrow W^B_{p_2}(\Omega)$ the inequality

$$|A|_1 - |B|_1 \geq (d/p_1 - d/2)_+ + (d/2 - d/p_2)_+$$

is a necessary requirement for 2-factorability. For Slobodeckij spaces this result is sharp, as we will show in Theorem 20.

4.4 Embedding Integral Reproducing Kernel Banach Spaces into RKHSs

Consider a single hidden layer neural network $f : \mathbb{R}^d \to \mathbb{R}$ with n hidden neurons, that is, f is of the form

$$f(x) = \sum_{i=1}^{n} \alpha_i \sigma(\langle x, w_i \rangle + b_i),$$

where $\sigma : \mathbb{R} \to \mathbb{R}$ is an activation function, $w_i \in \mathbb{R}^d$ are weight vectors and $b_i \in \mathbb{R}$ are biases. This network's outputs can be interpreted as integrals by observing that for the discrete signed measure

$$\mu = \sum_{i=1}^{n} \alpha_i \delta_{(w_i, b_i)}$$

on $\mathbb{R}^{d+1}$ the identity

$$f(x) = \int_{\mathbb{R}^{d+1}} \sigma(\langle (x, 1), (w, b) \rangle_{\mathbb{R}^{d+1}})\, d\mu(w, b)$$

holds for all $x \in \mathbb{R}^d$, where we write $(w, b) \in \mathbb{R}^{d+1}$ for the vector obtained by appending $b \in \mathbb{R}$ to $w \in \mathbb{R}^d$. This perspective motivates the theory of *integral reproducing kernel Banach spaces* (IRKBS), which can be used to describe such networks and their limits, see e.g. [6, Section 3.5].

In this section we investigate Question (1) for IRKBS, which are a certain class of proper BSF. Let us begin by recalling the definition of IRKBS.

Definition 16. *Let $(X, \mathcal{A})$ be a measurable space, $\Psi : X \times X \to \mathbb{R}$ be a measurable function, and $\mathcal{M}$ be a Banach space of finite signed measures on X. We call a bounded and measurable function $\beta : X \to \mathbb{R}$ a normalizing function of Ψ, if*

$$\sup_{y \in X} |\Psi(x, y)\beta(y)| < \infty \tag{18}$$

holds for all $x \in X$. In this case, we define $f_{\mu,\Psi,\beta} : X \to \mathbb{R}$ by

$$f_{\mu,\Psi,\beta}(x) := \int_X \Psi(x, y)\beta(y)\, d\mu(y)\,, \qquad x \in X \tag{19}$$

for each $\mu \in \mathcal{M}$. Then the set $E_{\mathcal{M},\Psi,\beta} := \{f_{\mu,\Psi,\beta} \mid \mu \in \mathcal{M}\}$ equipped with the norm

$$\|f_\mu\|_{E_{\mathcal{M},\Psi,\beta}} := \inf\{\|\nu\|_{\mathcal{M}} \mid \nu \in \mathcal{M}, f_\nu = f_\mu\}$$

is called an integral reproducing kernel Banach space.

In the following, $\mathcal{M}(X)$ denotes the space of finite signed measures on X equipped with the total variation (TV) norm. The next result, which is a variant of [6, Prop. 3.3], shows that under natural conditions $E_{\mathcal{M},\Psi,\beta}$ becomes a proper BSF.

Proposition 17. *Let X be a measurable space, $\Psi : X \times X \to \mathbb{R}$ be a measurable function and $\beta : X \to \mathbb{R}$ be a normalizing function of Ψ. Let $\mathcal{M}$ be a Banach space of finite signed measures on X with $\mathcal{M} \hookrightarrow \mathcal{M}(X)$. Then, $(E_{\mathcal{M},\Psi,\beta}, \|\cdot\|_{E_{\mathcal{M},\Psi,\beta}})$ is a proper BSF.*

In the following, we would like to reduce our considerations to the case $\beta = 1$ without loosing generality. To explain this, let $(X, \mathcal{A})$ be a measurable space, $\beta : X \to \mathbb{R}$ be bounded and measurable, and $\mathcal{M}$ be a Banach space of signed measures on X. For $\mu \in \mathcal{M}$, we define a new finite, signed measure $\beta\mu$ on X by

$$(\beta\mu)(A) := \int_A \beta(y)\, d\mu(y)\,, \qquad A \in \mathcal{A}.$$

Moreover, we write $\beta\mathcal{M} := \{\beta\mu \mid \mu \in \mathcal{M}\}$ and

$$\|\mu\|_{\beta\mathcal{M}} := \inf\{\|\nu\|_{\mathcal{M}} \mid \nu \in \mathcal{M}, \mu = \beta\nu\}\,.$$

It can be quickly verified, that $(\beta\mathcal{M}, \|\cdot\|_{\beta\mathcal{M}})$ is a Banach space of finite signed measures. Moreover, the embedding $\mathcal{M} \hookrightarrow \mathcal{M}(X)$ and the boundedness of $\beta : X \to \mathbb{R}$ imply $\beta\mathcal{M} \hookrightarrow \mathcal{M}(X)$. Under the assumptions of Proposition 17 we further have

$$f_{\mu,\Psi,\beta} = f_{\beta\mu,\Psi,1} \tag{20}$$

and also

$$\begin{aligned}\|f_{\mu,\Psi,\beta}\|_{E_{\mathcal{M},\Psi,\beta}} = \inf\{\|\nu\|_{\mathcal{M}} \mid \nu \in \mathcal{M}, f_{\nu,\Psi,\beta} = f_{\mu,\Psi,\beta}\} &= \inf\{\|\nu\|_{\mathcal{M}} \mid \nu \in \mathcal{M}, f_{\beta\nu,\Psi,1} = f_{\beta\mu,\Psi,1}\} \\ &= \inf\{\|\beta\nu\|_{\beta\mathcal{M}} \mid f_{\beta\nu,\Psi,1} = f_{\beta\mu,\Psi,1}\} \\ &= \|f_{\beta\mu,\Psi,1}\|_{E_{\beta\mathcal{M},\Psi,1}}\,. \end{aligned} \tag{21}$$

In other words, (20) and (21) make it possible to work with the constant normalizing function $\beta = 1$ by adapting the considered Banach space of signed measures.

Having introduced IRKBSs, our next goal is to investigate under which assumptions these spaces can be embedded into an RKHS. To this end, we assume that $\Psi : X \times X \to \mathbb{R}$ is a symmetric function that can be written as a difference of two kernels k_1 and k_2 on X, that is

$$\Psi = k_1 - k_2 .$$

In the literature, this is known as a *positive decomposition* of Ψ and it is discussed in e.g. [8] and the references mentioned therein. Moreover, positively decomposable functions are the foundation of reproducing kernel Krein spaces (RKKSs), a generalization of RKHSs, see e.g. [24].

The following lemma investigates Question (1) for IRKBS $E_{\mathcal{M},\Psi,1}$ in cases where Ψ is positively decomposable.

Lemma 18. *Let X be a measurable space and $\mathcal{M}$ be a Banach space of finite signed measures on X with $\mathcal{M} \hookrightarrow \mathcal{M}(X)$. Moreover, let k_1, k_2 be measurable kernels on X with RKHSs H_1, H_2, such that $\Psi := k_1 - k_2$ is bounded. Then, the IRKBS $E_{\mathcal{M},\Psi,1}$ is a proper BSF. If, in addition, we have*

$$H_1, H_2 \subset \mathcal{L}_1(\mu) \tag{22}$$

for all $\mu \in \mathcal{M}$, then the RKHS $H = H_1 + H_2$ of the kernel $k := k_1 + k_2$ fulfils

$$E_{\mathcal{M},\Psi,1} \hookrightarrow H.$$

An easy way to check (22) is provided by [39, Thm. 4.26], which shows that if k is a measurable kernel on X with RKHS H and μ is a finite signed measure on X with

$$\|k\|_{L_1(\mu)} := \int_X \sqrt{k(x,x)}\, d|\mu|(x) < \infty ,$$

where $|\mu| := \mu_+ - \mu_-$ denotes the total variation of μ, then $H \subset \mathcal{L}_1(\mu)$ follows.

To adapt this to our situation, we first note that (22) is obviously equivalent to $H_1 + H_2 \subset \mathcal{L}_1(\mu)$ for all $\mu \in \mathcal{M}$. Since $k_1 + k_2$ is the kernel of $H_1 + H_2$, we thus see that (22) is satisfied if

$$\int_X \sqrt{k_1(x,x) + k_2(x,x)}\, d\mu(x) < \infty , \qquad \mu \in \mathcal{M}. \tag{23}$$

For example, if k_1 and k_2 are bounded, (23) holds, and thus also (22). As a consequence, we then have the embeddings $E_{\mathcal{M},\Psi,1} \hookrightarrow H_1 + H_2 \hookrightarrow \ell_\infty(X)$.

To illustrate Lemma 18 and its crucial assumption (22) in a concrete setting, we fix a real sequence $(\lambda_i)_{i\geq 0}$ and consider the power series

$$\sigma_+(t) := \sum_{i=0}^{\infty} (\lambda_i)_+ \, t^i ,$$
$$\sigma_-(t) := \sum_{i=0}^{\infty} (\lambda_i)_- \, t^i$$

as well as their convergence radii R_+ and R_-. Moreover, we write $R := \min\{R_+, R_-\}$ and assume $R > 0$.

For $X \subset \mathring{B}(0, \sqrt{R})$ we define $k_1, k_2 : X \times X \to \mathbb{R}$ by $k_1(x,y) \coloneqq \sigma_+(\langle x,y \rangle)$ and $k_2(x,y) \coloneqq \sigma_-(\langle x,y \rangle)$. Then, k_1 and k_2 are continuous and [39, Lemma 4.8.] shows that k_1 and k_2 are kernels on X since all coefficients are non-negative. We further define $\Psi : X \times X \to \mathbb{R}$ by $\Psi \coloneqq k_1 - k_2$, that is,

$$\Psi(x,y) = \sum_{i=0}^{\infty} \lambda_i \langle x,y \rangle^i , \qquad x, y \in X.$$

Let us assume for simplicity that Ψ is bounded, so that we may choose $\beta \equiv 1$ as a normalizing function of Ψ.

Then, Lemma 18 shows that the IRKBS $E_{\mathcal{M}(X),\Psi,1}$ is a proper BSF. To obtain a positive answer to Question (1) from Lemma 18 it remains to check (22). Here we first note that the sufficient condition (23) for (22) reads as

$$\sqrt{\sum_{i=0}^{\infty} |\lambda_i| \| \cdot \|^{2i}} \in \mathcal{L}_1(\mu) , \qquad \mu \in \mathcal{M}.$$

Unfortunately, this condition is not necessary. To derive a necessary condition, we first note that (22) implies $k_i(\,\cdot\,, x) \in \mathcal{L}_1(\mu)$ for all $x \in X$, $\mu \in \mathcal{M}$, and $i = 1, 2$. In our situation, the latter reads as

$$\sum_{i=0}^{\infty} |\lambda_i| \langle \,\cdot\, , x \rangle^i \in \mathcal{L}_1(\mu) \tag{24}$$

for all $x \in X$ and $\mu \in \mathcal{M}$.

Finally, the following example shows that the condition (18) ensuring that $E_{\mathcal{M},\Psi,1}$ is a proper BSF is in general substantially weaker than the condition (22) ensuring a surrounding RKHS.

Example 19. *Let $X = \mathbb{R}^d$, $\sigma : \mathbb{R} \to \mathbb{R}$ be given by*

$$\sigma(t) = \cos(t) = \sum_{i=0}^{\infty} (-1)^i \frac{t^{2i}}{(2i)!} ,$$

and $\Psi : X \times X \to \mathbb{R}$ be given by $\Psi(x,y) = \sigma(\langle x,y \rangle)$. Then, Ψ is bounded and for $\mathcal{M} := \mathcal{M}(X)$ we know by Lemma 18 that the IRKBS $E_{\mathcal{M},\Psi,1}$ is a proper BSF. Moreover, our discussion above shows that a positive decomposition of Ψ is given by the kernels

$$k_1(x,y) = \sum_{i=0}^{\infty} \frac{t^{4i}}{(4i)!} \qquad \text{and} \qquad k_2(x,y) = \sum_{i=0}^{\infty} \frac{t^{4i+2}}{(4i+2)!}.$$

By (24) *we thus see that we need at least*

$$\cosh(\langle \,\cdot\, , x \rangle) \in \mathcal{L}_1(\mu)$$

for all $\mu \in \mathcal{M}, x \in X$ in order to apply Lemma 18. Consequently, we can only find an RKHS H with $E_{\mathcal{M},\Psi,1} \hookrightarrow H$ with the help of Lemma 18, if we substantially reduce the space $\mathcal{M}$ of considered measures.

4.5 Triebel-Besov-Lizorkin spaces

In this section we consider Triebel-Lizorkin spaces and Besov spaces. To quickly introduce those spaces, let us fix a bounded domain $\Omega \subset \mathbb{R}^d$ with smooth boundary [29, Sect. 2.4.1, Def. 1]. We denote *Triebel-Lizorkin spaces* by $F^s_{p,q}(\Omega)$, where $s \geq 0$ is the *smoothness*, $p \in [1, \infty)$ is the *integration index*, and $q \in [1, \infty]$ is the *fein index*. Moreover, *Besov spaces* are denoted as $B^s_{p,q}(\Omega)$ where $s \geq 0$ and $p, q \in [1, \infty]$.

In the following, we say that $X^s_{p,q}(\Omega)$ is a Besov-Triebel-Lizorkin space, if either $X^s_{p,q}(\Omega) = F^s_{p,q}(\Omega)$ with $p < \infty$ or $X^s_{p,q}(\Omega) = B^s_{p,q}(\Omega)$ with $p \in [1, \infty]$.

Recall from e.g. [29, Sect. 2.2.4] that for $s > d/p$, the spaces $F^s_{p,q}(\Omega)$ and $B^s_{p,q}(\Omega)$ can be continuously embedded into $C^0(\Omega)$ and consequently, they are proper BSFs. Nonetheless, we will also investigate the case $s \leq d/p$, where we can still sharply answer the question of 2-factorability. Finally, some additional embedding theorems for Besov-Triebel-Lizorkin spaces are recalled in Theorem 32.

Many well known function spaces are special cases of Besov-Triebel-Lizorkin spaces [41, Prop. 2.3.5], especially the following identities hold up to norm-equivalence:

The fractional Sobolev spaces $H^s_p(\Omega)$, which are sometimes also referred to as Bessel potential spaces, can be identified via

$$H^s_p(\Omega) = F^s_{p,2}(\Omega)\,, \qquad\qquad s \geq 0, p \in (1, \infty). \tag{25}$$

Here we recall that $H^s_2(\Omega)$ is a *Hilbert space* which can be quickly inferred from their definition with the help of the Fourier transformation, see e.g. [29, p. 13]. Combining this with the above mentioned embedding into $C^0(\Omega)$, the spaces $H^s_2(\Omega)$ are RKHSs whenever $s > d/2$.

For $s > 0$ with $s \notin \mathbb{N}$, the Hölder spaces $\mathrm{Höl}^s(\Omega)$ of $\lfloor s \rfloor$-times differentiable, bounded functions, for which all these derivatives are bounded and the derivatives of order $\lfloor s \rfloor$ are $s - \lfloor s \rfloor$-Hölder continuous, can be identified via

$$\mathrm{Höl}^s(\Omega) = B^s_{\infty,\infty}(\Omega). \tag{26}$$

Note that for $s \in (0, 1)$, these spaces coincide with the Hölder spaces discussed in Section 4.2. In contrast to Section 4.2, however, we can now also deal with smoothness $s > 1$.

For $s > 0$ and $p \in [1, \infty)$, the Slobodeckij spaces $W^s_p(\Omega)$, whose definition is recalled in Section 5.2.4 can be identified via

$$W^s_p(\Omega) = F^s_{p,q}(\Omega), \qquad\qquad \text{where } q = \begin{cases} 2, & \text{if } s \in \mathbb{N} \text{ and } p > 1, \\ p, & \text{if } s \notin \mathbb{N}. \end{cases} \tag{27}$$

Besov spaces and Triebel-Lizorkin spaces coincide when their integration and fein index are equal, that is

$$B^s_{p,p}(\Omega) = F^s_{p,p}(\Omega), \qquad\qquad p < \infty. \tag{28}$$

This is the only scenario in which Besov spaces and Triebel-Lizorkin spaces are equal, see [41, 2.3.9].

Our first result in this section investigates 2-factorability of embeddings between Slobodeckij spaces. Modulo limiting cases, it provides a characterization.

Theorem 20. *Let $\Omega \subset \mathbb{R}^d$ be a bounded domain with smooth boundary, and let $0 \leq t < s < \infty$ and $1 \leq p_1, p_2 < \infty$. If $s \in \mathbb{N}$, let $p_1 > 1$ and if $t \in \mathbb{N}_0$, let $p_2 > 1$. Furthermore, let $s - t \geq d(1/p_1 - 1/p_2)$ hold. Then, the embedding $W^s_{p_1}(\Omega) \hookrightarrow W^t_{p_2}(\Omega)$ exists and the following statements hold true:*

i) ***Sufficient requirement.*** *If $s-t>(d/p_1-d/2)_+ + (d/2-d/p_2)_+$, then the embedding 2-factorizes as*

$$W^s_{p_1}(\Omega) \hookrightarrow H^u_2(\Omega) \hookrightarrow W^t_{p_2}(\Omega) \tag{29}$$

for any $u \in (t+(d/2-d/p_2)_+ \, , s-(d/p_1-d/2)_+ \,)$.

ii) ***Necessary requirement.*** *If the embedding $W^s_{p_1}(\Omega) \hookrightarrow W^t_{p_2}(\Omega)$ is 2-factorable, then we have*

$$s-t \geq (d/p_1-d/2)_+ + (d/2-d/p_2)_+ \, .$$

Lemma 33, a result showing that Slobodeckij spaces are "dense" in the Besov-Triebel-Lizorkin spaces in a suitable way, allows analyzing 2-factorability of embeddings of Besov-Triebel-Lizorkin spaces based on Theorem 20, yielding a central result which is our strongest answer to (1).

Theorem 21. *Let $\Omega \subset \mathbb{R}^d$ be a smooth bounded domain and let $0 \leq t < s$. Let $X^s_{p_1,q_1}(\Omega)$ and $Y^t_{p_2,q_2}(\Omega)$ be spaces*[1] *of Triebel-Lizorkin type $F^s_{p,q}(\Omega)$ or of Besov type $B^s_{p,q}(\Omega)$. If both spaces are of Triebel-Lizorkin type, assume $s-t \geq d/p_1 - d/p_2$, and otherwise let $s-t > d/p_1-d/p_2$. Then, the embedding*

$$X^s_{p_1,q_1}(\Omega) \hookrightarrow Y^t_{p_2,q_2}(\Omega)$$

exists and the following statements hold true:

i) ***Sufficient requirement.*** *If $s-t>(d/p_1-d/2)_+ + (d/2-d/p_2)_+$, then the embedding 2-factorizes as*

$$X^s_{p_1,q_1}(\Omega) \hookrightarrow H^u_2(\Omega) \hookrightarrow Y^t_{p_2,q_2}(\Omega) \tag{30}$$

for any $u \in (t+(d/2-d/p_2)_+ \, , s-(d/p_1-d/2)_+ \,)$. If both spaces are of Triebel Lizorkin type, then the 2-factorization in (30) *even exists for all $u \in [t+(d/2-d/p_2)_+ \, , s-(d/p_1-d/2)_+ \,] \setminus \{s,t\}$.*

ii) ***Necessary requirement.*** *If $t>0$ and the embedding $X^s_{p_1,q_1}(\Omega) \hookrightarrow Y^t_{p_2,q_2}(\Omega)$ is 2-factorable, then we have*

$$s-t \geq (d/p_1-d/2)_+ + (d/2-d/p_2)_+ \, .$$

Finally, we answer Question (1) for $F = \ell_\infty(\Omega)$, characterizing of the existence of an encompassing RKHS H consisting of bounded functions for Besov-Triebel-Lizorkin spaces modulo border cases.

Theorem 22. *Let $\Omega \subset \mathbb{R}^d$ be a smooth bounded domain and let $s>0$. Let $X^s_{p,q}(\Omega)$ be a space of Triebel-Lizorkin type $F^s_{p,q}(\Omega)$ or of Besov type $B^s_{p,q}(\Omega)$. Let $s > d/p$. Then, the embedding*

$$X^s_{p,q}(\Omega) \hookrightarrow C^0(\Omega)$$

exists and the following statements hold true:

i) ***Sufficient requirement.*** *If $s > (d/p-d/2)_+ + d/2$, then the embedding 2-factorizes as*

$$X^s_{p,q}(\Omega) \hookrightarrow H^u_2(\Omega) \hookrightarrow C^0(\Omega)$$

for any $u \in (d/2, s-(d/p-d/2)_+ \,)$.

[1]Mixed cases, such as for example $X^s_{p_1,q_1}(\Omega) = B^s_{p_1,q_1}(\Omega)$ and $Y^t_{p_2,q_2}(\Omega) = F^t_{p_2,q_2}(\Omega)$, are possible.

ii) ***Necessary requirement.*** *If the embedding $X^s_{p,q}(\Omega) \hookrightarrow C^0(\Omega)$ is 2-factorable, then we have*

$$s \geq (d/p - d/2)_+ + d/2.$$

Theorem 22 *i)* shows that over a *bounded* smooth domain $\Omega \subseteq \mathbb{R}^d$, we find a space E consisting of sufficiently smooth functions of *finite smoothness* such that for $F = C^0(\Omega)$ our Question (1) allows a positive answer. This is a strong contrast to an *unbounded* domain $\Omega \subseteq \mathbb{R}^d$, where Theorem 13 shows that (1) is not even feasible for $E = C^\infty(\Omega) \subset \ell_\infty(\Omega) = F$, which stresses the necessity of the boundedness assumption.

5 Proofs

5.1 Proofs for Section 3

For the following results recall that a BSF E on a set X is called proper if the point evaluation $\delta_x : E \to \mathbb{R}, f \mapsto f(x)$ is continuous for all $x \in X$, see Definition 1.

Lemma 23. *Let F be a proper BSF on X and E be a BSF on X with $E \subset F$. Then the following statements are equivalent:*

i) E is proper.

ii) We have $E \hookrightarrow F$.

In this case we further have $\|\delta_x\|_{E'} \leq \|\operatorname{id} : E \to F\| \cdot \|\delta_x\|_{F'}$ for all $x \in X$.

Proof of Lemma 23: *ii)* $\Rightarrow$ *i)*. We write $c := \|\operatorname{id} : E \to F\|$ and fix an $x \in X$. Then we have $B_E \subset cB_F$, and hence we obtain

$$\|\delta_x\|_{E'} = \sup_{f \in B_E} |\delta_x f| \leq \sup_{f \in cB_F} |\delta_x f| = c \sup_{f \in B_F} |\delta_x f| = c\|\delta_x\|_{F'} \,,$$

and in particular $\delta_x : E \to \mathbb{R}$ is continuous.

i) $\Rightarrow$ *ii)*. We will apply the Closed Graph Theorem, see e.g. [21, Theorem 1.6.11]. To this end let us fix a sequence $(f_n) \subset E$ with $f_n \to f$ in E and $f_n \to g$ in F. This yields both $f_n \to f$ pointwise and $f_n \to g$ pointwise, and hence we conclude $f = g$, that is $\operatorname{id} f = g$. The Closed Graph Theorem then gives the continuity of $\operatorname{id} : E \to F$. $\square$

Proof of Proposition 2: Let us write $\|f\|_E := \|f\|_F$ for all $f \in E$. Then $(E, \|\cdot\|_E)$ is complete since $E \subset F$ is closed, and consequently E is a BSF on X. Moreover, we obviously have $E \hookrightarrow F$, and hence E is proper by Lemma 23.

i) $\Rightarrow$ *ii)*. Since E, H, and F are proper BSFs an application of Lemma 23 yields $E \hookrightarrow H \hookrightarrow F$. Our next goal is to show that E is a closed subspace of H. To this end, let us fix a sequence $(f_n) \subset E$ and an $f \in H$ with $\|f_n - f\|_H \to 0$. Since $H \hookrightarrow F$, we then have $\|f_n - f\|_F \to 0$, and in particular, (f_n) is a Cauchy sequence with respect to $\|\cdot\|_F$. However, we have $\|f\|_E = \|f\|_F$ for all $f \in E$, and therefore (f_n) is also a Cauchy sequence in E. Since $(E, \|\cdot\|_E)$ is complete, there thus exists a $g \in E$ with $\|f_n - g\|_E \to 0$. Moreover, since both H and E are proper, we additionally have $f_n(x) \to f(x)$ and $f_n(x) \to g(x)$ for all $x \in X$, and therefore $f = g$. This shows $f \in E$.

Finally, since E is a closed subspace of H, we see that $(E, \|\cdot\|_H)$ is a Hilbert space, and in particular complete. In addition, $E \hookrightarrow H$ shows that $\operatorname{id} : (E, \|\cdot\|_E) \to (E, \|\cdot\|_H)$ is continuous, and obviously, it is also bijective. The Open Mapping Theorem then guarantees that the inverse is also continuous, see e.g. [21, Cor. 1.6.6].

ii) ⇒ i). If E is isomorphic to a Hilbert space, we can define a Hilbert space norm on E that is equivalent to $\|\cdot\|_E$. Then E equipped with this Hilbert space norm is an RKHS H that satisfies $E = H$. □

Lemma 24. *Let E be a proper BSF on X and $Y \subset X$ be non-empty. Then*

$$E_{|Y} := \{f_{|Y} : f \in E\}$$

equipped with the norm

$$\|g\|_{E_{|Y}} := \inf\{\|f\|_E \,:\, f \in E \text{ and } f_{|Y} = g\}\,, \qquad\qquad g \in E_{|Y}$$

is a proper BSF on Y with

$$\|\delta_y\|_{(E_{|Y})'} \le \|\delta_y\|_{E'}\,, \qquad\qquad y \in Y. \tag{31}$$

Moreover, the restriction operator $\cdot_{|Y} : E \to E_{|Y}$ is linear and continuous with $\|\cdot_{|Y} : E \to E_{|Y}\| \le 1$.

Proof of Lemma 24: Obviously, $E_{|Y}$ is a vector space consisting of functions $Y \to \mathbb{R}$ and the restriction operator is linear.

To verify that $\|\cdot\|_{E_{|Y}}$ is a norm, we first note that $\|\cdot\|_{E_{|Y}} \ge 0$ and $\|0\|_{E_{|Y}} = 0$ are obvious. Moreover, if $g \in E_{|Y}$ satisfies $\|g\|_{E_{|Y}} = 0$, then there exists a sequence $(f^{(n)}) \subset E$ with $f^{(n)}_{|Y} = g$ and $\|f^{(n)}\| \to 0$. The latter gives $f^{(n)}(x) \to 0$ for all $x \in X$, and hence we find

$$g(y) = f^{(n)}_{|Y}(y) = f^{(n)}(y) \to 0\,, \qquad\qquad y \in Y.$$

This shows $g = 0$. In addition, the homogeneity follows from

$$\|\alpha g\|_{E_{|Y}} = \inf\{\|f\|_E \,:\, f \in E \text{ and } f_{|Y} = \alpha g\} = \inf\{\|\alpha f\|_E \,:\, f \in E \text{ and } f_{|Y} = g\} = \alpha\|g\|_{E_{|Y}}\,.$$

To verify the triangle inequality, we fix some $g_1, g_2 \in E_{|Y}$. Since for all $f^{(1)}, f^{(2)} \in E$ with $f^{(i)}_{|Y} = g_i$ we have $(f^{(1)} + f^{(2)})_{|Y} = g_1 + g_2$ we then obtain

$$\begin{aligned}\|g_1 + g_2\|_{E_{|Y}} &\le \inf\{\|f^{(1)} + f^{(2)}\|_E \,:\, f^{(i)} \in E \text{ and } f^{(i)}_{|Y} = g_i\}\\ &\le \inf\{\|f^{(1)}\|_E + \|f^{(2)}\|_E \,:\, f^{(i)} \in E \text{ and } f^{(i)}_{|Y} = g_i\}\\ &= \|g_1\|_{E_{|Y}} + \|g_2\|_{E_{|Y}}\,.\end{aligned}$$

Let us now show that restriction operator $\cdot_{|Y} : E \to E_{|Y}$ is continuous. To this end, we fix an $f \in E$. Then we easily find

$$\|f_{|Y}\|_{E_{|Y}} = \inf\{\|\tilde f\|_E \,:\, \tilde f \in E \text{ and } \tilde f_{|Y} = f_{|Y}\}\,, \le \|f\|_E\,.$$

and hence we have the desired continuity with $\|\cdot_{|Y} : E \to E_{|Y}\| \le 1$.

Next, we show that the norm is complete. To this end, let $(g_n) \subset E_{|Y}$ be a sequence with

$$\sum_{n=1}^{\infty} \|g_n\|_{E_{|Y}} < \infty.$$

By a well-known characterization of *complete* normed spaces, see e.g. [21, Thm. 1.3.9], it suffices to show that $\sum_{n=1}^\infty g_n$ converges in $E_{|Y}$. To this end, we choose $f^{(n)} \in E$ with $f^{(n)}_{|Y} = g_n$ and $\|f^{(n)}\|_E \leq \|g_n\|_{E_{|Y}} + n^{-2}$ for all $n \geq 1$. This gives

$$\sum_{n=1}^{\infty} \|f^{(n)}_{|Y}\|_E < \infty,$$

and since E is complete, we see with the help of [21, Thm. 1.3.9] that $f \coloneqq \sum_{n=1}^\infty f^{(n)}$ exists with convergence in E. Using the already established continuity of the restriction operator $\cdot_{|Y} : E \to E_{|Y}$ we conclude that

$$\biggl(\sum_{n=1}^{\infty} f^{(n)}\biggr)_{|Y} = \biggl(\lim_{m\to\infty} \sum_{n=1}^{m} f^{(n)}\biggr)_{|Y} = \lim_{m\to\infty} \biggl(\sum_{n=1}^{m} f^{(n)}\biggr)_{|Y} = \lim_{m\to\infty} \sum_{n=1}^{m} f^{(n)}_{|Y} = \sum_{n=1}^{\infty} g_n\,,$$

where the convergence in the last three expressions take place in $E_{|Y}$ by the continuity of the restriction operator.

Finally, to check that $E_{|Y}$ is proper, we fix an $y \in Y$. For $g \in E_{|Y}$ and an $f \in E$ with $f_{|Y} = g$ we then obtain

$$|\delta_y g| = |g(y)| = |f_{|Y}(y)| = |\delta_y f| \leq \|\delta_y\|_{E'} \cdot \|f\|_E\,,$$

and by taking the infimum over all $f \in E$ with $f_{|Y} = g$ we then see that $|\delta_y g| \leq \|\delta_y\|_{E'} \cdot \|g\|_{E_{|Y}}$. The latter also shows (31). □

Proof of Proposition 3: We first show that $E_{|Y} \subset F_{|Y}$. To this end, we fix some $g \in E_{|Y}$. Then there exists an $f \in E$ with $f_{|Y} = g$. We then know that $f \in F$, which in turn implies $g = f_{|Y} \in F_{|Y}$. Moreover, we find

$$\|g\|_{F_{|Y}} = \inf\bigl\{\|\tilde f\|_F \,:\, \tilde f \in F \text{ and } \tilde f_{|Y} = g\bigr\} \leq \|f\|_F \leq \|\,\mathrm{id} : E \to F\| \cdot \|f\|_E\,,$$

and by taking the infimum over all $f \in E$ with $f_{|Y} = g$ we then find $\|g\|_{F_{|Y}} \leq \|\,\mathrm{id} : E \to F\| \cdot \|g\|_{E_{|Y}}$. □

Proof of Lemma 5: Let us fix a 2-factorization (10). Then we can expand this to

$$\begin{tikzcd} E \arrow[r, "\mathrm{id}"] \arrow[d, "U"] & F \\ H \arrow[r, "\mathrm{id}_H"] & H \arrow[u, "V"'] \end{tikzcd}$$

Since H is of type 2, so is id_H, and therefore (6) shows that A is of type 2, too. Analogously, we find that A is of cotype 2. □

Proof of Theorem 6: If we have an RKHS H with $E \subset H \subset F$, then Lemma 23 shows $E \hookrightarrow H_0 \hookrightarrow F$, and this in turn directly provides a 2-factorization. To show the converse, assume a 2-factorization

$$\begin{tikzcd} E \arrow[rr, hook, "\mathrm{id}"] \arrow[dr, "U"'] & & F \\ & H_0 \arrow[ur, "V"'] & \end{tikzcd}\,.$$

We define the feature map $\phi : X \to H_0$ by

$$\phi(x) \coloneqq V^*\delta_{x,F}\,,$$

where $\delta_{x,F} \in F'$ denotes the evaluation functional at x acting on F and $V^* : F' \to H_0$ is the adjoint operator of V that is uniquely determined by

$$\langle V^* f', h \rangle_{H_0} = \langle f', Vh \rangle_{F',F} , \qquad\qquad h \in H_0, f' \in F'.$$

The corresponding kernel $k : X \times X \to \mathbb{R}$ is

$$k(x, x') := \langle \phi(x), \phi(x') \rangle_{H_0}$$

and by [39, Thm. 4.21] its RKHS is given by

$$H = \left\{ \langle h, \phi(\cdot) \rangle_{H_0} \,\middle|\, h \in H_0 \right\} . \tag{32}$$

Our next goal is to show $H = \operatorname{ran} V$. To this end, we first observe that for $h \in H_0$ and $x \in X$ we have

$$\langle h, \phi(x) \rangle_{H_0} = \langle h, V^* \delta_{x,F} \rangle_{H_0} = \langle V^* \delta_{x,F}, h \rangle_{H_0} = \langle \delta_{x,F}, Vh \rangle_{F',F} = (Vh)(x) .$$

Now, if choose an $f \in H$, then by (32) there exists an $h \in H_0$ with $f = \langle h, \phi(\cdot) \rangle_{H_0}$, and hence our observation yields $f = Vh \in \operatorname{ran} V$. Conversely, if we fix an $f \in \operatorname{ran} V$, then there exists an $h \in H_0$ with $Vh = f$, and our observation gives $f = \langle h, \phi(\cdot) \rangle_{H_0} \in H$ by (32). We conclude that $H = \operatorname{ran} V$, which in particular implies $H \subset F$.

To establish $E \subset H$, we simply observe that the 2-factorization together with $H = \operatorname{ran} V$ yields $E = VUE \subset \operatorname{ran} V = H$. □

Proof of Theorem 7: *i)* By Pietsch's factorization theorem, see e.g. [11, Cor. 2.16], we know that $\mathrm{id} : E \to F$ is 2-factorable, and hence the assertion follows by Theorem 6.

ii) By Kwapien's theorem, see e.g. [11, Thm. 12.19], we know that $\mathrm{id} : E \to F$ is 2-factorable, and therefore the assertion again follows by Theorem 6. □

Proof of Corollary 9: By Theorem 6 we know that $\mathrm{id} : E \to F$ is 2-factorable, and hence the assertion follows by Lemma 5. □

5.2 Proofs for Section 4

5.2.1 Proofs for Section 4.2

Before we present the proofs of Section 4.2, we recall that given a metric space (X, d) and some $\alpha \in (0, 1]$, the map $d^\alpha : X \times X \to [0, \infty)$ is another metric on X, which generates the same topology as d does. Moreover, an $f : X \to \mathbb{R}$ is α-Hölder continuous with respect to d, if and only if f is 1-Hölder continuous with respect to d^α, and we have

$$\|f\|_{\mathrm{Höl}_d^\alpha} = \|f\|_{\mathrm{Höl}_{d^\alpha}^1} . \tag{33}$$

Finally, an elementary calculation shows $N_{d^\alpha}^{\mathrm{pack}}(X, \delta) = N_d^{\mathrm{pack}}(X, \delta^{1/\alpha})$ for all $\delta > 0$.

We construct suitable functions with disjoint support, for which in this section a cotype 2 argument yields the results. To this end, consider, for fixed center $t \in X$, radius $\delta > 0$, and $\alpha \in (0, 1]$ the bump function $f_{t,\delta,\alpha} : X \to \mathbb{R}$ defined by

$$f_{t,\delta,\alpha}(s) := (\delta - d^\alpha(t, s))_+ , \qquad\qquad s \in X. \tag{34}$$

Now, the basic idea is to consider centers $t_1, \ldots, t_n$ and a radius δ, for which the corresponding functions $f_{t_i,\delta,\alpha}$ have disjoint support, since in this case the Rademacher norms can be suitably controlled as the following result shows.

Proposition 25. *Let (X,d) be a metric space and $0<\beta\leq\alpha\leq 1$. Moreover, let $t_1,\dots,t_n\in X$ and $0<\delta\leq 1$. Then the following statements hold true for the bump functions $f_{t_i,\delta,\alpha}$ defined in* (34)*:*

i) We have $\operatorname{supp} f_{t_i,\delta,\alpha}\subset B(t_i,\delta^{1/\alpha})$ *and* $\|f_{t_i,\delta,\alpha}\|_\infty=\delta$*.*

ii) If $d^\alpha(t_i,t_j)\geq 3\delta$ holds for all $i\neq j$, then, for all $\varepsilon_1,\dots,\varepsilon_n\in\{-1,1\}$, we have

$$\Big\|\sum_{i=1}^n\varepsilon_i f_{t_i,\delta,\alpha}\Big\|_{\mathrm{Höl}_d^\alpha}\leq 1\,.$$

iii) If there exist $s_1,\dots,s_n\in X$ with $d(s_i,t_i)=\delta^{1/\alpha}$, then $\|f_{t_i,\delta,d^\alpha}\|_{\mathrm{Höl}_d^\beta}\geq\delta^{(\alpha-\beta)/\alpha}$ for all $i=1,\dots,n$.

Proof of Proposition 25. i) The construction of $f_{t_i,\delta,\alpha}$ gives $\{f_{t_i,\delta,\alpha}\neq 0\}\subset B(t_i,\delta^{1/\alpha})$, and since the latter ball is closed, we obtain the first assertion. The second assertion is trivial.

ii) To emphasize the role of the metric in the definition of the bump function, we write $f_{t_i,\delta,\alpha,d}\coloneqq f_{t_i,\delta,\alpha}$ for a moment. By switching to the metric d^α we then have $f_{t_i,\delta,\alpha,d}=f_{t_i,\delta,1,d^\alpha}$. Consequently, Eq. (33) shows

$$\Big\|\sum_{i=1}^n\varepsilon_i f_{t_i,\delta,\alpha}\Big\|_{\mathrm{Höl}_d^\alpha}=\Big\|\sum_{i=1}^n\varepsilon_i f_{t_i,\delta,\alpha,d}\Big\|_{\mathrm{Höl}_d^\alpha}=\Big\|\sum_{i=1}^n\varepsilon_i f_{t_i,\delta,1,d^\alpha}\Big\|_{\mathrm{Höl}_{d^\alpha}^1}\,.$$

Without loss of generality, it thus suffices to investigate the case $\alpha=1$ for

$$f\coloneqq\sum_{i=1}^n\varepsilon_i f_{t_i,\delta,1}\,.$$

Here we first note that *i)* together with our assumption on $t_1,\dots,t_n$ implies $d(\operatorname{supp} f_{t_i,\delta,1},\operatorname{supp} f_{t_j,\delta,1})\geq\delta$ for all $i\neq j$. This gives $\|f\|_\infty\leq\delta\leq 1$ by *i)*. It remains to show

$$|f(x)-f(y)|\leq d(x,y)\,,\qquad\qquad x,y\in X\,. \tag{35}$$

To this end, we define the "center identifying function" $\iota: X\to\{0,\dots,n\}$ by

$$\iota(x)=\begin{cases}i\,, & \text{if there exists an } i\in\{1,\dots,n\} \text{ such that } d(x,t_i)<\delta\\ 0\,, & \text{if } d(x,t_i)\geq\delta \text{ for all } i\in\{1,\dots,n\}.\end{cases}$$

With the help of this notation we can express f by

$$f(x)=\begin{cases}\varepsilon_{\iota(x)}\big(\delta-d(t_{\iota(x)},x)\big)\,, & \text{if } \iota(x)>0,\\ 0\,, & \text{if } \iota(x)=0.\end{cases}$$

In the following we thus investigate (35) for the possible values of $\iota(x),\iota(y)$.

Case $\iota(x)=\iota(y)=0$. Trivial.

Case $\iota(x)=\iota(y)>0$. Here we have $|f(x)-f(y)|=\big|d(t_{\iota(x)},x)-d(t_{\iota(x)},y)\big|\leq d(x,y)$.

Case $\iota(y)\neq\iota(x)$ and $\iota(x)=0$. Here, $\iota(x)=0$ implies $\delta\leq d(x,t_{\iota(y)})\leq d(x,y)+d(y,t_{\iota(y)})$, and hence we find $|f(x)-f(y)|=\delta-d(y,t_{\iota(y)})\leq d(x,y)$ as desired.

Case $\iota(x)\neq\iota(y)$ and $\iota(x),\iota(y)>0$. Here we first note that $d(x,y)\geq\delta$. If $\varepsilon_{\iota(x)}=\varepsilon_{\iota(y)}$, we thus find

$$|f(x)-f(y)|=\big|d(y,t_{\iota(y)})-d(x,t_{\iota(x)})\big|\leq\max\{d(y,t_{\iota(y)}),d(x,t_{\iota(x)})\}\leq\delta\leq d(x,y)\,.$$

Moreover, if $\varepsilon_{\iota(x)} \neq \varepsilon_{\iota(y)}$, we obtain

$$|f(x) - f(y)| = \delta - d(x, t_{\iota(x)}) + \delta - d(y, t_{\iota(y)}) \leq 3\delta - d(x, t_{\iota(x)}) - d(y, t_{\iota(y)}) \leq d(x, y)\,,$$

where in the last step we used $3\delta \leq d(t_{\iota(x)}, t_{\iota(y)}) \leq d(x, t_{\iota(x)}) + d(x, y) + d(y, t_{\iota(y)})$.

iii) For simplicity, we write $s \coloneqq s_i$ and $t \coloneqq t_i$. Then we have

$$\|f_{t,\delta,\alpha}\|_{\mathrm{H\ddot{o}l}^\beta_d} = \|f_{t,\delta,\alpha}\|_{\mathrm{H\ddot{o}l}^1_{d^\beta}} \geq \frac{|f_{t,\delta,\alpha}(t) - f_{t,\delta,\alpha}(s)|}{d^\beta(s,t)} = \frac{|\delta - (\delta - d^\alpha(s,t))|}{d^\beta(s,t)} = \frac{\delta}{\delta^{\beta/\alpha}} = \delta^{(\alpha-\beta)/\alpha}\,,$$

where in the first step we used (33). $\square$

Lemma 26. *Let (X, d) be a connected metric space, $\delta_0 > 0$, and $t \in X$. If there exists an $s_0 \in X$ with $d(s_0, t) > \delta_0$, then for all $\delta \in (0, \delta_0]$ there there exists an $s \in X$ with $d(s, t) = \delta$.*

Proof of Lemma 26. Let $\delta \in (0, \delta_0)$ and assume that for all $s \in X$ we have $d(s,t) \neq \delta$. Then we obtain a partition

$$X = \{s \in X \mid d(s,t) < \delta\} \cup \{s \in X \mid d(s,t) > \delta\}$$

of X into two non-empty, open sets. This contradicts the connectedness assumption. $\square$

Proof of Theorem 12. If X has only one element, the claim is trivial. Hence, we assume that X has at least two elements, which in turn implies $\operatorname{diam} X > 0$. We define $\delta_0 \coloneqq \min\{1, (\operatorname{diam} X)^\alpha\}/3$.

For $0 < \delta < \delta_0$ there then exist $s, t \in X$ with $d^\alpha(s,t) > 3\delta$, and hence we obtain $n \coloneqq N^{\mathrm{pack}}_{d^\alpha}(X, 3\delta) \geq 2$. Let us fix a 3δ-packing $t_1, \ldots, t_n \in X$. For a fixed $i \in \{1, \ldots, n\}$ and all $j \in \{1, \ldots, n\}$ with $j \neq i$ we then have $d^\alpha(t_j, t_i) \geq 3\delta > \delta$, where we note that $n \geq 2$ ensures that there actually exists such a $j \neq i$. Consequently, Lemma 26 gives an $s_i \in X$ with $d^\alpha(s_i, t_i) = \delta$. For $1 \leq i \leq n$, we write $f_i \coloneqq f_{t_i,\delta,\alpha}$, where $f_{t_i,\delta,\alpha}$ is the bump function defined in (34). Then, by Proposition 25 we have

$$\|(f_1, \ldots, f_n)\|_{\mathrm{Rad}^n(\mathrm{H\ddot{o}l}^\alpha_d(X))} = \mathbb{E}_{\varepsilon\sim\mathrm{Rad}} \Big\| \sum_{i=1}^n \varepsilon_i f_i \Big\|_{\mathrm{H\ddot{o}l}_{\alpha,d}} \leq 1$$

and

$$\|(f_1, \ldots, f_n)\|_{\ell_2^n(\mathrm{H\ddot{o}l}^\beta_d(X))} = \Big(\sum_{i=1}^n \|f_i\|^2_{\mathrm{H\ddot{o}l}^\beta_d} \Big)^{1/2} \geq n^{1/2} \delta^{(\alpha-\beta)/\alpha}\,.$$

Combining both estimates yields

$$C \coloneqq \|\,\mathrm{H\ddot{o}l}^\alpha_d(X) \hookrightarrow \mathrm{H\ddot{o}l}^\beta_d(X)\|_{\mathrm{cotype}_2} \geq \frac{\|(f_1, \ldots, f_n)\|_{\ell_2^n(\mathrm{H\ddot{o}l}^\beta_d(X))}}{\|(f_1, \ldots, f_n)\|_{\mathrm{Rad}^n(\mathrm{H\ddot{o}l}^\alpha_d(X))}} \geq n^{1/2}\delta^{(\alpha-\beta)/\alpha}.$$

Now Corollary 9 shows $C < \infty$, and using the definition of n we get

$$N^{\mathrm{pack}}_d(X, (3\delta)^{1/\alpha}) = N^{\mathrm{pack}}_{d^\alpha}(X, 3\delta) = n \leq C^2 \delta^{2(\alpha-\beta)/\alpha}\,,$$

where the first identity was already mentioned around (33). A simple variable transformation then yields the assertion for all $\delta < \min\{1, \operatorname{diam} X\}$.

To establish the second assertion, we note that even without the connectivity assumption

$$\|(f_1, \dots, f_n)\|_{\ell_2^n(\ell_\infty(X))} = \Big(\sum_{i=1}^n \|f_i\|_{\ell_\infty(X)}^2\Big)^{1/2} = n^{1/2}\delta$$

holds. Repeating the cotype 2 argument used above with $\beta = 0$ then yields the second assertion. $\square$

Proof of Theorem 13. Since Ω is unbounded we have $N^{\mathrm{pack}}(\Omega, \delta) = \infty$ for all $\delta > 0$: Indeed, if we had $N^{\mathrm{pack}}(\Omega, \delta) < \infty$, then there would be a maximal, finite δ-packing $x_1, \dots, x_n \in \Omega$. For every $x \in \Omega$ there would thus exist an x_i with $d(x, x_i) < \delta$ and $\Omega \subset B(x_1, \delta) \cup \dots \cup B(x_n, \delta)$ would be bounded, which contradicts our assumption. We now set $\delta := 3$ and choose an infinite δ-packing $x_1, \dots \in \Omega$. Moreover, we consider the bump function $f \in C^\infty(\mathbb{R}^d)$ from Eq. (36) and define $f_i : \Omega \to \mathbb{R}$ by $f_i(x) := f(x - x_i)$. For every $n \in \mathbb{N}$ we observe

$$\|(f_1, \dots, f_n)\|_{\mathrm{Rad}^n(C^\infty(\Omega))} \le \|f\|_{C^\infty(\mathbb{R}^d)} < \infty$$

as the supports of $f_1, \dots, f_n$ are disjoint by construction and $f \in \mathcal{S}(\mathbb{R}^d)$. Furthermore, we have $\|f_i\|_\infty = f(0) = 1$ for all $i = 1, \dots, n$, and hence we obtain $\|(f_1, \dots, f_n)\|_{\ell_2^n(\ell_\infty(\Omega))} = n^{1/2}$. We observe

$$\|C^\infty(\Omega) \hookrightarrow \ell_\infty(\Omega)\|_{\mathrm{cotype}_2} \ge \frac{\|(f_1, \dots, f_n)\|_{\ell_2^n(\ell_\infty(\Omega))}}{\|(f_1, \dots, f_n)\|_{\mathrm{Rad}^n(C^\infty(\Omega))}} \ge \frac{n^{1/2}}{\|f\|_{C^\infty(\mathbb{R}^d)}}$$

for all $n \ge 1$. This implies $\|C^\infty(\Omega) \hookrightarrow \ell_\infty(\Omega)\|_{\mathrm{cotype}_2} = \infty$ and by Corollary 9, there cannot exists an RKHS H with $C^\infty(\Omega) \subset H \subset \ell_\infty(\Omega)$, that is, there is no RKHS H with bounded kernel and $C^\infty(\Omega) \subset H$. $\square$

5.2.2 Proofs for Section 4.3

The proofs of Sections 4.3 and 4.5are based on the same bump function $f : \mathbb{R}^d \to \mathbb{R}$ given by

$$f(x) := \begin{cases} \frac{1}{e}\exp\Big(-\frac{1}{1-\|x\|^2}\Big), & \text{if } x \in B(0,1), \\ 0, & \text{otherwise.} \end{cases} \tag{36}$$

Obviously, we have $\mathrm{supp}(f) \subset B(0,1)$. Moreover, we have $f \in \mathcal{S}(\mathbb{R}^d)$, that is, f is a Schwartz function [1]. Finally, for all multi-indices $\alpha \in \mathbb{N}_0^d$ it holds $\partial_\alpha f \neq 0$.

Lemma 27. *Let $\Omega \subset \mathbb{R}^d$ be open such that $B(0,1) \subset \Omega$. Let $\alpha \in \mathbb{N}_0^d$, and $p \in [1, \infty)$. Moreover, for $\delta \in (0, 1/2]$ and suitable $n \ge 1$ let $x_1, \dots, x_n \in B(0, 1/2)$ be a 3δ-packing. We define $f_i : \Omega \to \mathbb{R}$ by $f_i(x) := f(\delta^{-1}x - \delta^{-1}x_i)$, where $f : \mathbb{R}^d \to \mathbb{R}$ is the bump function from Eq.* (36)*. Moreover, for $\varepsilon_1, \dots, \varepsilon_n \in \{-1, 1\}$ we define $h : \Omega \to \mathbb{R}$ by*

$$h(x) := \sum_{i=1}^n \varepsilon_i f_i(x)\,.$$

Then the functions $f_1, \dots, f_n$ have disjoint support and the function h fulfills

$$\|\partial_\alpha h\|_{L_p(\Omega)} = n^{1/p}\delta^{d/p - |\alpha|_1}\|\partial_\alpha f\|_{L_p(\Omega)}\,.$$

Proof. We define $z_i := \delta^{-1}x_i$. For $x \in B(0,1)$ we then have $\|x+z_i\| \le \|x\| + \|z_i\| \le 1+\delta^{-1}/2 \le \delta^{-1}$, that is $B(0,1) \subset B(-z_i,\delta^{-1})$. Moreover, we observe the identity

$$(\partial_\alpha h)(x) = \delta^{-|\alpha|_1} \sum_{i=1}^n \varepsilon_i \partial_\alpha f\big(\delta^{-1}x - z_i\big)\,, \qquad\qquad x \in \Omega.$$

We define $D_i := \operatorname{supp}\big(x \mapsto \partial_\alpha f\big(\delta^{-1}x - z_i\big)\big)$. By $\operatorname{supp}(\partial_\alpha f) \subset \operatorname{supp}(f) \subset B(0,1)$ and our construction we have $D_i \subset B(x_i,\delta)$, and especially the sets $D_1,\dots,D_n$ are pairwise disjoint. This directly yields

$$\|\partial_\alpha h\|^p_{L_p(\Omega)} = \delta^{-p|\alpha|_1} \int_\Omega \Big|\sum_{i=1}^n \varepsilon_i \partial_\alpha f\big(\delta^{-1}x - z_i\big)\Big|^p dx = \sum_{i=1}^n \delta^{-p|\alpha|_1} \int_\Omega \Big|\varepsilon_i \partial_\alpha f\big(\delta^{-1}x - z_i\big)\Big|^p dx\,.$$

Moreover, using the substitution $y := \delta^{-1}x - z_i$, that is $x = \delta(y+z_i) = \delta y + x_i$, we find

$$\begin{aligned}
\int_\Omega \Big|\varepsilon_i \partial_\alpha f\big(\delta^{-1}x - z_i\big)\Big|^p dx = \delta^d \int_{\mathbb{R}^d} \mathbf{1}_\Omega(\delta y + x_i)|\partial_\alpha f(y)|^p\,dy &= \delta^d \int_{-z_i+\delta^{-1}\Omega} |\partial_\alpha f(y)|^p\,dy \\
&= \delta^d \int_{B(0,1)} |\partial_\alpha f(y)|^p\,dy \\
&= \delta^d \int_\Omega |\partial_\alpha f(y)|^p\,dy\,,
\end{aligned}$$

where in the last two steps we used the inclusions $\operatorname{supp}(\partial_\alpha f) \subset \operatorname{supp}(f) \subset B(0,1)$ and $B(0,1) \subset B(-z_i,\delta^{-1}) \subset -z_i + \delta^{-1}\Omega$ as well as $B(0,1) \subset \Omega$. Combining both calculations then yields the assertion. $\square$

Lemma 28. *Let $\Omega \subset \mathbb{R}^d$ be open such that $B(0,1) \subset \Omega$ and let $p \in [1,\infty)$. Moreover, for $\delta \in (0,1/2]$ and suitable $n \ge 1$ let $x_1,\dots,x_n \in B(0,1/2)$ be a 3δ-packing, and let $f_i : \Omega \to \mathbb{R}$ be defined by*

$$f_i(x) := f(\delta^{-1}x - \delta^{-1}x_i)\,,$$

where f is the bump function from Eq. (36). *Then for all coherent $A \subset \mathbb{N}_0^d$ and $\alpha_0 \in A$ with $s := |\alpha_0|_1 = |A|_1$ we have*

$$\|\partial_{\alpha_0} f\|_{L_p(\Omega)} \cdot n^{1/p}\delta^{d/p-s} \le \|(f_1,\dots,f_n)\|_{\mathrm{Rad}^n(W_p^A(\Omega))} \le \sup_{\alpha\in A} \|\partial_\alpha f\|_{L_p(\Omega)} \cdot n^{1/p}\delta^{d/p-s}\,, \tag{37}$$

and

$$\|\partial_{\alpha_0} f\|_{L_p(\Omega)} \cdot n^{1/2}\delta^{d/p-s} \le \|(f_1,\dots,f_n)\|_{\ell_2^n(W_p^A(\Omega))} \le \sup_{\alpha\in A} \|\partial_\alpha f\|_{L_p(\Omega)} \cdot n^{1/2}\delta^{d/p-s}\,.$$

Proof. From Lemma 27 we directly obtain

$$\|(f_1,\dots,f_n)\|_{\mathrm{Rad}^n(W_p^A(\Omega))} = \mathbb{E}_{\varepsilon\sim\mathrm{Rad}}\|\varepsilon_1 f_1 + \dots + \varepsilon f_n\|_{W_p^A(\Omega)} = \sup_{\alpha\in A} n^{1/p}\delta^{d/p-|\alpha|_1} \cdot \|\partial_\alpha f\|_{L_p(\Omega)}\,.$$

Now the lower bound on $\|(f_1,\dots,f_n)\|_{\mathrm{Rad}^n(W_p^A(\Omega))}$ is obvious, and the upper bound quickly follows by remembering $\delta \le 1/2$.

Moreover, applying Lemma 27 in the case $n = 1$ shows $\|\partial_\alpha f_i\|_{L_p(\Omega)} = \delta^{d/p-|\alpha|_1}\|\partial_\alpha f\|_{L_p(\Omega)}$ for all $i = 1,\dots,n$, which in turn implies

$$\|f_i\|_{W_p^A(\Omega)} = \sup_{\alpha\in A} \delta^{d/p-|\alpha|_1}\|\partial_\alpha f\|_{L_p(\Omega)}\,.$$

Hence we obtain

$$\|(f_1,\dots,f_n)\|_{\ell_2^n(W_p^A(\Omega))} = n^{1/2}\sup_{\alpha\in A} \delta^{d/p-|\alpha|_1}\|\partial_\alpha f\|_{L_p(\Omega)}\,.$$

The rest of the proof is analogous to the proof of (37). □

Proof of Theorem 14. By a simple translation and scaling argument we may assume without loss of generality that $B(0,1)\subset\Omega$. Let us fix a $0<\delta\leq 1/2$. We write $n := N^{\mathrm{pack}}(B(0,1/2),3\delta)$ and choose a 3δ-packing $x_1,\dots,x_n\in B(0,1/2)$. Moreover, we define $f_1,\dots,f_n$ as in Lemma 28. By Lemma 28 there then exists a constant C that is independent on n and δ such that

$$\|W_{p_1}^A(\Omega)\hookrightarrow W_{p_2}^B(\Omega)\|_{\mathrm{type}_2} \geq \frac{\|(f_1,\dots,f_n)\|_{\mathrm{Rad}^n(W_{p_2}^B(\Omega))}}{\|(f_1,\dots,f_n)\|_{\ell_2^n(W_{p_1}^A(\Omega))}} \geq Cn^{1/p_2-1/2}\delta^{s-t+d(1/p_2-1/p_1)}\,.$$

Using the packing number bound (8) we thus find some constant $\widehat{C}>0$ such that

$$\|W_{p_1}^A(\Omega)\hookrightarrow W_{p_2}^B(\Omega)\|_{\mathrm{type}_2} \geq \widehat{C}\delta^{s-t+d(1/2-1/p_1)}\,.$$

holds for all $\delta\in(0,1/2]$. Now Lemma 5 implies that $s-t\geq(d/p_1-d/2)_+$ holds. Using a cotype 2 argument, we analogously obtain the requirement $s-t\geq(d/2-d/p_2)_+$.

It remains to show that $s-t\geq(d/p_1-d/2)_+ + (d/2-d/p_2)_+$. However, if $(d/p_1-d/2)_+ = 0$ or $(d/2-d/p_2)_+ = 0$, then the claim immediately follows from our previous considerations. Moreover, if both are positive, we have

$$(d/p_1-d/2)_+ + (d/2-d/p_2)_+ = d(1/p_1-1/p_2)\,,$$

and the claim holds by assumption. □

Proof of Theorem 15. Recall that if there exists an RKHS H with $W_p^A(\Omega)\subset H\subset\ell_\infty(\Omega)$, then $W_p^A(\Omega)\hookrightarrow\ell_\infty(\Omega)$ is 2-factorable, and therefore of type 2 and cotype 2.

By a simple translation and scaling argument we may assume without loss of generality that $B(0,1)\subset\Omega$. Let us fix a $0<\delta\leq 1/2$. We write $n := N^{\mathrm{pack}}(B(0,1/2),3\delta)$ and choose a 3δ-packing $x_1,\dots,x_n\in B(0,1/2)$. For the functions $f_1,\dots,f_n$ considered in Lemma 27 we then have

$$\|(f_1,\dots,f_n)\|_{\mathrm{Rad}^n(\ell_\infty(\Omega))} = \mathbb{E}_{\varepsilon\sim\mathrm{Rad}}\|\varepsilon_1 f_1+\cdots+\varepsilon_n f_n\|_{\ell_\infty(\Omega)} = \|f_1\|_\infty = 1,$$

since the functions $f_1,\dots,f_n$ are disjointedly supported translated copies of each other. Lemmas 27 and 28 yield

$$\|W_p^A(\Omega)\hookrightarrow\ell_\infty(\Omega)\|_{\mathrm{type}_2} \geq \frac{\|(f_1,\dots,f_n)\|_{\mathrm{Rad}^n(\ell_\infty(\Omega))}}{\|(f_1,\dots,f_n)\|_{\ell_2^n(W_p^A(\Omega))}} \geq Cn^{-1/2}\delta^{s-d/p}$$

for some suitable constant $C>0$, and using the packing number bound (8) Lemma 5 gives $s\geq(d/p-d/2)_+$.

Analogously, a cotype 2 argument shows $s\geq d/2$. We conclude the claim by considering the cases $(d/p_1-d/2)_+ = 0$ and $(d/p_1-d/2)_+ = d(1/p-1/2)$. □

5.2.3 Proofs for Section 4.4

Proof of Proposition 17. Let $\mu \in \mathcal{M}$ and denote $|\mu| := \mu_+ - \mu_-$ for its total variation. For $x \in x$ we define $c_x := \sup_{y\in X}|\Psi(x,y)\beta(y)| < \infty$ and observe

$$|f_\mu(x)| \leq \int_X |\Psi(x,y)\beta(y)|\, d|\mu|(y) \leq c_x\|\mu\|_{\mathrm{TV}} \leq c_x\|\mu\|_{\mathcal{M}}\|\mathrm{id} : \mathcal{M} \to \mathcal{M}(X)\| < \infty. \tag{38}$$

Clearly, $E_{\mathcal{M},\Psi,\beta}$ is a vector space of functions $X \to \mathbb{R}$ and $\|\cdot\|_{E_{\mathcal{M},\Psi,\beta}}$ defines a norm on $E_{\mathcal{M},\Psi,\beta}$. We show that $(E_{\mathcal{M},\Psi,\beta}, \|\cdot\|_{E_{\mathcal{M},\Psi,\beta}})$ is complete. To this end, define the subspace $\mathcal{N} := \{\nu \in \mathcal{M} \mid f_\nu = 0\} \subset \mathcal{M}$ and observe that $\mathcal{N}$ is closed in $\mathcal{M}$ as

$$\mathcal{N} = \bigcap_{x\in X} \ker \phi_x$$

holds. Therefore, the quotient space $\mathcal{M}/\mathcal{N}$ is a Banach space. By construction, the operator $A : \mathcal{M}/\mathcal{N} \to E_{\mathcal{M},\Psi,\beta}$ given by $[\mu]_\sim \mapsto f_\mu$ is well-defined and bijective. Furthermore, A is an isometry since for all $\mu \in \mathcal{M}$ we have

$$\|f_\mu\|_{E_{\mathcal{M},\Psi,\beta}} = \inf\{\|\nu\|_{\mathcal{M}} \mid \nu \in \mathcal{M}, f_\nu = f_\mu\} = \inf\{\|\nu\|_{\mathcal{M}} \mid \nu \in \mathcal{M}, \nu - \mu \in \mathcal{N}\} = \big\|[\mu]_{\mathcal{M}/\mathcal{N}}\big\|_{\mathcal{M}/\mathcal{N}}.$$

Since $\mathcal{M}/\mathcal{N}$ is a Banach space, we hence conclude that $(E_{\mathcal{M},\Psi,\beta}, \|\cdot\|_{E_{\mathcal{M},\Psi,\beta}})$ is a Banach space, too.

Finally, we show that $E_{\mathcal{M},\Psi,\beta}$ is a proper BSF: Let $x \in X$ and $f_\mu \in E_{\mathcal{M},\Psi,\beta}$. By (38) we then have

$$|f_\mu(x)| \leq c_x\|\mathrm{id} : \mathcal{M} \to \mathcal{M}(X)\| \cdot \|\nu\|_{\mathcal{M}}$$

for all $\nu \in \mathcal{M}$ with $f_\nu = f_\mu$. This leads to

$$|f_\mu(x)| \leq c_x\|\mathrm{id} : \mathcal{M} \to \mathcal{M}(X)\| \cdot \|f_\mu\|_{E_{\mathcal{M},\Psi,\beta}}$$

by the definition of the norm $\|\cdot\|_{E_{\mathcal{M},\Psi,\beta}}$, namely $\|f_\mu\|_{E_{\mathcal{M},\Psi,\beta}} = \inf\{\|\nu\|_{\mathcal{M}} \mid \nu \in \mathcal{M}, f_\nu = f_\mu\}$. □

Proof of Lemma 18. Since Ψ is measurable and bounded, Proposition 17 shows that $E_{\mathcal{M},\Psi,1}$ is a proper BSF. By [4, Sect. I.6], the kernel $k = k_1 + k_2$ indeed corresponds to the RKHS $H_1 + H_2$.

For $i \in \{1,2\}$ and $\mu \in \mathcal{M}$, our assumption (22) ensures that the map $x \mapsto k_i(\,\cdot\,, x)$ is weakly μ-integrable with Pettis-integral $\int k_i(\,\cdot\,, y)\, d\mu(y) \in H_i$, see [40, Section 2] for details. For all $x \in X$, this yields

$$\begin{aligned}\Big(\int k_i(\,\cdot\,, y)\, d\mu(y)\Big)(x) = \Big\langle k_i(\,\cdot\,, x), \int k_i(\,\cdot\,, y)\, d\mu(y)\Big\rangle_{H_i} &= \int \langle k_i(\,\cdot\,, x), k_i(\,\cdot\,, y)\rangle_{H_i}\, d\mu(y)\\ &= \int k_i(x,y)\, d\mu(y)\\ &= f_{\mu,k_i,1}(x)\,,\end{aligned}$$

and hence we find $f_{\mu,k_i,1} \in H_i$. We conclude $f_{\mu,\Psi,1} = f_{\mu,k_1,1} - f_{\mu,k_2,1} \in H_1 + H_2 = H$ and $E_{\mathcal{M},\Psi,1} \subset H$. Since both $E_{\mathcal{M},\Psi,1}$ and H are proper BSF, Proposition 3 yields the continuous embedding $E_{\mathcal{M},\Psi,1} \hookrightarrow H$. □

5.2.4 Proofs for Section 4.5

For the definition of Slobodeckij spaces, we essentially follow [29, 2.1.2], but we note that our version of the Slobodeckij norm is actually only norm-equivalent to the norm given there. As discussed previously, neither our Question (1) nor 2-factorability is affected by switching to equivalent norms.

To begin with, let $\Omega \subset \mathbb{R}^d$ be a domain with smooth boundary and let $s \geq 0, p \geq 1$. For the coherent set of multi-indices $A := \{\alpha \in \mathbb{N}_0^d \mid |\alpha|_1 \leq \lfloor s \rfloor\}$, let $W^s_{\text{aux}}(\Omega)$ be the set of A-times weakly differentiable functions defined in Eq. (15). If $s \in \mathbb{N}$, then we define for $f \in W^s_{\text{aux}}(\Omega)$ the Sobolev norm just as in Eq. (16), that is

$$\|f\|_{W_p^s(\Omega)} := \sup_{\alpha \in \mathbb{N}_0^d, |\alpha|_1 \leq s} \|\partial_\alpha f\|_{L_p(\Omega)}\,.$$

Recall that the classical Sobolev norm considers the p-sum of the involved terms $\|\partial_\alpha f\|_{L_p(\Omega)}$ instead. Our definition gives, of course, an equivalent norm.

Moreover, if $s \notin \mathbb{N}$, we write $\theta := s - \lfloor s \rfloor \in (0,1)$ and define, for $f \in W^s_{\text{aux}}(\Omega)$, the Slobodeckij norm by

$$\|f\|_{W_p^s(\Omega)} := \max\Big\{ \sup_{\alpha \in \mathbb{N}_0^d, |\alpha|_1 \leq \lfloor s \rfloor} \|\partial_\alpha f\|_{L_p(\Omega)}, \sup_{\alpha \in \mathbb{N}_0^d, |\alpha|_1 = \lfloor s \rfloor} \|\partial_\alpha f\|_{\theta,p,\Omega} \Big\},$$

where for a measurable function $g : \Omega \to \mathbb{R}$ the semi-norm $\|g\|_{\theta,p,\Omega}$ is defined as

$$\|g\|_{\theta,p,\Omega} := \Big(\int_\Omega \int_\Omega \frac{|g(x) - g(y)|^p}{\|x - y\|^{\theta p + d}}\, dy\, dx \Big)^{1/p}. \tag{39}$$

In both cases, we define the Slobodeckij space by

$$W_p^s(\Omega) := \Big\{ f \in W^s_{\text{aux}}(\Omega) : \|f\|_{W_p^s(\Omega)} < \infty \Big\}\,.$$

To bound the semi-norm $\|g\|_{\theta,p,\Omega}$, the following well-known formula, see e.g. [14, Satz 14.8], which holds for all measurable functions $f : [0,\infty) \to [0,\infty)$ and all $0 \leq a < b \leq \infty$, turns out to be useful:

$$\int_{\mathbb{R}^d} \mathbf{1}_{[a,b]}(\|x\|) f(\|x\|)\, dx = dV_d \int_a^b f(r) r^{d-1}\, dr\,. \tag{40}$$

Here $V_d := \operatorname{vol}(B(0,1))$ is the Volume of the d-dimensional unit ball. For later use we further note that for $z \in \mathbb{R}^d$ and $a \geq 1$, $\beta > 0$ this formula yields

$$\begin{aligned}
\int_{B(z,a)} \int_{\mathbb{R}^d \setminus B(z,a+1/2)} \frac{1}{\|x-y\|^{\beta+d}}\, dy\, dx &= \int_{B(0,a)} \int_{\mathbb{R}^d \setminus B(0,a+1/2)} \frac{1}{\|x-y\|^{\beta+d}}\, dy\, dx \\
&\leq \int_{B(0,a)} \int_{\mathbb{R}^d \setminus B(0,a+1/2)} \big|\|x\| - \|y\|\big|^{-\beta-d}\, dy\, dx \\
&\leq \int_{B(0,a)} \int_{\mathbb{R}^d \setminus B(0,a+1/2)} \big(\|y\| - a\big)^{-\beta-d}\, dy\, dx \\
&\leq a^d d V_d^2 \int_{a+1/2}^{\infty} (r-a)^{-\beta-d} r^{d-1}\, dr \\
&< \frac{2^{\beta+d} a^{2d} d V_d^2}{\beta}\,,
\end{aligned} \tag{41}$$

where in the last step we used $r < 2a(r-a)$ for all $r \geq a + 1/2$.

Lemma 29. *For all $d \geq 1$, $p \in [1, \infty)$, and $\theta \in (0,1)$, there exists a constant $c_{d,p,\theta} \geq 0$ such that for all open and bounded $\Omega \subset \mathbb{R}^d$ and all $f : \Omega \to \mathbb{R}$ that are continuously differentiable with bounded derivative, we have*

$$\|f\|_{\theta,p,\Omega} \leq c_{d,p,\theta} |f|_1 (\mathrm{vol}(\Omega))^{1/p} (\operatorname{diam} \Omega)^{1-\theta} .$$

Proof. Since the derivative of f is bounded, f is Lipschitz continuous with constant $|f|_1 < \infty$. We write $\delta \coloneqq \operatorname{diam} \Omega$. For fixed $x \in \Omega$ we then have $\|x - y\| \leq \delta$ for all $y \in \Omega$, that is $\Omega \subset B(x, \delta)$. This yields

$$\begin{aligned}
\|f\|_{\theta,p,\Omega}^p = \int_\Omega \int_\Omega \frac{|f(x) - f(y)|^p}{\|x - y\|^{\theta p + d}} \, dy \, dx &\leq |f|_1^p \int_\Omega \int_\Omega \frac{1}{\|x - y\|^{(\theta-1)p+d}} \, dy \, dx \\
&\leq |f|_1^p \int_\Omega \int_{B(x,\delta)} \frac{1}{\|x - y\|^{(\theta-1)p+d}} \, dy \, dx \\
&= |f|_1^p \int_\Omega \int_{B(0,\delta)} \frac{1}{\|z\|^{(\theta-1)p+d}} \, dz \, dx .
\end{aligned}$$

Moreover, with the help of (40) we obtain

$$\int_{B(0,\delta)} \frac{1}{\|z\|^{(\theta-1)p+d}} \, dz = dV_d \int_0^\delta r^{-1+(1-\theta)p} \, dr = dV_d \frac{\delta^{(1-\theta)p}}{(1-\theta)p} ,$$

and inserting this into the previous estimate then yields the assertion since Ω is bounded. $\square$

Analogous to Lemma 27, we establish the foundation for estimating the type 2 norm and cotype 2 norm of embeddings $W^s_{p_1}(\Omega) \hookrightarrow W^t_{p_2}(\Omega)$.

Lemma 30. *Let $\Omega \subset \mathbb{R}^d$ be open such that $B(0,1) \subset \Omega$, and let $\alpha \in \mathbb{N}_0^d$, $\theta \in (0,1)$, and $p \in [1, \infty)$. Moreover, for $\delta \in (0, 1/2]$ and suitable $n \geq 1$ let $x_1, \dots, x_n \in B(0, 1/2)$ be a 3δ-packing, and $\varepsilon_1, \dots, \varepsilon_n \in \{-1, 1\}$. We define $h : \Omega \to \mathbb{R}$ by*

$$h(x) \coloneqq \sum_{i=1}^n \varepsilon_i f(\delta^{-1} x - \delta^{-1} x_i) ,$$

where f is the bump function from Eq. (36). Moreover, we write $s \coloneqq \theta + |\alpha|_1$. Then we have $\|\partial_\alpha f\|_{\theta,p,B(0,1)} > 0$ and there exists a constant $c_{d,p,\theta} > 0$ only depending on d, p, θ, such that

$$\|\partial_\alpha f\|_{\theta,p,B(0,1)} n^{1/p} \delta^{d/p - s} \leq \|\partial_\alpha h\|_{\theta,p,\Omega} \leq c_{d,p,\theta} \big(1 + \|\partial_\alpha f\|_{\theta,p,B(0,3/2)}\big) n^{1/p} \delta^{d/p - s} .$$

Proof of Lemma 30. Before we begin with the actual proof, we note that for any suitable set $\Theta \subset \mathbb{R}^d$ and any sufficiently differentiable function $a : \Theta \to \mathbb{R}$ one has $\mathrm{supp}(\partial_\alpha a) \subset \mathrm{supp}(a)$, for any suitable $\alpha \in \mathbb{N}_0^d$. This inclusion will be used several times.

Just as in the proof of Lemma 27 we define $z_i \coloneqq \delta^{-1} x_i \in B(0, \delta^{-1}/2) \subset \delta^{-1}\Omega$. Note that we have $\|z_i - z_j\| \geq 3$ for $i \neq j$ as well as the identity

$$(\partial_\alpha h)(x) = \delta^{-|\alpha|_1} \sum_{i=1}^n \varepsilon_i \partial_\alpha f\big(\delta^{-1} x - z_i\big) .$$

We define $f_i \coloneqq \varepsilon_i \partial_\alpha f(\,\cdot\, - z_i)$ and $g \coloneqq \sum_{i=1}^n f_i$. For later use we note that $\operatorname{supp} f_i \subset B(z_i, 1)$, and in particular the functions $f_1, \dots, f_n$ have mutually disjoint support. Moreover, by substitution we obtain

$$\begin{aligned}
\|\partial_\alpha h\|_{\theta,p,\Omega}^p &= \int_\Omega \int_\Omega \frac{|\partial_\alpha h(x) - \partial_\alpha h(y)|^p}{\|x-y\|^{\theta p + d}} \, dy \, dx \\
&= \delta^{-p|\alpha|_1} \int_\Omega \int_\Omega \frac{\bigl|\sum_{i=1}^n \varepsilon_i \partial_\alpha f\bigl(\delta^{-1}x - z_i\bigr) - \sum_{j=1}^n \varepsilon_j \partial_\alpha f\bigl(\delta^{-1} y - z_j\bigr)\bigr|^p}{\|x-y\|^{\theta p + d}} \, dy \, dx \\
&= \delta^{2d - p|\alpha|_1} \int_{\delta^{-1}\Omega} \int_{\delta^{-1}\Omega} \frac{\bigl|g(x) - g(y)\bigr|^p}{\|\delta x - \delta y\|^{\theta p + d}} \, dy \, dx \\
&= \delta^{d - ps} \int_{\delta^{-1}\Omega} \int_{\delta^{-1}\Omega} \frac{\bigl|g(x) - g(y)\bigr|^p}{\|x-y\|^{\theta p + d}} \, dy \, dx \, . \qquad (42)
\end{aligned}$$

Our first goal is to establish the upper bound on $\|\partial_\alpha h\|_{\theta,p,\Omega}^p$. To this end, we write $H(x,x) \coloneqq 0$ and

$$H(x,y) \coloneqq \frac{|g(x) - g(y)|^p}{\|x-y\|^{\theta p + d}} \, , \qquad x, y \in \mathbb{R}^d \text{ with } x \neq y.$$

Moreover, we write $B_i \coloneqq B(z_i, 3/2)$. Then, we have $\operatorname{supp} f_i \subset B_i$ and $\lambda^d(B_i \cap B_j) = 0$ for all $i \neq j$. Finally, we define $B_0 \coloneqq \mathbb{R}^d \setminus (B_1 \cup \dots \cup B_n)$. Since we have $H(x,y) = H(y,x) \geq 0$ for all $x, y \in \mathbb{R}^d$ and $H(x,y) = 0$ for all $x, y \in B_0$ we then obtain

$$\begin{aligned}
\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} H(x,y) \, dy \, dx &= \sum_{i=0}^n \sum_{j=0}^n \int_{B_i} \int_{B_j} H(x,y) \, dy \, dx \\
&= \sum_{i=1}^n \int_{B_i} \int_{B_i} H(x,y) \, dy \, dx + \sum_{i=0}^n \sum_{j \neq i} \int_{B_i} \int_{B_j} H(x,y) \, dy \, dx \\
&= n \|\partial_\alpha f\|_{\theta,p,B(0,3/2)} + \sum_{i=0}^n \sum_{j \neq i} \int_{B_i} \int_{B_j} H(x,y) \, dy \, dx \, ,
\end{aligned}$$

where in the last step we used

$$\begin{aligned}
\int_{B_i} \int_{B_i} H(x,y) \, dy \, dx &= \int_{B(z_i,3/2)} \int_{B(z_i,3/2)} \frac{|\partial_\alpha f(x - z_i) - \partial_\alpha f(y - z_i)|^p}{\|x-y\|^{\theta p + d}} \, dy \, dx \\
&= \int_{B(0,3/2)} \int_{B(0,3/2)} \frac{|\partial_\alpha f(x) - \partial_\alpha f(y)|^p}{\|x-y\|^{\theta p + d}} \, dy \, dx \\
&= \|\partial_\alpha f\|_{\theta,p,B(0,3/2)}^p \, .
\end{aligned}$$

To treat the double sum, we first observe

$$\begin{aligned}
\sum_{i=0}^n \sum_{j \neq i} \int_{B_i} \int_{B_j} H(x,y) \, dy \, dx &= \sum_{j=1}^n \int_{B_0} \int_{B_j} H(x,y) \, dy \, dx + \sum_{i=1}^n \sum_{j \neq i} \int_{B_i} \int_{B_j} H(x,y) \, dy \, dx \\
&\leq \sum_{j=1}^n \int_{\mathbb{R}^d \setminus B_j} \int_{B_j} H(x,y) \, dy \, dx + \sum_{i=1}^n \int_{B_i} \int_{\mathbb{R}^d \setminus B_i} H(x,y) \, dy \, dx
\end{aligned}$$

$$= 2\sum_{i=1}^{n}\int_{B_i}\int_{\mathbb{R}^d\setminus B_i} H(x,y)\,dy\,dx\,.$$

In addition, we have

$$\begin{aligned}\int_{B_i}\int_{\mathbb{R}^d\setminus B_i} H(x,y)\,dy\,dx &= \int_{B(z_i,1)}\int_{\mathbb{R}^d\setminus B_i} H(x,y)\,dy\,dx \\ &\quad + \int_{B_i\setminus B(z_i,1)}\int_{\mathbb{R}^d\setminus B(z_i,2)} H(x,y)\,dy\,dx \\ &\quad + \int_{B_i\setminus B(z_i,1)}\int_{B(z_i,2)\setminus B_i} H(x,y)\,dy\,dx\,.\end{aligned}$$

In the following, we estimate these three double integrals. The first one can be estimated by

$$\begin{aligned}\int_{B(z_i,1)}\int_{\mathbb{R}^d\setminus B_i} H(x,y)\,dy\,dx &= \int_{B(z_i,1)}\int_{\mathbb{R}^d\setminus B(z_i,3/2)} \frac{|g(x)-g(y)|^p}{\|x-y\|^{\theta p+d}}\,dy\,dx \\ &\le \int_{B(z_i,1)}\int_{\mathbb{R}^d\setminus B(z_i,3/2)} \frac{2^p\|g\|_\infty^p}{\|x-y\|^{\theta p+d}}\,dy\,dx \\ &\le \frac{2^{p(1+\theta)+d} d V_d^2}{\theta p}\,,\end{aligned}$$

where we used $\|g\|_\infty = \|f\|_\infty = 1$ and (41). The second double integral can be analogously estimated by

$$\begin{aligned}\int_{B_i\setminus B(z_i,1)}\int_{\mathbb{R}^d\setminus B(z_i,2)} H(x,y)\,dy\,dx &= \int_{B_i\setminus B(z_i,1)}\int_{\mathbb{R}^d\setminus B(z_i,2)} \frac{|g(x)-g(y)|^p}{\|x-y\|^{\theta p+d}}\,dy\,dx \\ &= \int_{B_i\setminus B(z_i,1)}\int_{\mathbb{R}^d\setminus B(z_i,2)} \frac{|g(y)|^p}{\|x-y\|^{\theta p+d}}\,dy\,dx \\ &\le \int_{B(z_i,3/2)}\int_{\mathbb{R}^d\setminus B(z_i,2)} \frac{1}{\|x-y\|^{\theta p+d}}\,dy\,dx \\ &\le \frac{2^{\theta p+3d} d V_d^2}{\theta p}\,.\end{aligned}$$

For the third double integral we observe that for $x \in B_i \setminus B(z_i,1)$ we have $g(x) = 0$ and for $y \in B(z_i,2)\setminus B_i$ we also have $g(y) = 0$. This shows

$$\int_{B_i\setminus B(z_i,1)}\int_{B(z_i,2)\setminus B_i} H(x,y)\,dy\,dx = 0\,.$$

Combining these considerations, we find

$$\|\partial_\alpha h\|_{\theta,p,\Omega}^p \le \delta^{d-ps} n \|\partial_\alpha f\|_{\theta,p,B(0,3/2)}^p + 2\delta^{d-ps} n\Big(\frac{2^{p(1+\theta)+d} d V_d^2}{\theta p} + \frac{2^{\theta p+3d} d V_d^2}{\theta p}\Big)\,.$$

Finally we establish the lower bound for the Slobodeckij norm of h using the fact that $f_1,\dots,f_n$ have mutually disjoint support:

$$\int_{\delta^{-1}\Omega}\int_{\delta^{-1}\Omega} \frac{|g(x)-g(y)|^p}{\|x-y\|^{\theta p+d}}\,dy\,dx \ge \sum_{i=1}^{n}\int_{B(z_i,1)}\int_{B(z_i,1)} \frac{|g(x)-g(y)|^p}{\|x-y\|^{\theta p+d}}\,dy\,dx$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} \int_{B(z_i,1)} \int_{B(z_i,1)} \frac{|\varepsilon_i \partial_\alpha f(x-z_i) - \varepsilon_i \partial_\alpha f(y-z_i)|^p}{\|x-y\|^{\theta p + d}} \, dy \, dx \\
&= n \int_{B(0,1)} \int_{B(0,1)} \frac{|\partial_\alpha f(x) - \partial_\alpha f(y)|^p}{\|x-y\|^{\theta p + d}} \, dy \, dx \\
&= n \|\partial_\alpha f\|_{\theta,p,B(0,1)}^p .
\end{aligned}
$$

Moreover, since $\partial_\alpha f$ is continuous and not constant, we have $\|\partial_\alpha f\|_{\theta,p,B(0,1)} > 0$. □

Now, we are able to estimate the Rademacher sequence norm and the ℓ_2-sequence norm for bump functions in Slobodeckij spaces.

Lemma 31. *Let $\Omega \subset \mathbb{R}^d$ be open such that $B(0,1) \subset \Omega$ and let $s \geq 0$, $p \in [1,\infty)$. For $\delta \in (0,1/2]$ and suitable $n \geq 1$, let $x_1, \ldots, x_n \in B(0,1/2)$ be a 3δ-packing and define $f_i : \Omega \to \mathbb{R}$ by*

$$
f_i(x) \coloneqq f(\delta^{-1}x - \delta^{-1}x_i),
$$

where f is the bump function considered in Eq. (36). Then, there exist constants $c > 0$ and $C > 0$ that are independent of δ and n such that we have

$$
cn^{1/p}\delta^{d/p-s} \leq \|(f_1,\ldots,f_n)\|_{\mathrm{Rad}^n(W_p^s(\Omega))} \leq Cn^{1/p}\delta^{d/p-s}
$$

and

$$
cn^{1/2}\delta^{d/p-s} \leq \|(f_1,\ldots,f_n)\|_{\ell_2^n(W_p^s(\Omega))} \leq Cn^{1/2}\delta^{d/p-s} .
$$

Proof. In the case $s \in \mathbb{N}$, the assertion has already be shown in Lemma 28. In the following we thus assume $s \notin \mathbb{N}$ and define $\theta \coloneqq s - \lfloor s \rfloor$. Finally, we fix some $\alpha_0 \in \mathbb{N}_0^d$ with $|\alpha_0|_1 = \lfloor s \rfloor$.

Now Lemma 30 directly yields the lower bound on the Rademacher norm

$$
\|\partial_{\alpha_0} f\|_{\theta,p,B(0,1)} n^{1/p}\delta^{d/p-s} \leq \mathbb{E}_{\varepsilon\sim\mathrm{Rad}}\|\varepsilon_1 f_1 + \cdots + \varepsilon_n f_n\|_{W_p^s(\Omega)} = \|(f_1,\ldots,f_n)\|_{\mathrm{Rad}^n(W_p^s(\Omega))}.
$$

For the corresponding upper bound we write $C_{\alpha,d,p,\theta} \coloneqq c_{d,p,\theta}(1+\|\partial_\alpha f\|_{\theta,p,B(0,3/2)})$ From Lemmas 27 and 30 we then obtain

$$
\begin{aligned}
\|(f_1,\ldots,f_n)\|_{\mathrm{Rad}^n(W_p^s(\Omega))} &= \mathbb{E}_{\varepsilon\sim\mathrm{Rad}}\|\varepsilon_1 f_1 + \cdots + \varepsilon_n f_n\|_{W_p^s(\Omega)} \\
&\leq \max\{ \sup_{|\alpha|_1 \leq \lfloor s \rfloor} \|\partial_\alpha f\|_{L_p(\Omega)} n^{1/p}\delta^{d/p-|\alpha|_1}, \sup_{|\alpha|_1 = \lfloor s \rfloor} C_{\alpha,d,p,\theta} n^{1/p}\delta^{d/p-s} \}.
\end{aligned}
$$

Since $\delta \leq 1/2$, we see that the asserted upper bound for the Rademacher norm holds for sufficiently large $C > 0$.

Moreover, applying Lemmas 27 and 30 in the case $n = 1$ shows for all $i = 1, \ldots, n$ the estimate

$$
\begin{aligned}
\|\partial_{\alpha_0} f\|_{\theta,p,B(0,1)} \delta^{d/p-s} \leq \|f_i\|_{W_p^s(\Omega)} &\leq \max\{ \sup_{|\alpha|_1 \leq \lfloor s \rfloor} \|\partial_\alpha f\|_{L_p(\Omega)} \delta^{d/p-|\alpha|_1}, \sup_{|\alpha|_1 = \lfloor s \rfloor} \delta^{d/p-s} C_{\alpha,d,p,\theta} \} \\
&\leq C\delta^{d/p-s} .
\end{aligned}
$$

From this we easily obtain the bounds on the ℓ_2-sequence norms. □

The following theorem relates different spaces of Besov-Triebel-Lizorkin type to each other and is at the heart of the constructive statement in Theorem 20.

Theorem 32. *Let $\Omega \subset \mathbb{R}^d$ be a bounded domain with smooth boundary and let $s, t \geq 0$, $p, p_1, p_2 \in [1, \infty]$, and $q, q_1, q_2 \in [1, \infty]$. Then the following statements hold true:*

i) ***Lowering the smoothness to increase the integration index.*** *If $s > t$, then we have*

$$F^s_{p_1,q_1}(\Omega) \hookrightarrow F^t_{p_2,q_2}(\Omega)\,, \qquad \text{if } s - t \geq \frac{d}{p_1} - \frac{d}{p_2} \text{ and } p_1, p_2 < \infty,$$
$$B^s_{p_1,q_1}(\Omega) \hookrightarrow B^t_{p_2,q_2}(\Omega)\,, \qquad \text{if } s - t > \frac{d}{p_1} - \frac{d}{p_2}.$$

ii) ***Lowering the integration index.*** *If $p_1 \geq p_2$, then we have*

$$F^s_{p_1,q}(\Omega) \hookrightarrow F^s_{p_2,q}(\Omega)\,, \qquad \text{if } p_1, p_2, q < \infty,$$
$$B^s_{p_1,q}(\Omega) \hookrightarrow B^s_{p_2,q}(\Omega)\,.$$

iii) ***Changing the fein index.*** *If $q_1 \leq q_2$, then for all $\varepsilon > 0$ we have*

$$F^s_{p,q_1}(\Omega) \hookrightarrow F^s_{p,q_2}(\Omega)\,, \qquad \text{if } p < \infty,$$
$$F^{s+\varepsilon}_{p,\infty}(\Omega) \hookrightarrow F^s_{p,q}(\Omega)\,, \qquad \text{if } p < \infty,$$
$$B^s_{p,q_1}(\Omega) \hookrightarrow B^s_{p,q_2}(\Omega),$$
$$B^{s+\varepsilon}_{p,\infty}(\Omega) \hookrightarrow B^s_{p,q}(\Omega).$$

iv) ***Changing between Besov and Triebel spaces.*** *If $s > t$ and $s - t > d(1/p_1 - 1/p_2)$, then we have*

$$F^s_{p_1,q_1}(\Omega) \hookrightarrow B^t_{p_2,q_2}(\Omega), \qquad \text{if } p_1 < \infty,$$
$$B^s_{p_1,q_1}(\Omega) \hookrightarrow F^t_{p_2,q_2}(\Omega) \qquad \text{if } p_2 < \infty.$$

Proof. i) See e.g. [41, Parts (i) and (ii) of the theorem on p. 196/7].

ii) See e.g. [41, Part (iii) of the theorem on p. 196/7].

iii) In the case $\Omega = \mathbb{R}^d$, these embeddings can be found in e.g. [41, Prop. 2 on p. 47]. The restriction to bounded domains is discussed in e.g. [29, Sec. 2.4.4].

iv) We only show $F^s_{p_1,q_1}(\Omega) \hookrightarrow B^t_{p_2,q_2}(\Omega)$, the other embedding can be proven analogously.

Recall the identity $F^s_{p,p}(\Omega) = B^s_{p,p}(\Omega)$ for $p < \infty$, see Eq. (28). We set $\varepsilon := s - t - d(1/p_1 - 1/p_2) > 0$. By *i)* we obtain $F^s_{p_1,q_1}(\Omega) \hookrightarrow F^{s-\varepsilon/2}_{p_1,p_1}(\Omega) = B^{s-\varepsilon/2}_{p_1,p_1}(\Omega)$ as well as $B^{s-\varepsilon/2}_{p_1,p_1}(\Omega) \hookrightarrow B^t_{p_2,q_2}(\Omega)$. □

Proof of Theorem 20. We first note that Part *i)* of Theorem 32 together with Eq. (27) gives

$$W^s_{p_1}(\Omega) = F^s_{p_1,q_1}(\Omega) \hookrightarrow F^t_{p_2,q_2}(\Omega) = W^t_{p_2}(\Omega)\,, \tag{43}$$

where q_1 and q_2 are defined according to Eq. (27).

i) Let us fix a $u \in (t + (d/2 - d/p_2)_+\,, s - (d/p_1 - d/2)_+\,)$. Then Part *i)* of Theorem 32 shows

$$F^s_{p_1,q_1}(\Omega) \hookrightarrow F^u_{2,2}(\Omega) \hookrightarrow F^t_{p_2,q_2}(\Omega)\,.$$

Combining this with (43) and (25) yields the assertion.

ii) By a translation and scaling argument, we can assume $B(0,1) \subset \Omega$ without loss of generality. Let us fix a $0 < \delta \le 1/2$. We define $n := N^{\mathrm{pack}}(B(0,1/2),\delta)$ and choose $f_1,\dots,f_n$ as in Lemma 31. Then Lemma 31 gives

$$\|W^s_{p_1}(\Omega) \hookrightarrow W^t_{p_2}(\Omega)\|_{\mathrm{type}_2} \ge \frac{\|(f_1,\dots,f_n)\|_{\mathrm{Rad}^n(W^s_{p_1}(\Omega))}}{\|(f_1,\dots,f_n)\|_{\ell_2^n(W^t_{p_2}(\Omega))}} \ge \hat{C} n^{1/p_2-1/2}\delta^{s-t-d(1/p_1-1/p_2))}$$

for constant $\hat{C} > 0$ that is independent of δ and n. Using the packing number bound (8) we thus find a constant $\bar{C} > 0$ such that

$$\|W^s_{p_1}(\Omega) \hookrightarrow W^t_{p_2}(\Omega)\|_{\mathrm{type}_2} \ge \bar{C}\delta^{s-t+d(1/2-1/p_1)}$$

holds for all $\delta \in (0,1/2]$. Now, Lemma 5 in combination with $s > t$ implies $s - t \ge (d/p_1 - d/2)_+$. Using a cotype 2 argument, we analogously obtain the requirement $s - t \ge (d/2 - d/p_2)_+$.

It remains to show that $s - t \ge (d/p_1 - d/2)_+ + (d/2 - d/p_2)_+$ holds. If $(d/p_1 - d/2)_+ = 0$ or $(d/2 - d/p_2)_+ = 0$, this follows directly from our previous considerations. Moreover, if both expressions are positive, we have

$$(d/p_1 - d/2)_+ + (d/2 - d/p_2)_+ = d(1/p_1 - 1/p_2),$$

and the claim holds by assumption. □

The following lemma shows that Slobodeckij spaces are *dense* in the family of Besov-Triebel-Lizorkin spaces in a suitable way.

Lemma 33. *Let $\Omega \subset \mathbb{R}^d$ be a bounded domain with smooth boundary, $s > 0$, and let $X^s_{p,q}(\Omega)$ be a Besov-Triebel-Lizorkin space. Then, for any $\varepsilon_1, \varepsilon_2 > 0$ such that $\min\{1,s\} > \varepsilon_1 > \varepsilon_2$ there exist non-integer smoothness parameters $0 < \check{s} < s < \hat{s}$ and integration parameters $\hat{p},\check{p} \in [1,\infty)$ fulfilling*

$$W^{\hat{s}}_{\hat{p}}(\Omega) \hookrightarrow X^s_{p,q}(\Omega) \hookrightarrow W^{\check{s}}_{\check{p}}(\Omega), \tag{44}$$

and

$$\varepsilon_1 \ge \max\{\hat{s}-s, s-\check{s}\} \ge \min\{\hat{s}-s, s-\check{s}\} \ge \varepsilon_2 \ge \max\{|d/p - d/\hat{p}|, |d/p - d/\check{p}|\}\,. \tag{45}$$

Proof. We choose $\hat{s} \in (s+\varepsilon_2, s+\varepsilon_1) \setminus \mathbb{N}$ and $\check{s} \in (s-\varepsilon_1, s-\varepsilon_2) \setminus \mathbb{N}$ which gives $\hat{s} > s > \check{s}$.

In the case $p < \infty$ we define $\hat{p} = \check{p} = p$. Then, (45) is fulfilled and (27) provides the identities $W^{\hat{s}}_{\hat{p}}(\Omega) = F^{\hat{s}}_{\hat{p},\hat{p}}(\Omega)$ and $W^{\check{s}}_{\check{p}}(\Omega) = F^{\check{s}}_{\check{p},\check{p}}(\Omega)$. Now, in the Triebel-Lizorkin case, part *i)* of Theorem 32 yields

$$F^{\hat{s}}_{\hat{p},\hat{p}}(\Omega) \hookrightarrow F^s_{p,q}(\Omega) \hookrightarrow F^{\check{s}}_{\check{p},\check{p}}(\Omega)\,.$$

Analogously, in the Besov case, part *iv)* of Theorem 32 gives $F^{\hat{s}}_{\hat{p},\hat{p}}(\Omega) \hookrightarrow B^s_{p,q}(\Omega) \hookrightarrow F^{\check{s}}_{\check{p},\check{p}}(\Omega)$.

In the case $p = \infty$ we have $X^s_{p,q}(\Omega) = B^s_{\infty,q}(\Omega)$. Define $\hat{p} = \check{p} := 2d/\varepsilon_2 \in [2,\infty)$. Then, (45) is fulfilled and (44) follows from part *iv)* of Theorem 32, namely $F^{\hat{s}}_{\hat{p},\hat{p}}(\Omega) \hookrightarrow B^s_{\infty,q}(\Omega) \hookrightarrow F^{\check{s}}_{\check{p},\check{p}}(\Omega)$. □

Proof of Theorem 21. By Theorem 32 *i)* and *iv)*, the assumption $s - t > d(1/p_1 - 1/p_2)$ or, if both spaces are of Triebel-Lizorkin type the assumption $s - t \ge d(1/p_1 - 1/p_2)$, ensures that the embedding $X^s_{p_1,q_1}(\Omega) \hookrightarrow Y^t_{p_2,q_2}(\Omega)$ exists.

i) The existence of embeddings $X^s_{p_1,q_1}(\Omega) \hookrightarrow H^u_2(\Omega)$ and $H^u_2(\Omega) \hookrightarrow Y^t_{p_2,q_2}(\Omega)$ constituting the 2-factorization follows directly from Theorem 32 *i)* and *iv)*, using the identity $H^u_2(\Omega) = F^u_{2,2}(\Omega)$, see Eq. (25).

ii) Let the embedding $X^s_{p_1,q_1}(\Omega) \hookrightarrow Y^t_{p_2,q_2}(\Omega)$ be 2-factorable as

$$\begin{tikzcd} X^s_{p_1,q_1}(\Omega) \arrow[rr, hook, "\mathrm{id}"] \arrow[dr, "U"'] & & Y^t_{p_2,q_2}(\Omega) \\ & H \arrow[ur, "V"'] & \end{tikzcd},$$

where H is a Hilbert space and U and V are bounded linear operators.

Choose $t > \varepsilon > 0$. We apply Lemma 33, where we set $\varepsilon_1 := \varepsilon$ and $\varepsilon_2 := \varepsilon/2$, to find Slobodeckij spaces $W^{\hat{s}}_{\hat{p_1}}(\Omega) \hookrightarrow X^s_{p_1,q_1}(\Omega) \hookrightarrow W^{\check{s}}_{\check{p_1}}(\Omega)$ and $W^{\hat{t}}_{\hat{p_2}}(\Omega) \hookrightarrow Y^t_{p_2,q_2}(\Omega) \hookrightarrow W^{\check{t}}_{\check{p_2}}(\Omega)$ with non-integer smoothness parameters $0 < \check{t} \le t < s \le \hat{s}$ and integration parameters $1 \le \hat{p_1}, \check{p_2}$ fulfilling

$$\varepsilon \ge \max\{\hat{s} - s, t - \check{t}\} \ge \min\{\hat{s} - s, t - \check{t}\} \ge \max\{|d/p_1 - d/\hat{p_1}|, |d/p_2 - d/\check{p_2}|\}. \tag{46}$$

We obtain the embedding $W^{\hat{s}}_{\hat{p_1}}(\Omega) \hookrightarrow W^{\check{t}}_{\check{p_2}}(\Omega)$, which can be 2-factorized as

$$\begin{tikzcd} W^{\hat{s}}_{\hat{p_1}}(\Omega) \arrow[r, hook, "\mathrm{id}"] \arrow[drr, "U\circ\mathrm{id}"'] & X^s_{p_1,q_1}(\Omega) \arrow[rr, hook, "\mathrm{id}"] \arrow[dr, "U"] & & Y^t_{p_2,q_2}(\Omega) \arrow[r, hook, "\mathrm{id}"] & W^{\check{t}}_{\check{p_2}}(\Omega) \\ & & H \arrow[ur, "V"] \arrow[urr, "\mathrm{id}\circ V"'] & & \end{tikzcd}.$$

Now, we estimate

$$\begin{aligned} \hat{s} - \check{t} - d(1/\hat{p_1} - 1/\check{p_2}) &= \hat{s} - s + t - \check{t} + d/p_1 - d/\hat{p_1} + d/\check{p_2} - d/p_2 + s - t - d(1/p_1 - 1/p_2) \\ &\ge \hat{s} - s + t - \check{t} - |d/p_1 - d/\hat{p_1}| - |d/p_2 - d/\check{p_2}| + s - t - d(1/p_1 - 1/p_2) \\ &\ge s - t - d(1/p_1 - 1/p_2), \end{aligned}$$

using (46) in the last step. The assumption $s - t \ge d(1/p_1 - 1/p_2)$ yields

$$\hat{s} - \check{t} \ge d(1/\hat{p_1} - 1/\check{p_2})\,,$$

and hence Theorem 20 shows $\hat{s} - \check{t} \ge (d/\hat{p_1} - d/2)_+ + (d/2 - d/\check{p_2})_+$. We now leverage from the general estimate $(a+b)_+ \le a_+ + |b|$, which holds for all $a, b \in \mathbb{R}$, and the inequalities in (46) to estimate

$$\begin{aligned} & s - t - (d/p_1 - d/2)_+ - (d/2 - d/p_2)_+ \\ & \ge \hat{s} - \check{t} - (d/\hat{p_1} - d/2)_+ - (d/2 - d/\check{p_2})_+ - |s - \hat{s}| - \left|\check{t} - t\right| - |d/p_1 - d/\hat{p_1}| - |d/p_2 - d/\check{p_2}| \\ & \ge \hat{s} - \check{t} - (d/\hat{p_1} - d/2)_+ - (d/2 - d/\check{p_2})_+ - 4\varepsilon \\ & \ge -4\varepsilon\,. \end{aligned}$$

Since $\varepsilon > 0$ is arbitrarily small, the claim follows. □

Proof of Theorem 22. We choose $\hat{s} \in (0, s - d/p)$. Then (26) states $B^{\check{s}}_{\infty,\infty}(\Omega) = \mathrm{H\ddot{o}l}^{\check{s}}(\Omega)$ and we clearly have $\mathrm{H\ddot{o}l}^{\check{s}}(\Omega) \hookrightarrow C^0(\Omega)$. Parts *i)* and *iv)* of Theorem 32 yield the existence of the embedding

$$X^s_{p,q}(\Omega) \hookrightarrow B^{\hat{s}}_{\infty,\infty}(\Omega) \hookrightarrow C^0(\Omega).$$

i) Let $u \in (d/2, s - (d/p - d/2)_+)$ and choose $\hat{s} := (u - d/2)/2 > 0$. Then, by $H^u_2(\Omega) = F^u_{2,2}(\Omega)$, see (25), and Theorem 21 we have

$$X^s_{p,q}(\Omega) \hookrightarrow H^u_2(\Omega) \hookrightarrow B^{\hat{s}}_{\infty,\infty}(\Omega) \hookrightarrow C^0(\Omega)$$

and the claim follows.

ii) Let the embedding $X^s_{p,q}(\Omega) \hookrightarrow C^0(\Omega)$ be 2-factorable as

$$\begin{tikzcd}
X^s_{p,q}(\Omega) \arrow[rr, hook, "\mathrm{id}"] \arrow[dr, "U"'] & & C^0(\Omega) \\
& H \arrow[ur, "V"'] &
\end{tikzcd},$$

where H is a Hilbert space and U and V are bounded linear operators. Fix some $\varepsilon > 0$. By Lemma 33 we find non-integer $\hat{s} > s$ and some $\hat{p} \geq 1$ such that

$$\varepsilon > \hat{s} - s \geq |d/p - d/\hat{p}| \tag{47}$$

holds and such that the embedding $W^{\hat{s}}_{\hat{p}}(\Omega) \hookrightarrow X^s_{p,q}(\Omega)$ exists. Especially, the embedding $W^{\hat{s}}_{\hat{p}}(\Omega) \hookrightarrow C^0(\Omega)$ is 2-factorable as

$$\begin{tikzcd}
W^{\hat{s}}_{\hat{p}}(\Omega) \arrow[r, hook, "\mathrm{id}"] \arrow[drr, "U \circ \mathrm{id}"'] & X^s_{p,q}(\Omega) \arrow[rr, hook, "\mathrm{id}"] \arrow[dr, "U"] & & C^0(\Omega) \\
& & H \arrow[ur, "V"'] &
\end{tikzcd}.$$

By a translation and scaling argument, we can now assume $B(0,1) \subset \Omega$ without loss of generality.

Let us fix a $0 < \delta \leq 1/2$. We define $n := N^{\mathrm{pack}}(B(0,1/2), \delta)$ and choose $f_1, \dots, f_n$ as in Lemma 31. Since those functions are disjointedly supported translated copies of each other, we have

$$\|(f_1, \dots, f_n)\|_{\mathrm{Rad}^n(C^0(\Omega))} = \mathbb{E}_{\varepsilon \sim \mathrm{Rad}} \|\varepsilon_1 f_1 + \dots + \varepsilon_n f_n\|_{C^0(\Omega)} = \|f_1\|_\infty = 1.$$

This observation and Lemma 31 give

$$\|W^{\hat{s}}_{\hat{p}}(\Omega) \hookrightarrow C^0(\Omega)\|_{\mathrm{type}_2} \geq \frac{\|(f_1, \dots, f_n)\|_{\mathrm{Rad}^n(C^0(\Omega))}}{\|(f_1, \dots, f_n)\|_{\ell_2^n(W^{\hat{s}}_{\hat{p}}(\Omega))}} \geq \hat{C} n^{-1/2} \delta^{\hat{s} - d/\hat{p}}$$

for constant $\hat{C} > 0$ that is independent of δ and n. Using the packing number bound (8) we thus find a constant $\bar{C} > 0$ such that

$$\|W^{\hat{s}}_{\hat{p}}(\Omega) \hookrightarrow C^0(\Omega)\|_{\mathrm{type}_2} \geq \bar{C} \delta^{\hat{s} - d(1/\hat{p} - 1/2)}$$

holds for all $\delta \in (0, 1/2]$. Now, Lemma 5 implies $\hat{s} \geq (d/\hat{p} - d/2)_+$. Using a cotype 2 argument, we analogously obtain the requirement $\hat{s} \geq d/2$, where we use $\|(f_1, \dots, f_n)\|_{\ell_2^n(C^0(\Omega))} = n^{1/2}$.

Since ε is arbitrarily small, we conclude $s \geq (d/p - d/2)_+$ and $s \geq d/2$ from Eq. (47). It remains to show that $s \geq (d/p_1 - d/2)_+ + d/2$ holds. If $(d/p - d/2)_+ = 0$, this follows directly from our previous considerations. If $(d/p - d/2)_+ > 0$, then we have

$$(d/p - d/2)_+ + d/2 = d/p,$$

and the claim holds by assumption. □

Acknowledgements This work is funded by the Deutsche Forschungsgemeinschaft (DFG) in the project STE 1074/5-1, within the DFG priority programme SPP 2298 "Theoretical Foundations of Deep Learning". We would like to thank Jens Wirth and David Holzmüller for valuable discussions and insights that contributed to this work.

References

[1] R. A. Adams and J. J. F. Fournier. *Sobolev Spaces*. Academic Press, New York, 2nd edition, 2003.

[2] J. Agler and J. E. McCarthy. *Pick Interpolation and Hilbert Function Spaces*. American Mathematical Society, Providence, RI, 2002.

[3] N. Arcozzi, R. Rochberg, E.T. Sawyer, and B.D. Wick. *The Dirichlet Space and Related Function Spaces*. American Mathematical Society, Providence, RI, 2019.

[4] N. Aronszajn. Theory of reproducing kernels. *Trans. Amer. Math. Soc.*, 68:337–404, 1950.

[5] F. Bach. Breaking the curse of dimensionality with convex neural networks. *J. Mach. Learn. Res.*, 18:1–53, 2017.

[6] F. Bartolucci, E. De Vito, L. Rosasco, and S. Vigogna. Understanding neural networks with reproducing kernel Banach spaces. *Appl. Comput. Harmon. Anal.*, 62:194–236, 2023.

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

[8] N. Da Costa, C. Mostajeran, J. P. Ortega, and S. Said. Geometric learning with positively decomposable kernels, 2023.

[9] F. Cucker and D. X. Zhou. *Learning Theory: An Approximation Theory Viewpoint*. Cambridge University Press, Cambridge, 2007.

[10] J. Dick, P. Kritzer, and F. Pillichshammer. *Lattice Rules–Numerical Integration, Approximation, and Discrepancy*. Springer, Cham, 2022.

[11] J. Diestel, H. Jarchow, and A. Tonge. *Absolutely Summing Operators*. Cambridge University Press, Cambridge, 1995.

[12] R. M. Dudley. *Real Analysis and Probability*. Cambridge University Press, Cambridge, 2002.

[13] M. Eberts and I. Steinwart. Optimal regression rates for SVMs using Gaussian kernels. *Electron. J. Stat.*, 7:1–42, 2013.

[14] O. Forster. *Analysis 3*. Springer Spektrum, Wiesbaden, 8th revised edition, 2017.

[15] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola. A kernel two-sample test. *J. Mach. Learn. Res.*, 13:723–773, 2012.

[16] A. Gretton, K.M. Borgwardt, M. Rasch, B. Schölkopf, and A.J. Smola. A kernel method for the two-sample-problem. In P. B. Schölkopf, J. C. Platt, and T. Hoffman, editors, *Advances in Neural Information Processing Systems 19*, pages 513–520. MIT Press, 2007.

[17] A. F. Gualtierotti. *Detection of random signals in dependent Gaussian noise*. Springer, Cham, 2015.

[18] P. E. T. Jorgensen. *Harmonic Analysis*. American Mathematical Society, Providence, RI, 2018.

[19] D. H. Leung. Factoring operators through Hilbert space. *Israel J. Math.*, 71:225–227, 1990.

[20] F. Liu, L. Shi, X. Huang, J. Yang, and J. A. K. Suykens. Generalization properties of hyper-RKHS and its applications. *J. Mach. Learn. Res.*, 22:140, 2021.

[21] R. E. Megginson. *An Introduction to Banach Space Theory*. Springer, New York, 1998.

[22] C. A. Micchelli, Y. Xu, and H. Zhang. Universal kernels. *J. Mach. Learn. Res.*, 7:2651–2667, 2006.

[23] A. Müller. Integral probability metrics and their generating classes of functions. *Adv. in Appl. Probab.*, 29:429–443, 1997.

[24] C. S. Ong, X. Mary, S. Canu, and A. J. Smola. Learning with non-positive kernels. July 2004.

[25] C. S. Ong, A. J. Smola, and R. C. Williamson. Learning the kernel with hyperkernels. *J. Mach. Learn. Res.*, 6:1043–1071, 2005.

[26] S. Pereverzyev. *An Introduction to Artificial Intelligence based on Reproducing Kernel Hilbert Spaces*. Birkhäuser/Springer, Cham, 2022.

[27] G. Pisier. Un théorème sur les opérateurs linéaires entre espaces de Banach qui se factorisent par un espace de Hilbert. *Ann. Sci. École Norm. Sup.*, 13:23–43, 1980.

[28] G. Pisier. *Factorization of linear operators and geometry of Banach spaces*. American Mathematical Society, Providence, RI, 1986.

[29] T. Runst and W. Sickel. *Sobolev Spaces of Fractional Order, Nemytski Operators, and Nonlinear Partial Differential Equations*. de Gruyter, Berlin, 1996.

[30] S. Saitoh. *Theory of Reproducing Kernels and Applications*. Longman Scientific & Technical, Harlow, 1988.

[31] B. Schölkopf and A. J. Smola. *Learning with Kernels*. MIT Press, Cambridge, MA, 2002.

[32] C.-J. Simon-Gabriel and B. Schölkopf. Kernel distribution embeddings: Universal kernels, characteristic kernels and kernel metrics on distributions. *J. Mach. Learn. Res.*, (19):1–29, 2018.

[33] I. H. Sloan and H. Woźniakowski. When are quasi-Monte Carlo algorithms efficient for high-dimensional integrals? *J. Complexity*, 14:1–33, 1998.

[34] S. Smale and D.-X. Zhou. Estimating the approximation error in learning theory. *Anal. Appl.*, 1:17–41, 2003.

[35] B. K. Sriperumbudur, K. Fukumizu, and G. R. G. Lanckriet. Universality, characteristic kernels and RKHS embedding of measures. *J. Mach. Learn. Res.*, 12:2389–2410, 2011.

[36] B. K. Sriperumbudur, A. Gretton, K. Fukumizu, B. Schölkopf, and G. R. G. Lanckriet. Hilbert space embeddings and metrics on probability measures. *J. Mach. Learn. Res.*, 11:1517–1561, 2010.

[37] I. Steinwart. On the influence of the kernel on the consistency of support vector machines. *J. Mach. Learn. Res.*, 2:67–93, 2001. `http://www.jmlr.org/papers/volume2/steinwart01a/steinwart01a.pdf`.

[38] I. Steinwart. Reproducing kernel Hilbert spaces cannot contain all continuous functions on a compact metric space. *Arch. Math.*, 122:553–557, 2024.

[39] I. Steinwart and A. Christmann. *Support Vector Machines*. Springer, New York, 2008.

[40] I. Steinwart and J.F. Ziegel. Strictly proper kernel scores and characteristic kernels on compact spaces. *Appl. Comput. Harmon. Anal.*, 51:510–542, 2021. `https://arxiv.org/abs/1712.05279`.

[41] H. Triebel. *Theory of Function Spaces*. Akademische Verlagsgesellschaft Geest & Portig, Leipzig, 1983.

[42] H. Wendland. *Scattered Data Approximation*. Cambridge University Press, Cambridge, 2005.