J Theor Probab (2007) 20: 177–199
DOI 10.1007/s10959-007-0058-1

Bracketing Metric Entropy Rates and Empirical Central Limit Theorems for Function Classes of Besov- and Sobolev-Type

Richard Nickl · Benedikt M. Pötscher

Received: 6 October 2004 / Revised: 16 June 2006 /
Published online: 18 April 2007

Abstract We derive $\mathcal{L}^r(\mu)$-bracketing metric and sup-norm metric entropy rates of bounded subsets of general function spaces defined over $\mathbb{R}^d$ or, more generally, over Borel subsets thereof, by adapting results of Haroske and Triebel (Math. Nachr. **167**, 131–156, 1994; **278**, 108–132, 2005). The function spaces covered are of (weighted) Besov, Sobolev, Hölder, and Triebel type. Applications to the theory of empirical processes are discussed. In particular, we show that (norm-)bounded subsets of the above mentioned spaces are Donsker classes uniformly in various sets of probability measures.

Keywords Metric entropy with bracketing · Uniform metric entropy · Sobolev, Besov, Hölder, Triebel, and Bessel potential spaces · Uniform Donsker class · Glivenko-Cantelli

1 Introduction

In the theory of empirical processes and its applications (see, e.g., [11, 36]), the size of the function class indexing the empirical process plays a central role. The size, more precisely, the degree of compactness of the function class in some relevant topology, is measured by concepts like metric entropy (with or without bracketing). For example, many limit theorems for empirical processes (e.g., Glivenko-Cantelli and Donsker type results) are based on entropy conditions. The bracketing metric

An earlier version of this paper was circulated under the title "Bracketing Metric Entropy Rates for Function Classes of Besov- and Sobolev-Type Defined on Borel Sets of $\mathbb{R}^d$".

R. Nickl · B.M. Pötscher (✉)
Department of Statistics, University of Vienna, Universitätsstrasse 5/3, 1010 Vienna, Austria
e-mail: benedikt.poetscher@univie.ac.at

R. Nickl
e-mail: nickl@math.uconn.edu

entropy concept has proved to be particularly useful in this context; see Dudley [9], Alexander [2], Ossiander [25], Andersen, Giné, Ossiander, and Zinn [3]. Another prototypical application of bracketing metric entropy is in the study of convergence rates and lower risk bounds of statistical estimators (e.g., [5, 32, 33, 37]).

Classical results on (sup-norm) metric entropy bounds for 'smooth' function classes on bounded subsets of Euclidean space can be found in Kolmogorov and Tihomirov [21] and Birman and Solomyak [7]; see also Birgé and Massart [6]. [For a general exposition we refer to Chap. 15 in Lorentz, v.Golitschek, Makovoz [22] and Chap. 8 in Dudley [11].] Bracketing metric entropy bounds for such function classes can then immediately be obtained from these sup-norm metric entropy bounds. The only result regarding bracketing metric entropy bounds for classes of 'smooth' functions with *unbounded* support we are aware of is given in van der Vaart [30], which covers only a specific class of functions obtained from suitably pasting together functions that are of Lipschitz-type on bounded convex subsets; see Remark 5 for further discussion. Also note that bracketing metric entropy bounds for classes of functions of bounded variation that are defined on the real line are available; see van de Geer [31] and Remark 4 below.

In the present paper we provide bounds for the bracketing metric entropy of a large variety of 'smooth' function classes such as subsets of (weighted) Besov, Sobolev, Hölder, and Triebel spaces defined on $\mathbb{R}^d$ or on arbitrary Borel subsets thereof. We exploit results from the Fourier-analysis of such spaces [14, 19, 20]. In Sect. 3 we obtain sup-norm metric entropy, $\mathcal{L}^r(\mu)$-bracketing metric entropy, and uniform metric entropy bounds for (norm-)bounded subsets of the aforementioned spaces. Here the measure μ is not necessarily a finite measure. The entropy rates obtained are always of order $\varepsilon^{-\alpha}$ for some positive α, where α depends on the 'degree of smoothness' as well as on the behavior at infinity of the function class (and is connected to the measure μ via a suitable integrability condition in the bracketing case). These results also allow for classes of unbounded functions.

In Sect. 4 we provide sufficient conditions for any (norm-)bounded subset in a (weighted) Besov, Sobolev, Hölder, or Triebel space to be a Glivenko-Cantelli or Donsker class (uniformly in various sets of probability measures). These results are based on the bracketing bounds obtained in Sect. 3 on the one hand, and on a combination of embedding theorems for Besov spaces with results in Marcus [23] and Dudley [10] on the other hand. In Nickl [24] it is shown that these sufficient conditions are essentially sharp (at least in the unweighted case).

The focus of the paper is on real (weighted) Besov spaces. As discussed in Sect. 3.3, the results for Besov spaces immediately imply corresponding results for (weighted) Sobolev, Hölder, and Triebel spaces by using well-known embedding theorems.

2 Besov Spaces: Definition and Basic Properties

Let Ω be a (non-empty) Borel set in $\mathbb{R}^d$, $d \in \mathbb{N}$, and denote by $\mathcal{L}^0(\Omega)$ the set of real-valued $\mathcal{B}_\Omega$-measurable functions on Ω, where $\mathcal{B}_\Omega$ represents the σ-field of Borel sets of Ω. For $h \in \mathcal{L}^0(\Omega)$ and μ a (nonnegative) Borel measure on Ω, we set $\|h\|_{r,\mu} := (\int_\Omega |h|^r d\mu)^{1/r}$ for $1 \le r \le \infty$ where $\|h\|_{\infty,\mu}$ denotes the μ-essential supremum of $|h|$. [We always use the term Borel measure to mean a *nonnegative* Borel measure, not necessarily finite or σ-finite.] As usual, we denote by $\mathcal{L}^r(\Omega, \mu)$, or sometimes only $\mathcal{L}^r(\mu)$, the vector space of all $h \in \mathcal{L}^0(\Omega)$ that satisfy $\|h\|_{r,\mu} < \infty$. Furthermore, $L^r(\Omega, \mu)$ (or $L^r(\mu)$) denotes the corresponding Banach space of equivalence classes $[h]_\mu$, $h \in \mathcal{L}^r(\Omega, \mu)$, modulo equality μ-almost everywhere. The symbol λ will be used to denote Lebesgue-measure on the Borel sets of $\mathbb{R}^d$, and $\lambda|\Omega$ will denote the restriction of λ to Ω.

Furthermore, let $\mathsf{C}(\mathbb{R}^d)$ be the vector space of bounded continuous real-valued functions on $\mathbb{R}^d$ normed by the sup-norm $\|\cdot\|_\infty$. Also, let $\mathsf{UC}(\mathbb{R}^d)$ be the closed subspace of $\mathsf{C}(\mathbb{R}^d)$ that consists of all uniformly continuous functions, again equipped with the sup-norm. Attaching the subscript 0 to any of these two spaces denotes the respective closed subspace of functions satisfying $\lim_{\|x\|\to\infty} f(x) = 0$ where $\|\cdot\|$ always denotes the Euclidean norm. [Clearly, $\mathsf{UC}_0(\mathbb{R}^d) = \mathsf{C}_0(\mathbb{R}^d)$.]

We follow Edmunds and Triebel ([14] 2.2.1) in defining Besov spaces: Let φ_0 be a complex-valued C^∞-function on $\mathbb{R}^d$ with $\varphi_0(x) = 1$ if $\|x\| \le 1$ and $\varphi_0(x) = 0$ if $\|x\| \ge 3/2$. Define $\varphi_1(x) = \varphi_0(x/2) - \varphi_0(x)$ and $\varphi_k(x) = \varphi_1(2^{-k+1}x)$ for $k \in \mathbb{N}$. Then the functions φ_k form a dyadic resolution of unity. Let $\mathcal{S}'(\mathbb{R}^d)$ denote the space of complex tempered distributions on $\mathbb{R}^d$ and let F denote the Fourier transform acting on this space (with scaling constant $(2\pi)^{-d/2}$). Since any $f \in \mathcal{L}^p(\mathbb{R}^d, \lambda)$ gives rise to an element of $\mathcal{S}'(\mathbb{R}^d)$, the quantity $F^{-1}(\varphi_k F f)$ is well-defined (for any k) as an element of $\mathcal{S}'(\mathbb{R}^d)$. [In fact, more is true: $F^{-1}(\varphi_k F T)$ is an entire analytic function on $\mathbb{R}^d$ for any $T \in \mathcal{S}'(\mathbb{R}^d)$ and any k by the Paley-Wiener-Schwartz theorem.]

Definition 1 (Besov spaces) Let $0 \le s < \infty$, $1 \le p \le \infty$, and $1 \le q \le \infty$, with $q = 1$ in case $s = 0$. For $f \in \mathcal{L}^p(\mathbb{R}^d, \lambda)$ define

$$\|f\|_{s,p,q,\lambda} := \left(\sum_{k=0}^{\infty} 2^{ksq} \left\| F^{-1}(\varphi_k F f) \right\|_{p,\lambda}^q \right)^{1/q}$$

with the modification in case $q = \infty$

$$\|f\|_{s,p,\infty,\lambda} := \sup_{0 \le k < \infty} 2^{ks} \left\| F^{-1}(\varphi_k F f) \right\|_{p,\lambda}.$$

Define further

$$\mathcal{B}^s_{pq}(\mathbb{R}^d) := \left\{ f \in \mathcal{L}^p(\mathbb{R}^d, \lambda) : \|f\|_{s,p,q,\lambda} < \infty \right\}.$$

The Besov space $(\mathcal{B}^s_{pq}(\mathbb{R}^d), \|\cdot\|_{s,p,q,\lambda})$ is a semi-normed vector space (over $\mathbb{R}$). Note that with each element $f \in \mathcal{B}^s_{pq}(\mathbb{R}^d)$ any $f^* \in [f]_\lambda$ also belongs to $\mathcal{B}^s_{pq}(\mathbb{R}^d)$. By taking the quotient w.r.t. the set $\{f : \|f\|_{s,p,q,\lambda} = 0\}$ one obtains the Banach space $(B^s_{pq}(\mathbb{R}^d), \|\cdot\|_{s,p,q,\lambda})$. The Besov spaces are independent of the choice of φ_0, and, in particular, different φ_0 result in equivalent norms, cf. Edmunds and Triebel ([14], 2.2.1).

Remark 1 (i) The norm symbol in Definition 1 is in fact well-defined for arbitrary complex tempered distributions $T \in \mathcal{S}'(\mathbb{R}^d)$, and $\{T \in \mathcal{S}'(\mathbb{R}^d) : \|T\|_{s,p,q,\lambda} < \infty\}$ defines the complex Besov space (for $-\infty < s < \infty$, $1 \le p \le \infty$, $1 \le q \le \infty$); see

Edmunds and Triebel [14, 2.2.1]. Restricting attention to real-valued tempered distributions T (i.e., $T=\bar{T}$, where $\bar{T}$ denotes the conjugate complex of T) leads to the real Besov space $\{T\in\mathcal{S}'(\mathbb{R}^d):T=\bar{T},\|T\|_{s,p,q,\lambda}<\infty\}$. Furthermore, in the case $s>0$ or $s=0$ but $q=1$, it follows from Triebel [28, 2.3.2/7,2.5.7/1,2], that the so-defined real Besov space coincides with the space $B^s_{pq}(\mathbb{R}^d)$ defined above; hence, the restriction to $\mathcal{L}^p(\mathbb{R}^d,\lambda)$ in the definition of $\mathcal{B}^s_{pq}(\mathbb{R}^d)$ above is natural.

(ii) We note that $\|T\|_{s,p,q,\lambda}<\infty$ if and only if $\|\bar{T}\|_{s,p,q,\lambda}<\infty$ for any $T\in\mathcal{S}'(\mathbb{R}^d)$. This shows that T is an element of the complex Besov space if and only if the real part as well as the imaginary part of T belong to the corresponding real Besov space. In fact,

$$\|T\|_{s,p,q,\lambda}\le\|\operatorname{Re}T\|_{s,p,q,\lambda}+\|\operatorname{Im}T\|_{s,p,q,\lambda}\le c\|T\|_{s,p,q,\lambda}$$

holds for some $1\le c<\infty$ and for every $T\in\mathcal{S}'(\mathbb{R}^d)$. [To see this, recall first that different choices of φ_0 result in equivalent norms. Then, choosing φ_0 real-valued and symmetric gives $F^{-1}(\varphi_k F\bar{T})=\overline{F^{-1}(\varphi_k FT)}$, and hence $\|T\|_{s,p,q,\lambda}=\|\bar{T}\|_{s,p,q,\lambda}$.] In particular, the real Besov space is a closed subset of the corresponding complex Besov space. As a consequence of these facts, one can easily carry over results for complex Besov spaces to real ones and vice versa. [We shall use this tacitly when using results for complex Besov spaces from the literature.]

Remark 2 There are many equivalent definitions of Besov spaces, see, e.g., Triebel [28, 2.5.3, 2.5.7, 2.5.12]. We mention only one of them: Let $\alpha=(\alpha_1,\dots,\alpha_d)$ be a multi-index of nonnegative integers α_i, set $|\alpha|=\sum_{i=1}^d\alpha_i$, and let $D^\alpha=\frac{\partial^{|\alpha|}}{(\partial x_1)^{\alpha_1}\dots(\partial x_d)^{\alpha_d}}$ denote the partial differential operator of order $|\alpha|$ in the sense of distributions. Furthermore, for a function $f:\mathbb{R}^d\to\mathbb{R}$, the difference operator Δ_z is defined by $\Delta_z f(\cdot)=f(\cdot+z)-f(\cdot)$ and $\Delta_z^2 f(\cdot)=\Delta_z(\Delta_z f(\cdot))$ for $z\in\mathbb{R}^d$. Let $0<s<\infty$ and decompose s as $s=[s]^-+\{s\}^+$ with $[s]^-$ integer and $0<\{s\}^+\le 1$. Let further $1\le p\le\infty$ and $1\le q\le\infty$. For $f\in\mathcal{L}^p(\mathbb{R}^d)$ with $\|D^\alpha f\|_{p,\lambda}<\infty$ for $0\le|\alpha|\le[s]^-$, define

$$\|f\|^*_{s,p,q,\lambda}:=\sum_{0\le|\alpha|\le[s]^-}\|D^\alpha f\|_{p,\lambda}+\sum_{|\alpha|=[s]^-}\left(\int_{\mathbb{R}^d}|z|^{-\{s\}^+q-d}\left\|\Delta_z^2 D^\alpha f\right\|^q_{p,\lambda}dz\right)^{1/q}$$

with the modification in case $q=\infty$

$$\|f\|^*_{s,p,\infty,\lambda}:=\sum_{0\le|\alpha|\le[s]^-}\|D^\alpha f\|_{p,\lambda}+\sum_{|\alpha|=[s]^-}\sup_{0\ne z\in\mathbb{R}^d}|z|^{-\{s\}^+}\left\|\Delta_z^2 D^\alpha f\right\|_{p,\lambda}.$$

Then $\{f\in\mathcal{L}^p(\mathbb{R}^d):\|f\|^*_{s,p,q,\lambda}<\infty\}$ coincides with $\mathcal{B}^s_{pq}(\mathbb{R}^d)$, and the (semi)norms $\|\cdot\|_{s,p,q,\lambda}$ and $\|\cdot\|^*_{s,p,q,\lambda}$ are equivalent on $\mathcal{B}^s_{pq}(\mathbb{R}^d)$.

In case $s>d/p$, it is well-known that each equivalence class $[f]_\lambda$, $f\in\mathcal{B}^s_{pq}(\mathbb{R}^d)$, contains a (unique) continuous representative. [This representative possesses in fact classical partial derivatives of order (at least) $s-d/p$, hence $s-d/p$ is called the *differential dimension* of the Besov space.] Consequently, in case $s > d/p$, we can define the (closely related) Banach space

$$\mathsf{B}^s_{pq}(\mathbb{R}^d) = \mathcal{B}^s_{pq}(\mathbb{R}^d) \cap \mathsf{C}(\mathbb{R}^d),$$

again equipped with the norm $\|\cdot\|_{s,p,q,\lambda}$; that is $\mathsf{B}^s_{pq}(\mathbb{R}^d)$ is obtained by collecting the continuous representatives. We note that any $f \in \mathsf{B}^s_{pq}(\mathbb{R}^d)$ is also bounded. Furthermore, if additionally $p < \infty$, any $f \in \mathsf{B}^s_{pq}(\mathbb{R}^d)$ satisfies $\lim_{\|x\|\to\infty} f(x) = 0$. For convenience of the reader these well-known properties are summarized in Proposition 3 in the Appendix. In this proposition it is also shown that the aforementioned properties continue to hold in case $s = d/p$ and $q = 1$.

To control tail behavior, we introduce weighted Besov spaces, again following Edmunds and Triebel [14, 4.2]. Define the polynomial weighting function $\langle x\rangle^\beta = (1 + \|x\|^2)^{\beta/2}$ parameterized by $\beta \in \mathbb{R}$, where x is an element of $\mathbb{R}^d$.

Definition 2 (Weighted Besov spaces) Let $0 \le s < \infty$, $1 \le p \le \infty$, and $1 \le q \le \infty$, with $q = 1$ in case $s = 0$. For $\beta \in \mathbb{R}$ define

$$\mathcal{B}^s_{pq}(\mathbb{R}^d, \langle x\rangle^\beta) := \left\{ f : \|f \cdot \langle x\rangle^\beta\|_{s,p,q,\lambda} < \infty \right\}.$$

Again, the pair

$$\left(\mathcal{B}^s_{pq}(\mathbb{R}^d, \langle x\rangle^\beta), \|(\cdot)\langle x\rangle^\beta\|_{s,p,q,\lambda}\right)$$

is a semi-normed vector space and $B^s_{pq}(\mathbb{R}^d, \langle x\rangle^\beta)$ denotes the corresponding quotient space. In the case $s > d/p$ or $s = d/p$ and $q = 1$ we also define

$$\mathsf{B}^s_{pq}(\mathbb{R}^d, \langle x\rangle^\beta) = \mathcal{B}^s_{pq}(\mathbb{R}^d, \langle x\rangle^\beta) \cap \{f : f \cdot \langle x\rangle^\beta \in \mathsf{C}(\mathbb{R}^d)\}.$$

Since $\langle x\rangle^\beta = (1 + \|x\|^2)^{\beta/2}$ is continuous and positive, $\mathsf{B}^s_{pq}(\mathbb{R}^d, \langle x\rangle^\beta)$ is a Banach space of continuous functions. Note that elements of $\mathsf{B}^s_{pq}(\mathbb{R}^d, \langle x\rangle^\beta)$ are always bounded functions if $\beta \ge 0$, but not necessarily if $\beta < 0$.

3 Bracketing Metric Entropy Bounds

Definition 3 For a (non-empty) subset $\mathcal{J}$ of a normed space $(X, \|\cdot\|_X)$, let $N(\varepsilon, \mathcal{J}, \|\cdot\|_X)$ denote the *minimal covering number*, i.e., the minimal number of closed balls of radius ε, $0 < \varepsilon < \infty$, (w.r.t. $\|\cdot\|_X$) needed to cover $\mathcal{J}$. In accordance, let $H(\varepsilon, \mathcal{J}, \|\cdot\|_X) = \log N(\varepsilon, \mathcal{J}, \|\cdot\|_X)$ be the *metric entropy* of the set $\mathcal{J}$, where log denotes the natural logarithm.

Definition 4 Let Ω be a (non-empty) Borel subset of $\mathbb{R}^d$. Given two Borel-measurable functions $l, u : \Omega \to \mathbb{R}$, the bracket $[l, u]$ is the set of all functions $f \in \mathcal{L}^0(\Omega)$ with $l \le f \le u$. Given a Borel-measure μ on Ω and $1 \le r \le \infty$, the $\mathcal{L}^r(\mu)$-size of the bracket $[l, u]$ is defined as $\|u - l\|_{r,\mu}$. The $\mathcal{L}^r(\mu)$-bracketing number $N_{[]}(\varepsilon, \mathcal{F}, \|\cdot\|_{r,\mu})$ of a (non-empty) set $\mathcal{F} \subseteq \mathcal{L}^0(\Omega)$ is the minimal number of brackets of $\mathcal{L}^r(\mu)$-size less than or equal to ε, $0 < \varepsilon < \infty$, necessary to cover $\mathcal{F}$. The logarithm of the bracketing number is called the $\mathcal{L}^r(\mu)$-bracketing metric entropy $H_{[]}(\varepsilon, \mathcal{F}, \|\cdot\|_{r,\mu})$.

In the above definitions it is implicitly understood that $N(\varepsilon, \mathcal{J}, \|\cdot\|_X)$ and $N_{[]}(\varepsilon, \mathcal{F}, \|\cdot\|_{r,\mu})$ are finite. For two real-valued functions $a(\cdot)$ and $b(\cdot)$, we write $a(\varepsilon) \lesssim b(\varepsilon)$ if there exists a positive (finite) constant c not depending on ε such that $a(\varepsilon) \leq cb(\varepsilon)$ holds for every $\varepsilon > 0$. If $a(\varepsilon) \lesssim b(\varepsilon)$ and $b(\varepsilon) \lesssim a(\varepsilon)$ both hold we write $a(\varepsilon) \sim b(\varepsilon)$. [In abuse of notation, we shall also use this notation for sequences a_k and b_k, $k \in \mathbb{N}$.] Furthermore, for two (semi)norms $\|\cdot\|_{X,1}$ and $\|\cdot\|_{X,2}$ on a vector space X, we write $\|\cdot\|_{X,1} \lesssim \|\cdot\|_{X,2}$ if $\|\cdot\|_{X,1} \leq c\|\cdot\|_{X,2}$ for a (finite) positive constant c, and we write $\|\cdot\|_{X,1} \sim \|\cdot\|_{X,2}$ if the (semi)norms are equivalent.

3.1 Bracketing Metric Entropy in Weighted Besov Spaces

We now give one of the main results of the paper. Note that the second part of this theorem, in case $\beta < 0$, also allows for function classes $\mathcal{F}$ that may contain unbounded functions.

Theorem 1 *Let* $1 \leq p \leq \infty$, $1 \leq q \leq \infty$, $\beta \in \mathbb{R}$, *and* $s - d/p > 0$. *Suppose* $\mathcal{F}$ *is a (non-empty) bounded subset of* $\mathsf{B}^s_{pq}(\mathbb{R}^d, \langle x\rangle^\beta)$. *Then the following holds*:

1. *Suppose* $\beta > 0$. *Then* $\mathcal{F}$ *is a relatively compact subset of* $\mathsf{UC}_0(\mathbb{R}^d)$. *Furthermore, for* $\beta > s - d/p$ *we have*
$$H(\varepsilon, \mathcal{F}, \|\cdot\|_\infty) \lesssim \varepsilon^{-d/s},$$
and for $\beta < s - d/p$ *we have*
$$H(\varepsilon, \mathcal{F}, \|\cdot\|_\infty) \lesssim \varepsilon^{-(\beta/d+1/p)^{-1}}.$$
2. *Suppose that for some Borel measure* μ *on* $\mathbb{R}^d$ *and some* $1 \leq r \leq \infty$ *the moment condition* $\|\langle x\rangle^{\gamma-\beta}\|_{r,\mu} < \infty$ *holds for some* $\gamma > 0$. *If* $\gamma > s - d/p$, *we have*
$$H_{[]}(\varepsilon, \mathcal{F}, \|\cdot\|_{r,\mu}) \lesssim \varepsilon^{-d/s};$$
if $\gamma < s - d/p$, *we have*
$$H_{[]}(\varepsilon, \mathcal{F}, \|\cdot\|_{r,\mu}) \lesssim \varepsilon^{-(\gamma/d+1/p)^{-1}}.$$

Proof Without loss of generality, we may assume that $\mathcal{F}$ is the closed unit ball in $\mathsf{B}^s_{pq}(\mathbb{R}^d, \langle x\rangle^\beta)$. To prove the first claim in Part 1 of the theorem, apply Proposition 2 with $\gamma = \beta > 0$ and Proposition 3, both in the Appendix, to obtain

$$\mathsf{B}^s_{pq}(\mathbb{R}^d, \langle x\rangle^\beta) \hookrightarrow \mathsf{B}^0_{\infty 1}(\mathbb{R}^d) \hookrightarrow \mathsf{UC}(\mathbb{R}^d),$$

where the first, and hence the composite embedding, are compact. [For the definition of the embedding symbol $\hookrightarrow$ see the Appendix.] Consequently, $\mathcal{F}$ is a precompact and hence a relatively compact subset of $\mathsf{UC}(\mathbb{R}^d)$; by the second part of Proposition 3 it even belongs to the closed subspace $\mathsf{UC}_0(\mathbb{R}^d)$. This proves the first claim in Part 1.

To prove the remaining assertions of the theorem we first establish a bound for the weighted sup-norm metric entropy of $\mathcal{F}$. Proposition 2 in the Appendix (and an obvious selection-of-continuous-representatives argument) gives the rate of the entropy number

$$e_k = e\big(k, \mathcal{F}, \|(\cdot)\langle x\rangle^{\beta-\gamma}\|_{0,\infty,1,\lambda}\big)$$

of $\mathcal{F}$ in the space $\mathsf{B}^0_{\infty 1}(\mathbb{R}^d, \langle x\rangle^{\beta-\gamma})$ for arbitrary $\beta \in \mathbb{R}$ and $\gamma > 0$. Observe that the sequence e_k is nonincreasing, converges to zero, and satisfies $0 < e_k < \infty$. Therefore, for every $k \in \mathbb{N}$ there exists a unique non-negative integer $l(k)$ such that $e_k = e_{k+l(k)}$ and $e_k > e_{k+l(k)+1}$ hold. From the definition of the covering numbers it follows for every $\eta > e_{k+l(k)+1}$ that

$$\log_2 N\big(\eta, \mathcal{F}, \|(\cdot)\langle x\rangle^{\beta-\gamma}\|_{0,\infty,1,\lambda}\big) \le k + l(k)$$

and thus

$$\log_2 N\big(e_k, \mathcal{F}, \|(\cdot)\langle x\rangle^{\beta-\gamma}\|_{0,\infty,1,\lambda}\big) \le k + l(k).$$

Fix $0 < \varepsilon \le e_1$. Then there exists a unique index $k = k(\varepsilon) \ge 2$ such that $e_k < \varepsilon \le e_{k-1}$. Consequently,

$$\log_2 N\big(\varepsilon, \mathcal{F}, \|(\cdot)\langle x\rangle^{\beta-\gamma}\|_{0,\infty,1,\lambda}\big) \le \log_2 N\big(e_k, \mathcal{F}, \|(\cdot)\langle x\rangle^{\beta-\gamma}\|_{0,\infty,1,\lambda}\big) \le k + l(k).$$

By Proposition 2 in the Appendix, there exists a positive and finite constant c such that $e_k = e_{k+l(k)} \le c(k + l(k))^{-\alpha}$ holds, where either $\alpha = s/d$ or $\alpha = \gamma/d + 1/p$. Consequently,

$$\log_2 N\big(\varepsilon, \mathcal{F}, \|(\cdot)\langle x\rangle^{\beta-\gamma}\|_{0,\infty,1,\lambda}\big) \le k + l(k) \le c^{1/\alpha} e_k^{-1/\alpha}.$$

Similarly, there exists a positive finite constant c' such that $c'k^{-\alpha} \le e_k$. Hence,

$$\begin{aligned}\log_2 N\big(\varepsilon, \mathcal{F}, \|(\cdot)\langle x\rangle^{\beta-\gamma}\|_{0,\infty,1,\lambda}\big) &\le (c/c')^{1/\alpha} k = (c/c')^{1/\alpha}[k - 1 + 1]\\ &\le (c/c')^{1/\alpha}\big[c^{1/\alpha} e_{k-1}^{-1/\alpha} + 1\big]\\ &\le (c/c')^{1/\alpha}\big[c^{1/\alpha}\varepsilon^{-1/\alpha} + 1\big],\end{aligned}$$

where in the final two steps we have used the relations $e_{k-1} \le c(k-1)^{-\alpha}$ and $e_{k-1} \ge \varepsilon$. This establishes

$$H\big(\varepsilon, \mathcal{F}, \|(\cdot)\langle x\rangle^{\beta-\gamma}\|_{0,\infty,1,\lambda}\big) \le C_1 \varepsilon^{-1/\alpha} \tag{1}$$

for $0 < \varepsilon \le e_1$ and a suitable real number C_1. In fact (1) holds for every $\varepsilon > 0$, since the metric entropy is zero for $\varepsilon > e_1$ as e_1 is the operator norm of the embedding.

By Part 1 of Proposition 3 in the Appendix we have, in particular, the embedding

$$\big(\mathsf{B}^0_{\infty 1}(\mathbb{R}^d), \|\cdot\|_{0,\infty,1,\lambda}\big) \hookrightarrow \big(\mathsf{C}(\mathbb{R}^d), \|\cdot\|_\infty\big). \tag{2}$$

Hence, the inequality

$$\|(\cdot)\langle x\rangle^{\beta-\gamma}\|_\infty \lesssim \|(\cdot)\langle x\rangle^{\beta-\gamma}\|_{0,\infty,1,\lambda}$$

holds on the space $\mathsf{B}^0_{\infty 1}(\mathbb{R}^d, \langle x\rangle^{\beta-\gamma})$. Lemma 2 in the Appendix now gives the bound

$$H\big(\varepsilon, \mathcal{F}, \|(\cdot)\langle x\rangle^{\beta-\gamma}\|_\infty\big) \le C_2\varepsilon^{-1/\alpha} \tag{3}$$

for every $\varepsilon > 0$ and a suitable real number C_2. [In the above, $\mathcal{F}$ is viewed as a subset of the linear space $\mathsf{C}(\mathbb{R}^d, \langle x\rangle^{\beta-\gamma}) = \{f : f \cdot \langle x\rangle^{\beta-\gamma} \in \mathsf{C}(\mathbb{R}^d)\}$ normed by $\|(\cdot)\langle x\rangle^{\beta-\gamma}\|_\infty$.]

The bound (3) immediately implies the remaining claims in the first part of the theorem upon noting that we may set $\gamma = \beta$ under the assumptions of Part 1.

For the proof of the second part, let B_i denote closed balls in $\mathsf{C}(\mathbb{R}^d, \langle x\rangle^{\beta-\gamma})$ of radius ε covering $\mathcal{F}$, where $i = 1, \dots, N(\varepsilon, \mathcal{F}, \|(\cdot)\langle x\rangle^{\beta-\gamma}\|_\infty)$. Observe that each such ball B_i (with center f_i) contains all (continuous) functions f for which

$$\sup_{x\in\mathbb{R}^d} |f(x) - f_i(x)|\langle x\rangle^{\beta-\gamma} \le \varepsilon.$$

These balls induce brackets

$$\big[f_i - \varepsilon\langle x\rangle^{\gamma-\beta}, f_i + \varepsilon\langle x\rangle^{\gamma-\beta}\big]$$

which obviously cover $\mathcal{F}$. The $\mathcal{L}^r(\mu)$-size of such a bracket is given by

$$\|2\varepsilon\langle x\rangle^{\gamma-\beta}\|_{r,\mu}$$

which is finite by assumption. Thus, using (3),

$$H_{[]}\big(\varepsilon\|2\langle x\rangle^{\gamma-\beta}\|_{r,\mu}, \mathcal{F}, \|\cdot\|_{r,\mu}\big) \le H\big(\varepsilon, \mathcal{F}, \|(\cdot)\langle x\rangle^{\beta-\gamma}\|_\infty\big) \lesssim \varepsilon^{-1/\alpha} \tag{4}$$

for every $\varepsilon > 0$ (provided $\mu(\mathbb{R}^d) > 0$). Inserting the definition of α and rescaling by $\|2\langle x\rangle^{\gamma-\beta}\|_{r,\mu}$ delivers the desired result. [If $\mu(\mathbb{R}^d) = 0$, the result is trivial since by Proposition 3 in the Appendix the single bracket $[-K\langle x\rangle^{-\beta}, K\langle x\rangle^{-\beta}]$ covers $\mathcal{F}$ and has size zero.] □

The proof has in fact established a bound for the weighted sup-norm metric entropy, cf. (3), for arbitrary $\beta \in \mathbb{R}$. This bound delivers the first part of the theorem as a special case upon setting $\gamma = \beta > 0$, and provides $\mathcal{L}^r(\mu)$-bracketing metric entropy bounds for general β. We furthermore note that the upper bound (3) is best possible in the sense that a lower bound of the same polynomial order can be established for the unit ball in $\mathsf{B}^s_{pq}(\mathbb{R}^d, \langle x\rangle^\beta)$ by a variation of the above proof.

Note that $\mathcal{F}$ possesses an envelope $K\langle x\rangle^{-\beta}$ (cf. Proposition 3 in the Appendix). The moment condition in the second part of the theorem constitutes a trade-off between the behavior of the envelope $K\langle x\rangle^{-\beta}$ for large $\|x\|$ and the tail-behavior of the measure μ. While the moment condition becomes more lenient the smaller γ is, the rate bound deteriorates as γ becomes smaller. If $\beta > 0$ and μ is a finite measure or $r = \infty$, the moment condition is always satisfied at least for $\gamma = \beta$. If μ has an exponential moment, the condition $\|\langle x\rangle^{\gamma-\beta}\|_{r,\mu} < \infty$ is of course satisfied for any γ and β (and $r < \infty$). If μ is Lebesgue-measure, the moment condition holds for any γ satisfying $0 < \gamma < \beta - d/r$.

A direct metric entropy bound for the case $\beta = s - d/p$ is possible, but involved. However, a simple (and only marginally worse) metric entropy bound for $\mathcal{F}$ in this case can be obtained by viewing $\mathcal{F}$ as a bounded subset of $\mathsf{B}^t_{pq}(\mathbb{R}^d, \langle x\rangle^\beta)$ with $t < s$, and by applying the above theorem with t in place of s. A similar remark applies to the case $\gamma = s - d/p$ in the second part of the theorem.

If $\mathsf{B}^s_{pq}(\mathbb{R}^d, \langle x\rangle^\beta)$ is reflexive (see 2.6.1/2 in Triebel [30]), if $\beta > 0$, and if $\mathcal{F}$ in Theorem 1 is not only bounded but also weakly closed, then $\mathcal{F}$ is not only relatively compact but even compact in $\mathsf{UC}_0(\mathbb{R}^d)$. This follows from reflexivity, Alaoglu's theorem, and the observation that the set of evaluation functionals $\{\delta_x : x \in \mathbb{R}^d\}$ is contained in the dual space of $\mathsf{B}^s_{pq}(\mathbb{R}^d, \langle x\rangle^\beta)$ for $s > d/p$ and $\beta > 0$.

Finally, as a corollary to Theorem 1, we obtain uniform (bracketing) metric entropy bounds.

Corollary 1 *Let* $1 \le p \le \infty$, $1 \le q \le \infty$, $\beta \in \mathbb{R}$, *and* $s - d/p > 0$. *Let* $\mathcal{F}$ *be a (non-empty) bounded subset of* $\mathsf{B}^s_{pq}(\mathbb{R}^d, \langle x\rangle^\beta)$. *Let furthermore* $\mathfrak{M}$ *be a (non-empty) family of Borel measures on* $\mathbb{R}^d$ *such that the condition* $\sup_{\mu\in\mathfrak{M}} \|\langle x\rangle^{\gamma-\beta}\|_{r,\mu} < \infty$ *holds for some* $\gamma > 0$ *and for some* $1 \le r \le \infty$. *Then*

$$\sup_{\mu\in\mathfrak{M}} H(\varepsilon, \mathcal{F}, \|\cdot\|_{r,\mu}) \le \sup_{\mu\in\mathfrak{M}} H_{[]}(2\varepsilon, \mathcal{F}, \|\cdot\|_{r,\mu}) \lesssim \begin{cases} \varepsilon^{-d/s} & \text{for } \gamma > s - d/p, \\ \varepsilon^{-(\gamma/d+1/p)^{-1}} & \text{for } \gamma < s - d/p. \end{cases}$$

Proof The first inequality is obvious. Inspection of (4) in the proof of the second part of Theorem 1 immediately gives the second inequality. □

In particular, if $\beta > 0$, we may set $\gamma = \beta$ in which case the uniform moment condition is satisfied, e.g., for $\mathfrak{M}$ the set of all probability measures.

3.2 Besov Spaces on Subsets of $\mathbb{R}^d$

Often classes of 'smooth' functions defined on Borel subsets Ω of $\mathbb{R}^d$ are of interest. If these function classes are defined through restricting elements of $\mathsf{B}^s_{pq}(\mathbb{R}^d, \langle x\rangle^\beta)$ to the set Ω, the results in Sect. 3.1 can immediately be used to deliver the following result for bracketing metric entropy. [A corresponding sup-norm metric entropy bound can similarly be obtained from Theorem 1.]

Corollary 2 *Let* Ω *be a (non-empty) Borel subset of* $\mathbb{R}^d$, *let* $1 \le p \le \infty$, $1 \le q \le \infty$, $\beta \in \mathbb{R}$, *and* $s - d/p > 0$. *Suppose* $\mathcal{F}$ *is a (non-empty) bounded subset of* $\mathsf{B}^s_{pq}(\mathbb{R}^d, \langle x\rangle^\beta)$. *Let* $\mathcal{F}|\Omega$ *be the set of restrictions* $f|\Omega$ *of elements* $f \in \mathcal{F}$ *to the set* Ω. *Let* $\mathfrak{M}$ *be a (non-empty) family of Borel measures on* Ω.

1. *Suppose that* $\sup_{\mu\in\mathfrak{M}} \|\langle x\rangle^{\gamma-\beta}\|_{r,\mu} < \infty$ *holds for some* $\gamma > 0$ *and for some* $1 \le r \le \infty$. *Then*

$$\sup_{\mu\in\mathfrak{M}} H_{[]}(\varepsilon, \mathcal{F}|\Omega, \|\cdot\|_{r,\mu}) \lesssim \begin{cases} \varepsilon^{-d/s} & \text{for } \gamma > s - d/p, \\ \varepsilon^{-(\gamma/d+1/p)^{-1}} & \text{for } \gamma < s - d/p. \end{cases}$$

2. *Suppose* Ω *is a bounded set. If* $1 \le r < \infty$ *and* $\sup_{\mu\in\mathfrak{M}} \mu(\Omega) < \infty$, *or if* $r = \infty$ *then*

$$\sup_{\mu\in\mathfrak{M}} H_{[]}(\varepsilon, \mathcal{F}|\Omega, \|\cdot\|_{r,\mu}) \lesssim \varepsilon^{-d/s}.$$

Proof To prove the first part, identify $\mathcal{F}|\Omega$ with $\mathcal{F}\mathbf{1}_\Omega = \{f\mathbf{1}_\Omega : f \in \mathcal{F}\}$ and view the measures $\mu \in \mathfrak{M}$ as Borel measures on $\mathbb{R}^d$ with $\mu(\mathbb{R}^d \backslash \Omega) = 0$. Clearly,

$$H_{[]}(\varepsilon, \mathcal{F}|\Omega, \|\cdot\|_{r,\mu}) = H_{[]}(\varepsilon, \mathcal{F}\mathbf{1}_\Omega, \|\cdot\|_{r,\mu}) \le H_{[]}(\varepsilon, \mathcal{F}, \|\cdot\|_{r,\mu})$$

holds. The first part now follows from Corollary 1. For the proof of the second part of the corollary choose $\beta' > s - d/p$. Since Ω is a bounded set we have that $\mathcal{F}|\Omega \subseteq \mathcal{F}'|\Omega$, for a suitable bounded subset $\mathcal{F}'$ of $\mathsf{B}^s_{pq}(\mathbb{R}^d, \langle x\rangle^{\beta'})$. [E.g., choose $\mathcal{F}' = \mathcal{F}h$ where $h : \mathbb{R}^d \to \mathbb{R}$ is an infinitely differentiable function with compact support that satisfies $h(x) = 1$ for $x \in \Omega$, and use the fact that $\mathsf{B}^s_{pq}(\mathbb{R}^d)$ is a multiplication algebra, cf. [28, 2.8.3].] Set $\gamma = \beta'$ and apply the first part of the corollary to the set $\mathcal{F}'|\Omega$. Note that the moment condition

$$\sup_{\mu\in\mathfrak{M}} \|\langle x\rangle^{\gamma-\beta'}\|_{r,\mu} = \sup_{\mu\in\mathfrak{M}} \|1\|_{r,\mu} < \infty$$

is now always satisfied. □

We note that the set $\mathcal{F}|\Omega$ can be viewed as a bounded subset of $\mathsf{B}^s_{pq}(\mathbb{R}^d, \langle x\rangle^\beta)|\Omega$, the "factor space" obtained by restriction to Ω, which is normed by

$$\|f\cdot\langle x\rangle^\beta\|_{s,p,q,|\Omega} := \inf\big\{\|g\cdot\langle x\rangle^\beta\|_{s,p,q,\lambda} : g \in \mathsf{B}^s_{pq}(\mathbb{R}^d, \langle x\rangle^\beta),\ g|\Omega = f\big\}.$$

However, for Ω a domain, i.e., an open subset of $\mathbb{R}^d$, the "intrinsic" definition of a (weighted) Besov space on Ω is not given via restricting the elements of $\mathcal{B}^s_{pq}(\mathbb{R}^d, \langle x\rangle^\beta)$ to the domain Ω. Rather, a norm similar in spirit to the one mentioned in Remark 2, but only involving the values of $f(x)$ for $x \in \Omega$, is typically used (cf. 3.4.2/6 in [28] and 1.10.3-4 in [29]) resulting in a (weighted) Besov space $\mathcal{B}^s_{pq}(\Omega, \langle x\rangle^\beta)$ on Ω. Passing from $\mathcal{B}^s_{pq}(\Omega, \langle x\rangle^\beta)$ to $\mathsf{B}^s_{pq}(\Omega, \langle x\rangle^\beta)$ is then analogously possible if $s > d/p$ and Ω has a sufficiently regular boundary.

Clearly, if the factor space $\mathsf{B}^s_{pq}(\mathbb{R}^d, \langle x\rangle^\beta)|\Omega$ *coincides* (with equivalent norms) with the intrinsically defined space $\mathsf{B}^s_{pq}(\Omega, \langle x\rangle^\beta)$, then Corollary 2 immediately applies: If $\mathcal{F}(\Omega)$ is a bounded subset of $\mathsf{B}^s_{pq}(\Omega, \langle x\rangle^\beta)$ then

$$H_{[]}(\varepsilon, \mathcal{F}(\Omega), \|\cdot\|_{r,\mu}) \le H_{[]}(\varepsilon, \mathcal{F}|\Omega, \|\cdot\|_{r,\mu})$$

holds for a suitable $\mathcal{F}$ that satisfies the conditions of Corollary 2. This inequality together with Corollary 2 can then immediately be used to deliver rates for the (uniform) bracketing metric entropy of $\mathcal{F}(\Omega)$. Sufficient geometrical conditions on Ω such that $\mathsf{B}^s_{pq}(\Omega, \langle x\rangle^\beta)$ and $\mathsf{B}^s_{pq}(\mathbb{R}^d, \langle x\rangle^\beta)|\Omega$ coincide are available in the literature; see, e.g., [28, 3.4.2] and [30, 2.9.3, 4.2.2, 4.2.3] for results covering, in particular, the case of open half-spaces as well as of bounded C^∞-domains.

Remark 3 (‘Besov’ classes on not necessarily open subsets of $\mathbb{R}^d$) So far Besov spaces have been defined on open sets. In applications often ‘smooth’ function classes defined on a non-open Borel set $\tilde{\Omega}$ are of interest. If such function classes are defined via restriction from $\mathsf{B}^s_{pq}(\mathbb{R}^d, \langle x\rangle^\beta)$ to $\tilde{\Omega}$, Corollary 2 can of course be directly applied. Alternatively, if one insists upon defining such function classes in a more ‘intrinsic’ fashion, one can proceed as follows (for $s > d/p$): Suppose one can find a (sufficiently regular) domain Ω such that $\Omega \subseteq \tilde{\Omega} \subseteq \mathrm{cl}(\Omega)$, where $\mathrm{cl}(\cdot)$ denotes the closure, and such that $\mathsf{B}^s_{pq}(\Omega, \langle x\rangle^\beta)$ and $\mathsf{B}^s_{pq}(\mathbb{R}^d, \langle x\rangle^\beta)|\Omega$ coincide. Then one can define $\mathsf{B}^s_{pq}(\tilde{\Omega}, \langle x\rangle^\beta)$ simply as the set of all continuous functions $f : \tilde{\Omega} \to \mathbb{R}$ the restrictions $f|\Omega$ of which belong to $\mathsf{B}^s_{pq}(\Omega, \langle x\rangle^\beta)$, where the norm of f is given by the intrinsic (weighted) Besov norm of $f|\Omega$. We then immediately have for every bounded subset $\mathcal{F}(\tilde{\Omega})$ of $\mathsf{B}^s_{pq}(\tilde{\Omega}, \langle x\rangle^\beta)$, for every $1 \le r \le \infty$, and for any Borel measure μ on $\tilde{\Omega}$ that

$$H_{[]}(\varepsilon, \mathcal{F}(\tilde{\Omega}), \|\cdot\|_{r,\mu}) \le H_{[]}(\varepsilon, \mathcal{F}|\tilde{\Omega}, \|\cdot\|_{r,\mu})$$

holds for a suitable $\mathcal{F} \subseteq \mathsf{B}^s_{pq}(\mathbb{R}^d, \langle x\rangle^\beta)$ that satisfies the conditions of Corollary 2 (with $\tilde{\Omega}$ in place of Ω). The r.h.s. of the above inequality can then directly be bounded by Corollary 2.

3.3 Hölder, Sobolev, Triebel, and Related Spaces

3.3.1 Hölder Spaces

We define for integer $s > 0$ the space $\mathsf{C}^s(\mathbb{R}^d)$ as the set of all real-valued functions f on $\mathbb{R}^d$ for which

$$\|f\|_{s,\infty} := \sum_{0\le|\alpha|\le s} \|D^\alpha f\|_\infty$$

is finite and the classical partial derivatives $D^\alpha f$ are uniformly continuous for $|\alpha| = s$; for a definition of the multi-index α see Remark 2 above. For real but non-integer $s > 0$ we define $\mathsf{C}^s(\mathbb{R}^d)$ as the set of all real-valued functions f on $\mathbb{R}^d$ for which

$$\|f\|_{s,\infty} := \sum_{0\le|\alpha|\le[s]} \|D^\alpha f\|_\infty + \sum_{|\alpha|=[s]} \sup_{x\ne y} \frac{|D^\alpha f(x) - D^\alpha f(y)|}{|x-y|^{s-[s]}}$$

is finite. Here $[s]$ denotes the integer part of s and the derivatives are again to be understood in the classical sense. [For non-integer s we refer to $\mathsf{C}^s(\mathbb{R}^d)$ as a Hölder space, but we prefer to avoid this terminology in case s is integer, because there seems to be no universally accepted notion of a Hölder space in this case.] For $\beta \in \mathbb{R}$ define the weighted space

$$\mathsf{C}^s(\mathbb{R}^d, \langle x\rangle^\beta) = \{f : f \cdot \langle x\rangle^\beta \in \mathsf{C}^s(\mathbb{R}^d)\}$$

equipped with the norm $\|(\cdot)\langle x\rangle^\beta\|_{s,\infty}$. The results in Sect. 3.1 then imply the following corollary (for a discussion of a related result in [34], see Remark 5 below).

Corollary 3 *Let $\beta \in \mathbb{R}$, and $s > 0$. Let $\mathcal{F}$ be a (non-empty) bounded subset of $\mathsf{C}^s(\mathbb{R}^d, \langle x\rangle^\beta)$. Then the following statements hold:*

1. *Suppose $\beta > 0$ and $\beta \neq s$. Then $\mathcal{F}$ is a relatively compact subset of $\mathsf{UC}_0(\mathbb{R}^d)$ satisfying*
$$H(\varepsilon, \mathcal{F}, \|\cdot\|_\infty) \lesssim \varepsilon^{-d/\min(s,\beta)}.$$
2. *Suppose $\mathfrak{M}$ is a (non-empty) family of Borel measures on $\mathbb{R}^d$ such that the condition $\sup_{\mu\in\mathfrak{M}} \|\langle x\rangle^{\gamma-\beta}\|_{r,\mu} < \infty$ holds for some $1 \le r \le \infty$ and for some $\gamma > 0$ satisfying $\gamma \neq s$. Then*
$$\sup_{\mu\in\mathfrak{M}} H_{[]}(\varepsilon, \mathcal{F}, \|\cdot\|_{r,\mu}) \lesssim \varepsilon^{-d/\min(s,\gamma)}.$$

Proof The set $\mathcal{F}$ is a bounded subset of $\mathsf{B}^s_{\infty\infty}(\mathbb{R}^d, \langle x\rangle^\beta)$ by either 2.5.7/9, 2.5.7/6 (s non-integer) or 2.5.7/11 (s integer) of Triebel [28]. The result now follows from Theorem 1 and Corollary 1. □

3.3.2 Sobolev Spaces

For $1 < p < \infty$ and real $s \ge 0$, let
$$\mathcal{H}^s_p(\mathbb{R}^d) = \left\{ f \in \mathcal{L}^p(\mathbb{R}^d) : \|f\|_{s,p,\lambda} := \|F^{-1}(\langle x\rangle^s \cdot Ff)\|_{p,\lambda} < \infty \right\}$$
be the (real) Sobolev space, also known as Bessel-Potential space. [Recall that F represents the Fourier-transform acting on the space of complex tempered distributions.] We recall the well-known fact that for integer $s \ge 0$ an equivalent (semi)norm on $\mathcal{H}^s_p(\mathbb{R}^d)$ is given by
$$\|f\| = \sum_{0\le|\alpha|\le s} \|D^\alpha f\|_{p,\lambda}, \tag{5}$$
where D^α here denotes partial derivatives in the sense of distributions. [We note here that for fractional s a different definition of a Sobolev space is sometimes in use in the literature. This definition is less common nowadays as these spaces are just special cases of Besov spaces, see, e.g., Adams and Fournier [1, 7.30–7.34].] Similarly as in Sect. 2, one can now define the Banach spaces $\mathsf{H}^s_p(\mathbb{R}^d)$ consisting of continuous functions in case $s > d/p$. Furthermore, define for $s - d/p > 0$ the weighted Sobolev space
$$\mathsf{H}^s_p(\mathbb{R}^d, \langle x\rangle^\beta) = \left\{ f : f \cdot \langle x\rangle^\beta \in \mathsf{H}^s_p(\mathbb{R}^d) \right\}.$$

Corollary 4 *Let $1 < p < \infty$, $\beta \in \mathbb{R}$, and $s - d/p > 0$. Let $\mathcal{F}$ be a (non-empty) bounded subset of $\mathsf{H}^s_p(\mathbb{R}^d, \langle x\rangle^\beta)$.*

1. *Suppose $\beta > 0$. Then $\mathcal{F}$ is a relatively compact subset of $\mathsf{UC}_0(\mathbb{R}^d)$ satisfying*
$$H(\varepsilon, \mathcal{F}, \|\cdot\|_\infty) \lesssim \begin{cases} \varepsilon^{-d/s} & \text{for } \beta > s - d/p, \\ \varepsilon^{-(\beta/d+1/p)^{-1}} & \text{for } \beta < s - d/p. \end{cases}$$

2. *Suppose* $\mathfrak{M}$ *is a (non-empty) family of Borel measures on* $\mathbb{R}^d$ *such that the condition* $\sup_{\mu\in\mathfrak{M}} \|\langle x\rangle^{\gamma-\beta}\|_{r,\mu} < \infty$ *holds for some* $1 \le r \le \infty$ *and for some* $\gamma > 0$. *Then*

$$\sup_{\mu\in\mathfrak{M}} H_{[]}(\varepsilon, \mathcal{F}, \|\cdot\|_{r,\mu}) \lesssim \begin{cases} \varepsilon^{-d/s} & \text{for } \gamma > s - d/p, \\ \varepsilon^{-(\gamma/d+1/p)^{-1}} & \text{for } \gamma < s - d/p. \end{cases}$$

Proof Observe that $\mathcal{F}$ is a bounded subset of $\mathsf{B}^s_{p\infty}(\mathbb{R}^d, \langle x\rangle^\beta)$ by Triebel [28, 2.5.6/2 and 2.3.2/9, 5]. The result now follows from Theorem 1 and Corollary 1. □

Remark 4 (i) As common in the literature, the definition of $\mathcal{H}^s_p(\mathbb{R}^d)$ via the Fourier-transform excludes the case $p = 1$. For integer $s \ge 0$ and $p = 1$, the definition via the norm (5) gives rise to the well-known spaces

$$\mathcal{G}^s(\mathbb{R}^d) = \left\{ f \in \mathcal{L}^0(\mathbb{R}^d) : \sum_{0\le|\alpha|\le s} \|D^\alpha f\|_{1,\lambda} < \infty \right\},$$

also sometimes subsumed under the scale of Sobolev spaces. Since these spaces (and their weighted extensions) are continuously injected into $\mathcal{B}^s_{1\infty}(\mathbb{R}^d, \langle x\rangle^\beta)$ [28, 2.5.7/10], the results in Sect. 3.1 immediately apply whenever $s > d$.

(ii) In the particular case $s \ge p = d = 1$, any $[f]_\lambda$ with $f \in \mathcal{G}^s(\mathbb{R})$ contains an absolutely continuous representative, and hence $\mathcal{G}^s(\mathbb{R})$ is—up to a section—embedded into the Banach space $\mathsf{V}_1(\mathbb{R})$ of real-valued functions of bounded variation. An $\mathcal{L}^2(\mu)$-bracketing metric entropy bound of order ε^{-1} for bounded subsets of $\mathsf{V}_1(\mathbb{R})$ is given in [31] for μ an arbitrary probability measure.

3.3.3 Further Spaces

Our general results for (weighted) Besov spaces immediately imply similar entropy bounds for bounded subsets of many other (weighted) function spaces. This is so for Triebel spaces $F^s_{pq}(\mathbb{R}^d)$, since they can be embedded into a Besov space $B^s_{p\max(p,q)}(\mathbb{R}^d)$ ([28, 2.3.2/9]). Hölder-Zygmund space are also covered as they are identical (up to an equivalent norm) to the Besov spaces $\mathsf{B}^s_{\infty\infty}(\mathbb{R}^d)$ (see [28, 2.5.7/6]).

3.3.4 Function Classes on Subsets of $\mathbb{R}^d$

Clearly, entropy bounds for function classes defined on Borel subsets Ω of $\mathbb{R}^d$ that are obtained by restriction from the function spaces discussed in 3.3.1–3.3.3 above can be derived similarly; cf. Sect. 3.2.

4 Applications to Empirical Process Theory

In this section the Borel measures will always be probability measures on a Borel set $\Omega \subseteq \mathbb{R}^d$. Given $\mathcal{F} \subseteq \mathcal{L}^0(\Omega)$, the empirical measure $\mathbb{P}_n = 1/n \sum_{i=1}^n \delta_{X_i}$ of n independent random variables $X_1, \ldots, X_n$ distributed according to a law $\mathbb{P}$ induces a map from $\mathcal{F} \to \mathbb{R}$ given by $f \longmapsto \mathbb{P}_n f = 1/n \sum_{i=1}^n f(X_i)$. With $\mathbb{P} f := \int_\Omega f \, d\mathbb{P}$, the centered and scaled version of this map is the $\mathcal{F}$-indexed *empirical process* ν_n given by

$$f \longmapsto \nu_n(f) = \sqrt{n}(\mathbb{P}_n - \mathbb{P})f = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(f(X_i) - \mathbb{P}f).$$

A function class $\mathcal{F} \subseteq \mathcal{L}^2(\mathbb{P})$ is said to be $\mathbb{P}$-Donsker if it is pregaussian and if ν_n converges in law in $\ell^\infty(\mathcal{F})$ to a zero-mean Gaussian process $\mathbb{G}$ with covariance function $\mathbb{P}((f - \mathbb{P}f)(g - \mathbb{P}g))$ where $f, g \in \mathcal{F}$; see p. 94 in [11]. Here $\ell^\infty(\mathcal{F})$ denotes the Banach space of all bounded real valued functions on $\mathcal{F}$. [As on p. 91 in [11], we always assume the standard model.]

Given a (non-empty) family $\mathfrak{P}$ of probability measures, a function class $\mathcal{F}$ is said to be $\mathfrak{P}$-universal Donsker if $\mathcal{F}$ is $\mathbb{P}$-Donsker for all $\mathbb{P} \in \mathfrak{P}$. The class $\mathcal{F}$ is said to be a $\mathfrak{P}$-uniform Donsker class if the convergence of ν_n to $\mathbb{G}$ in $\ell^\infty(\mathcal{F})$ is uniform in $\mathfrak{P}$ in a sense made precise in Giné and Zinn [18]; see also Sheehy and Wellner [26] and Talagrand [27]. The $\mathfrak{P}$-uniform Donsker property obviously implies the $\mathfrak{P}$-universal Donsker property. We shall omit the prefix $\mathfrak{P}$ in both of these concepts if $\mathfrak{P}$ is the set of all probability measures on the underlying set Ω.

4.1 Uniform Limit Theorems for Besov Classes

Before we discuss the (uniform) empirical central limit theorem for Besov classes in detail, we shortly outline the ramifications of our entropy bounds for Glivenko-Cantelli results. The bracketing metric entropy bounds for bounded subsets $\mathcal{F}$ of Besov (and related) spaces obtained in Sect. 3 together with Theorem 7.1.5 in [11] immediately imply such results for $\mathcal{F}$. Furthermore, Corollary 1 implies $\mathfrak{P}$-uniform Glivenko-Cantelli results for $\mathcal{F}$ by using Theorem 6 in [13]. Finally, the entropy bounds in the present paper also allow one to derive rates of convergence in Glivenko-Cantelli results, e.g., by using results in [38, 39].

The following results provide a comprehensive description of Donsker properties of bounded subsets of (weighted) Besov spaces $\mathsf{B}^s_{pq}(\mathbb{R}^d, \langle x\rangle^\beta)$. These results immediately carry over to bounded subsets of the function spaces treated in Sect. 3.3.

Corollary 5 *Let* $1 \le p \le \infty$, $1 \le q \le \infty$, $\beta \in \mathbb{R}$, *and* $s - d/p > 0$. *Let* $\mathcal{F}$ *be a (non-empty) bounded subset of* $\mathsf{B}^s_{pq}(\mathbb{R}^d, \langle x\rangle^\beta)$. *Let* $\mathfrak{P}$ *be a (non-empty) family of probability measures on* $\mathbb{R}^d$ *such that*

$$\sup_{\mathbb{P}\in\mathfrak{P}} \|\langle x\rangle^{\gamma-\beta}\|_{2,\mathbb{P}} < \infty \tag{6}$$

holds for some $\gamma > 0$. *If* $\gamma/d + 1/p > 1/2$ *and* $s/d > 1/2$, *then* $\mathcal{F}$ *is a* $\mathfrak{P}$*-uniform Donsker class.*

Proof Consider first the case where $\gamma/d + 1/p \neq s/d$. Then Corollary 1 implies that there exists a finite constant $C > 0$ such that

$$\sup_{\mathbb{P}\in\mathfrak{P}} H_{[]}(\varepsilon, \mathcal{F}, \|\cdot\|_{2,\mathbb{P}}) \le C\varepsilon^{-1/\alpha},$$

where $\alpha = s/d$ if $\gamma/d + 1/p > s/d$ and $\alpha = \gamma/d + 1/p$ if $\gamma/d + 1/p < s/d$. Hence, the condition on the convergence of the bracketing integral in Theorem 2.8.4 of van der Vaart and Wellner [36] is satisfied in both cases. Observe that (6) implies for the envelope $K\langle x\rangle^{-\beta}$ of $\mathcal{F}$ (cf. Part 3 of Proposition 3 in the Appendix)

$$\lim_{M\to\infty} \sup_{\mathbb{P}\in\mathfrak{P}} \left\| K\langle x\rangle^{-\beta} \cdot \mathbf{1}_{[K\langle x\rangle^{-\beta}>M]} \right\|_{2,\mathbb{P}} = 0,$$

which verifies the envelope condition in Theorem 2.8.4 of van der Vaart and Wellner [36].

In the case where $\gamma/d + 1/p = s/d$ we reduce this to the case just established: Choose $0 < \gamma' < \gamma$ such that $\gamma'/d + 1/p > 1/2$ still holds, and observe that (6) a fortiori holds with γ' replacing γ. □

An analogous result for bounded subsets of (weighted) Besov spaces defined over Borel subsets Ω of $\mathbb{R}^d$ can similarly be obtained from the results in Sect. 3.2.

We note that the moment condition (6) is trivially satisfied irrespective of the particular choice of $\mathfrak{P}$ in case $\gamma \le \beta$. Therefore, if $s > d/p$, $\beta > 0$, and $\min(\beta/d + 1/p, s/d) > 1/2$, the set $\mathcal{F}$ is in fact a uniform Donsker class. In particular, if $s > d/p$ and $p \le 2$, $\mathcal{F}$ is a uniform Donsker class for *every* $\beta > 0$.

In contrast, in case $p \le 2$ but $\beta = 0$, Corollary 5 requires a moment condition, albeit of arbitrary small order. The reason for this is that the bracketing methods used in the present paper prohibit the case $\gamma = 0$ in the above corollary. Nevertheless, at least the universal Donsker property can be proved by different methods. For this we use a result of Marcus [23], which in turn builds on Giné [15].

Proposition 1 *Suppose* $1 \le p \le 2$, $1 \le q \le \infty$, *and* $s - d/p > 0$. *Then any* (*non-empty*) *bounded subset* $\mathcal{F}$ *of* $\mathsf{B}^s_{pq}(\mathbb{R}^d)$ *is a universal Donsker class.*

Proof Note that $\mathcal{F}$ is a bounded subset of $\mathsf{B}^r_{2q}(\mathbb{R}^d)$ for some $r > d/2$ by Triebel [28, 2.7.1/1], and hence of $\mathsf{B}^t_{22}(\mathbb{R}^d)$ for some t, $r > t > d/2$, by Triebel [28, 2.3.2/7]. Now, $\mathsf{B}^t_{22}(\mathbb{R}^d)$ is equal to the Sobolev space $\mathsf{H}^t_2(\mathbb{R}^d)$ by Triebel [29, 1.3.2/7]. Bounded subsets of $\mathsf{H}^t_2(\mathbb{R}^d)$ with $t > d/2$ are universally Donsker by Theorem 1.3 in Marcus [23]. □

At least for $d = 1$ and $p < 2$, even the uniform Donsker property can be proved. This is done in the following theorem, which in fact establishes the uniform Donsker property also for the limiting case $s = 1/p$ and $q = 1$, not covered by any of the preceding results.

Theorem 2 *Suppose* $1 \le p < 2$ *and* $d = 1$. *If either* $1 \le q \le \infty$ *and* $s - 1/p > 0$, *or* $q = 1$ *and* $s = 1/p$, *then any* (*non-empty*) *bounded subset* $\mathcal{F}$ *of* $\mathsf{B}^s_{pq}(\mathbb{R})$ *is a uniform Donsker class.*

Proof First observe that $\mathsf{B}^s_{pq}(\mathbb{R})$ for $s - 1/p > 0$ is embedded into $\mathsf{B}^{1/p}_{p1}(\mathbb{R})$, see Triebel [28, 2.3.2/7]. Hence, it suffices to give the proof for the case $q = 1$ and $s = 1/p$.

For a measurable function $f:\mathbb{R}\to\mathbb{R}$ its π-variation ($1\le\pi<\infty$) is defined as

$$v_\pi(f)=\sup\left\{\sum_{i=1}^n |f(x_i)-f(x_{i-1})|^\pi : -\infty<x_0<x_1<\cdots<x_n<\infty,\ n\in\mathbb{N}\right\}$$

and $\mathsf{V}_\pi(\mathbb{R})$ denotes the Banach space of functions for which $v_\pi(f)<\infty$, endowed with the norm $\|f\|_{v_\pi}:=\|f\|_\infty+(v_\pi(f))^{1/\pi}$. We next show that $\mathsf{B}^{1/p}_{p1}(\mathbb{R})$ is embedded into $\mathsf{V}_p(\mathbb{R})$. Given this, the proof will then be complete since bounded subsets of $\mathsf{V}_p(\mathbb{R})$ with $1\le p<2$ are uniform Donsker classes in view of Theorem 2.2 in [10].

To prove the embedding consider first the case $p=1$. Note that $\mathsf{B}^1_{11}(\mathbb{R})$ is contained in the linear space of continuous functions f with a Lebesgue-integrable generalized derivative (i.e., the space of absolutely continuous functions), see Triebel [28, 2.5.7/10]. Hence, by Theorem 5.3.5 in [40], the essential total variation of every such f is finite. Since f is continuous, this implies that $v_1(f)$ is finite, and hence so is $\|f\|_{v_1}<\infty$. This establishes the set-inclusion $\mathsf{B}^1_{11}(\mathbb{R})\subseteq\mathsf{V}_1(\mathbb{R})$. That this inclusion is in fact an embedding now follows from the closed graph theorem (cf. the first paragraph in the Appendix). [In view of Part 1 of Proposition 3 in the Appendix and the definition of $\mathsf{V}_1(\mathbb{R})$, it follows that norm convergence in both spaces implies convergence everywhere.]

In case $1<p<2$, consider the following chain of embeddings

$$\mathsf{B}^{1/p}_{p1}(\mathbb{R})\hookrightarrow\dot{\mathcal{B}}^{1/p}_{p1}(\mathbb{R})\cap\mathsf{C}_0(\mathbb{R})\hookrightarrow V_p(\mathbb{R}),\tag{7}$$

where $\dot{\mathcal{B}}^{1/p}_{p1}(\mathbb{R})$ is the homogeneous Besov space defined in Sect. 4.1 of Bourdaud, de Cristoforis and Sickel [8]. The intersection $\dot{\mathcal{B}}^{1/p}_{p1}(\mathbb{R})\cap\mathsf{C}_0(\mathbb{R})$ in the above display is endowed with the restriction of the (semi)norm on $\dot{\mathcal{B}}^{1/p}_{p1}(\mathbb{R})$ and is a Banach space (Proposition 10 in Bourdaud, de Cristoforis and Sickel [8]). The space $V_p(\mathbb{R})$ is defined in Sect. 2.2 of the same reference (where it is denoted by $BV_p(\mathbb{R})$).

To prove the first embedding in (7) observe that

$$\mathsf{B}^{1/p}_{p1}(\mathbb{R})=\mathcal{L}^p(\mathbb{R},\lambda)\cap\dot{\mathcal{B}}^{1/p}_{p1}(\mathbb{R})\cap\mathsf{C}_0(\mathbb{R})$$

(cf. Definition 4 in Bourdaud, de Cristoforis and Sickel [8], Remark 2 in Sect. 2, and Proposition 3 in the Appendix), and hence

$$\mathsf{B}^{1/p}_{p1}(\mathbb{R})\subseteq\dot{\mathcal{B}}^{1/p}_{p1}(\mathbb{R})\cap\mathsf{C}_0(\mathbb{R})$$

holds. Both spaces are Banach spaces. In view of Part 1 of Proposition 3 in the Appendix and of the first part of Proposition 10 in Bourdaud, de Cristoforis and Sickel [8], it follows that norm convergence in both spaces implies convergence everywhere. The first embedding in (7) now follows from the closed graph theorem (cf. the Appendix).

The second embedding in (7) is to be understood as a continuous quotient map (i.e., it associates to each function $f\in\dot{\mathcal{B}}^{1/p}_{p1}(\mathbb{R})\cap\mathsf{C}_0(\mathbb{R})$ the equivalence class $[g]_\lambda$ such that $f\in[g]_\lambda\in V_p(\mathbb{R})$). The existence of this embedding follows from a modification of a theorem of Peetre given as Theorem 5 in Bourdaud, de Cristoforis and Sickel [8]. This establishes the chain of embeddings (7).

Now, by definition of the space $V_p(\mathbb{R})$, it follows that every $f \in \mathsf{B}^{1/p}_{p1}(\mathbb{R})$ coincides with a function g in $\mathsf{V}_p(\mathbb{R})$ outside of a Lebesgue null-set N. Fix $\varepsilon > 0$ and a grid of points such that $-\infty < x_0 < x_1 < \cdots < x_n < \infty$. Since f is uniformly continuous by Part 1 of Proposition 3 in the Appendix and, since $\mathbb{R}\backslash N$ is dense in $\mathbb{R}$, we can find $x_i^* \in \mathbb{R}\backslash N$ such that $|f(x_i) - f(x_i^*)| < (\varepsilon/n)^{1/p}$. Observing that f and g agree outside of N we obtain

$$\begin{aligned}
&\sum_{i=1}^{n} |f(x_i) - f(x_{i-1})|^p \\
&\quad \le \sum_{i=1}^{n} (|f(x_i) - f(x_i^*)| + |f(x_i^*) - f(x_{i-1}^*)| + |f(x_{i-1}^*) - f(x_{i-1})|)^p \\
&\quad \le 2^{2p-2}\left[\sum_{i=1}^{n} |f(x_i) - f(x_i^*)|^p + |f(x_i^*) - f(x_{i-1}^*)|^p + |f(x_{i-1}^*) - f(x_{i-1})|^p\right] \\
&\quad \le 2^{2p-1}\varepsilon + 2^{2p-2}\sum_{i=1}^{n} |g(x_i^*) - g(x_{i-1}^*)|^p \le 2^{2p-1}\varepsilon + 2^{2p-2} v_p(g).
\end{aligned}$$

This shows that $v_p(f)$ is finite. Since also $\|f\|_\infty < \infty$ holds by Proposition 3 in the Appendix, we have $\|f\|_{v_p} < \infty$. Summarizing, we have established the set-inclusion $\mathsf{B}^{1/p}_{p1}(\mathbb{R}) \subseteq \mathsf{V}_p(\mathbb{R})$ for $1 < p < 2$. That this inclusion is in fact an embedding now follows from the closed graph theorem (cf. the Appendix). □

The above result is proved by showing that the set $\mathcal{F}$ is of bounded p-variation on $\mathbb{R}$ under the particular conditions of Theorem 2 and by a subsequent application of a theorem due to Dudley [10] which establishes the uniform Donsker property for function classes of bounded p-variation, $p < 2$, on $\mathbb{R}$. Whether the unit ball in $\mathsf{B}^s_{2q}(\mathbb{R})$ with $s > 1/2$, which is universally Donsker by the proposition above, is also a *uniform* Donsker class is an open question. We conjecture that the answer is in the negative.

In case $\beta = 0$ and $p > 2$, the unit ball in $\mathsf{B}^s_{pq}(\mathbb{R}^d)$ can be shown not to be a universal Donsker class (even if $s/d > 1/2 > 1/p$) although it is uniformly sup-norm bounded, cf. [24]. [A similar result is true for the unit ball in $\mathsf{B}^s_{pq}(\mathbb{R}^d, \langle x\rangle^\beta)$ in case $\beta > 0$, $p > 2$, $s > d/p$ and $\min(\beta/d + 1/p, s/d) < 1/2$ (even if $s/d > 1/2 > 1/p$).]

Finally, in case $\beta < 0$, the moment condition in Corollary 5 can be expressed as an integrability condition on the envelope $F = K\langle x\rangle^{-\beta}$ of $\mathcal{F}$, i.e., (6) amounts to

$$\sup_{\mathbb{P}\in\mathfrak{P}} \int_{\mathbb{R}^d} F^{2+\delta} d\mathbb{P} < \infty \tag{8}$$

for $\delta = -2\gamma/\beta > 0$. [Note that in case $p \le 2$ the constant γ, and hence δ, can be chosen arbitrarily close to, but not equal to zero in Corollary 5.] This clearly shows that more than just $F \in \mathcal{L}^2(\mathbb{P})$ is required to obtain the $\mathbb{P}$-Donsker property from Corollary 5. A similar phenomenon has been observed by Dudley and Koltchinskii [12] for VC major classes; see also Corollary 3.1 in [35] and Theorems 9.5.4 and 9.5.5 in [11].

A message of the preceding two paragraphs is the following: The typical sufficient conditions for $\mathbb{P}$-Donsker theorems, namely $\mathcal{L}^2(\mathbb{P})$-integrability of the envelope together with convergence of the entropy integral, are not independent but are intertwined (at least if the sample space is $\mathbb{R}^d$). The point here is that a sufficiently fast decay at infinity of the envelope F beyond $F \in \mathcal{L}^2(\mathbb{P})$ is sometimes required to guarantee a convergent entropy integral. This is related to the fact that the bracketing metric entropy bounds obtained in Theorem 1 depend on the parameters β and p (governing the decay of the function class $\mathcal{F}$) on the one hand and on γ (governing the tail behavior of the probability measure $\mathbb{P}$) on the other hand; cf. also [24].

Remark 5 Extending results by Giné and Zinn [16, 17] and Arcones [4], van der Vaart [34] considers Lipschitz-type classes on $\mathbb{R}^d$ and obtains bounds for their $\mathcal{L}^r(\mathbb{P})$-bracketing metric entropy for $r < \infty$. [His function classes are such that the restrictions of each element to convex uniformly bounded subsets I_j partitioning $\mathbb{R}^d$ belong to a ball of radius M_j in the Lipschitz space over I_j.] His envelope conditions for these function classes to be $\mathbb{P}$-Donsker are similar to the moment condition in Corollary 5 for $p = q = \infty$, which covers bounded subsets of the closely related Hölder spaces as a special case; cf. Sect. 3.3. [It follows from this corollary that a bounded subset of the space $\mathsf{C}^s(\mathbb{R}^d, \langle x\rangle^\beta)$ is $\mathbb{P}$-Donsker, if $s > d/2$ and $\|\langle x\rangle^{\gamma-\beta}\|_{2,\mathbb{P}} < \infty$ holds for some $\gamma > d/2$.] Clearly, the "partitioning" idea underlying van der Vaart [34, 35] can also be combined with our results in Sect. 3.2 to construct further Donsker classes.

Acknowledgements We are most grateful to an anonymous referee and to V. Losert for a careful reading of a previous version and constructive criticism and suggestions which have helped to improve the paper. We also would like to acknowledge helpful comments from H.G. Feichtinger, E. Giné, H. Rindler, H. Triebel, and W.P. Ziemer.

Appendix

A normed space $(X, \|\cdot\|_X)$ is said to be *embedded* into the normed space $(Y, \|\cdot\|_Y)$ if X is a linear subspace of Y and if the identity map $\mathrm{id} : X \to Y$ is continuous. We shall write

$$(X, \|\cdot\|_X) \hookrightarrow (Y, \|\cdot\|_Y)$$

to denote such an embedding. We shall also use the symbol $\hookrightarrow$ more generally if the embedding of X in Y includes a quotient or section map. [In this case the identity map is to be replaced by the appropriate linear map.] The embedding is said to be compact if the image (under the embedding operator) of the unit ball of X is precompact (i.e., totally bounded) in Y. Furthermore, we note here the following well-known consequence of the closed graph theorem: Let X and Y be linear subspaces of the vector space of real-valued functions on a (non-empty) set $\Omega \subseteq \mathbb{R}^d$ satisfying $X \subseteq Y$. Let X and Y be equipped with norms $\|\cdot\|_X$ and $\|\cdot\|_Y$, respectively, such that they become Banach spaces. Furthermore, suppose that each of this norm-topologies has the property that any norm-convergent sequence possesses a subsequence that converges everywhere. Then the map $\mathrm{id} : X \to Y$ has a closed graph, and is thus continuous.

We recall the following definition of the *entropy numbers*:

Definition 5 Let $\mathcal{J}$ be a subset of the normed space $(Y, \|\cdot\|_Y)$, and let $U_Y = \{y \in Y : \|y\|_Y \le 1\}$ be the closed unit ball in Y. Then, for all natural numbers k, the k-th entropy number of $\mathcal{J}$ is defined as

$$e(k, \mathcal{J}, \|\cdot\|_Y) = \inf\left\{\varepsilon > 0 : \mathcal{J} \subseteq \bigcup_{j=1}^{2^{k-1}} (y_j + \varepsilon U_Y) \text{ for some } y_1, \dots, y_{2^{k-1}} \in Y\right\},$$

with the convention that the infimum equals $+\infty$ if the set over which it is taken is empty.

Suppose $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ are normed spaces such that X is a linear subspace of Y. Let U_X denote the closed unit ball in X. Then, $e(k, \mathrm{id}(U_X), \|\cdot\|_Y)$ is called the k-th entropy number of the operator $\mathrm{id} : X \to Y$. Clearly, $e(k, \mathrm{id}(U_X), \|\cdot\|_Y)$ is finite for all k if and only if X is embedded into Y, and the entropy numbers converge to zero as $k \to \infty$ if and only if the embedding is compact.

We next state a special case of more general results due to Haroske and Triebel [19, 20].

Proposition 2 (Haroske and Triebel) *Suppose* $1 \le p \le \infty$, $1 \le q \le \infty$, $s - d/p > 0$, $\beta \in \mathbb{R}$, *and* $\gamma > 0$ *hold. Then* $B^s_{pq}(\mathbb{R}^d, \langle x\rangle^\beta)$ *is compactly embedded into* $B^0_{\infty 1}(\mathbb{R}^d, \langle x\rangle^{\beta-\gamma})$. *Furthermore, the entropy numbers of this embedding satisfy*

$$e\big(k, \mathrm{id}(U_{B^s_{pq}(\mathbb{R}^d,\langle x\rangle^\beta)}), \|(\cdot)\langle x\rangle^{\beta-\gamma}\|_{0,\infty,1,\lambda}\big) \sim \begin{cases} k^{-s/d} & \text{for } \gamma > s - d/p, \\ k^{-(\gamma/d+1/p)} & \text{for } \gamma < s - d/p \end{cases}$$

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

Proof Clearly, it suffices to consider the case $\beta = \gamma > 0$ since $B^s_{pq}(\mathbb{R}^d, \langle x\rangle^\delta)$ and $B^s_{pq}(\mathbb{R}^d, \langle x\rangle^\vartheta)$ are isometrically isomorphic via the map $[f]_\lambda \longmapsto [f\langle x\rangle^{\delta-\vartheta}]_\lambda$. For complex Besov spaces the result now follows as a special case of Theorem 4.1 in [20], noting that the norms used in that reference are equivalent to the weighted norms used here; cf. Theorem 4.2.2 in [14]. In view of Remark 1 we may apply Lemma 1 given below with $X = B^0_{\infty 1}(\mathbb{R}^d)$,

$$Y = \big\{T \in \mathcal{S}'(\mathbb{R}^d) : \|T\|_{0,\infty,1,\lambda} < \infty\big\} = B^0_{\infty 1}(\mathbb{R}^d) + iB^0_{\infty 1}(\mathbb{R}^d),$$

$$A = \mathrm{id}\big(\big\{T \in \mathcal{S}'(\mathbb{R}^d) : \|\langle x\rangle^\beta T\|_{s,p,q,\lambda} \le 1\big\}\big),$$

$B = \mathrm{id}(U_{B^s_{pq}(\mathbb{R}^d,\langle x\rangle^\beta)})$, and $D = cB$, where c was defined in Remark 1(ii). This now shows that Theorem 4.1 in [20] carries over to the case of real Besov spaces. □

For a discussion of the case $\gamma = s - d/p$, see Sect. 4.2 in [20].

We next summarize some properties of weighted Besov spaces which have been used in the paper. In particular, the first part justifies the definition of $\mathsf{B}^s_{pq}(\mathbb{R}^d, \langle x\rangle^\beta)$ in Sect. 2.

Proposition 3 *Let $1 \le p \le \infty$, $1 \le q \le \infty$, and $\beta \in \mathbb{R}$. Suppose either $s - d/p > 0$, or $s - d/p = 0$ and $q = 1$.*

1. *If $\beta \ge 0$, we have that $B^s_{pq}(\mathbb{R}^d, \langle x\rangle^\beta)$ is embedded (up to a section) in the Hölder space $\mathsf{C}^{s-d/p}(\mathbb{R}^d)$ (with the convention that $\mathsf{C}^0(\mathbb{R}^d) = \mathsf{UC}(\mathbb{R}^d)$).*
2. *If $\beta > 0$, or $\beta = 0$ and $p < \infty$, we have that $\lim_{\|x\|\to\infty} f(x) = 0$ for all $f \in \mathsf{B}^s_{pq}(\mathbb{R}^d, \langle x\rangle^\beta)$.*
3. *Let $\mathcal{F}$ be a (non-empty) bounded subset of $\mathsf{B}^s_{pq}(\mathbb{R}^d, \langle x\rangle^\beta)$. Then $\sup_{f\in\mathcal{F}} |f(x)| \le K\langle x\rangle^{-\beta}$ holds for some real number $K \ge 0$.*

Proof Since $B^s_{pq}(\mathbb{R}^d, \langle x\rangle^\beta) \hookrightarrow B^s_{pq}(\mathbb{R}^d)$ for $\beta \ge 0$ (Sect. 4.2.3 in [14]), it is sufficient to prove Part 1 only for the case $\beta = 0$. Under the conditions of the proposition we have the following embedding (up to a section)

$$B^s_{pq}(\mathbb{R}^d) \hookrightarrow \mathsf{C}^{s-d/p}(\mathbb{R}^d) \hookrightarrow \mathsf{UC}(\mathbb{R}^d) \tag{9}$$

[28, 2.7.1/12-13, 2.2.2/18], which then implies the first part. To prove the third part, observe that (9) implies the existence of a real number c such that for all $f \in \mathcal{F}$

$$\sup_{x\in\mathbb{R}^d} |f(x)\cdot\langle x\rangle^\beta| = \|f\cdot\langle x\rangle^\beta\|_\infty \le c\|f\cdot\langle x\rangle^\beta\|_{s,p,q,\lambda} \le cb$$

holds, where $b < \infty$ is a Besov-norm bound for $\mathcal{F}$. This then gives $|f| \le K\langle x\rangle^{-\beta}$ for $K = cb \ge 0$ and all $f \in \mathcal{F}$. For $\beta > 0$, the second part now follows immediately.

It remains to prove the second part for $\beta = 0$ and $p < \infty$. Let $\mathcal{S}$ be the Schwartz space of rapidly decreasing infinitely differentiable complex-valued functions on $\mathbb{R}^d$. The set $\mathcal{S}_\mathbb{R} = \{f \in \mathcal{S} : f = \bar{f}\}$ is dense in the Banach space $\mathsf{B}^s_{pq}(\mathbb{R}^d)$ for $p < \infty$ and $q < \infty$ as a consequence of Theorem 2.3.3 in [28] (and Remark 1(ii)). Since $\|\cdot\|_\infty \lesssim \|\cdot\|_{s,p,q,\lambda}$ on $\mathsf{B}^s_{pq}(\mathbb{R}^d)$ by (9) above and since $\mathsf{C}_0(\mathbb{R}^d)$ is the $\|\cdot\|_\infty$-completion of $\mathcal{S}_\mathbb{R}$, the conclusion of the proposition follows for $q < \infty$. If $q = \infty$, under the assumptions of the proposition $s - d/p > 0$ follows, and hence $\mathsf{B}^s_{p\infty}(\mathbb{R}^d)$ is embedded into $\mathsf{B}^{d/p}_{p1}(\mathbb{R}^d)$ in view of Triebel [28, 2.3.2/7], thus reducing the case $q = \infty$ to the case $q = 1 < \infty$ just established. □

We finally state two auxiliary results (for the second one cf. also [36, Theorem 2.7.11]).

Lemma 1 *Let $(X, \|\cdot\|_X)$ be a normed space over $\mathbb{R}$ and let $Y = X + iX$ be its complexification. [That is, Y is a vector space over $\mathbb{C}$ where each $y \in Y$ can be written uniquely as $y = x_1 + ix_2$ with $x_1 =: \operatorname{Re} y \in X$ and $x_2 =: \operatorname{Im} y \in X$.] Let Y be equipped with a norm $\|\cdot\|_Y$ such that for some $0 < C_1 \le C_2 < \infty$*

$$C_1\|y\|_Y \le \|\operatorname{Re} y\|_X + \|\operatorname{Im} y\|_X \le C_2\|y\|_Y \tag{10}$$

holds for all $y \in Y$. Furthermore, let A be a (non-empty) subset of Y.

(i) *If $B \subseteq \operatorname{Re} A := \{\operatorname{Re} y : y \in A\} \subseteq X$ then for every $k \in \mathbb{N}$*

$$e(k, B, \|\cdot\|_X) \le C_2 e(k, A, \|\cdot\|_Y).$$

(ii) *If $D \subseteq X$ is such that $A \subseteq D + iD$ (e.g., $D = \operatorname{Re} A \cup \operatorname{Im} A$) then for every $k \in \mathbb{N}$*

$$e(2k-1, A, \|\cdot\|_Y) \le 2C_1^{-1} e(k, D, \|\cdot\|_X).$$

Proof If $e(k, A, \|\cdot\|_Y) = \infty$, Part (i) is trivial. For every $\varepsilon > e(k, A, \|\cdot\|_Y)$, every covering of A by 2^{k-1} closed balls in Y of radius ε induces (via the projection $\operatorname{Re}(\cdot)$) a covering of $\operatorname{Re} A$, and hence of B, by at most 2^{k-1} closed balls in X of radius $C_2\varepsilon$ in view of (10). This immediately proves (i). To prove (ii) we may assume that $e(k, D, \|\cdot\|_X) < \infty$. Let $\varepsilon > e(k, D, \|\cdot\|_X)$ be arbitrary. Observe that any covering of D by 2^{k-1} closed balls K_j in X with center x_j and of radius ε induces a covering of $D + iD \supseteq A$ by 2^{2k-2} closed balls in Y of radius $2C_1^{-1}\varepsilon$ with centers $x_j + ix_l$, $1 \le i, l \le 2^{k-1}$. This completes the proof. □

Lemma 2 *Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be normed spaces and let $\mathcal{F}$ be a (non-empty) subset of X with finite metric entropy $H(\varepsilon, \mathcal{F}, \|\cdot\|_X)$ for all ε, $0 < \varepsilon < \infty$. Let the map $A : \mathcal{F} \to (Y, \|\cdot\|_Y)$ satisfy $\|A(x_1) - A(x_2)\|_Y \le C\|x_1 - x_2\|_X^\sigma$ for all $x_1, x_2 \in \mathcal{F}$ and for some $\sigma > 0$, $0 < C < \infty$. We then have for the image $A(\mathcal{F})$ that*

$$H(\varepsilon, A(\mathcal{F}), \|\cdot\|_Y) \le H\big(2^{-1}C^{-1/\sigma}\varepsilon^{1/\sigma}, \mathcal{F}, \|\cdot\|_X\big)$$

for every ε, $0 < \varepsilon < \infty$. [If $C = 0$, then $H(\varepsilon, A(\mathcal{F}), \|\cdot\|_Y) = 0$ for every ε, $0 < \varepsilon < \infty$.]

Proof Let B_i, $i = 1, \ldots, N(2^{-1}C^{-1/\sigma}\varepsilon^{1/\sigma}, \mathcal{F}, \|\cdot\|_X)$ be closed balls of radius $2^{-1}C^{-1/\sigma}\varepsilon^{1/\sigma}$ covering $\mathcal{F}$. For each i choose x_i from $B_i \cap \mathcal{F}$, which obviously is non-empty. Then

$$B_i^* = \Big\{ y \in Y : \|y - A(x_i)\|_Y \le \sup_{x \in B_i \cap \mathcal{F}} \|A(x) - A(x_i)\|_Y \Big\}$$

is a closed ball in Y containing $A(B_i \cap \mathcal{F})$, hence the union of all balls B_i^* covers $A(\mathcal{F})$. The radius of B_i^* is less than or equal to ε since

$$\sup_{x \in B_i \cap \mathcal{F}} \|A(x) - A(x_i)\|_Y \le \sup_{x \in B_i \cap \mathcal{F}} C\|x - x_i\|_X^\sigma \le \varepsilon.$$

Thus $H(\varepsilon, A(\mathcal{F}), \|\cdot\|_Y) \le H(2^{-1}C^{-1/\sigma}\varepsilon^{1/\sigma}, \mathcal{F}, \|\cdot\|_X)$. The claim in parentheses is obvious. □

References

1. Adams, R.A., Fournier, J.F.: Sobolev Spaces, 2nd edn. Academic, New York (2003)
2. Alexander, K.S.: Probability inequalities for empirical processes and a law of the iterated logarithm. Ann. Probab. **12**, 1041–1067 (1984)
3. Andersen, N.T., Giné, E., Ossiander, M., Zinn, J.: The central limit theorem and the law of the iterated logarithm for empirical processes under local conditions. Probab. Theory Relat. Fields **77**, 271–305 (1988)
4. Arcones, M.A.: The central limit theorem for U-processes indexed by Hölder's functions. Stat. Probab. Lett. **20**, 57–62 (1994)

5. Birgé, L., Massart, P.: Rates of convergence of minimum contrast estimators. Probab. Theory Relat. Fields **97**, 113–150 (1993)
6. Birgé, L., Massart, P.: An adaptive compression algorithm in Besov spaces. Constr. Approx. **16**, 1–36 (2000)
7. Birman, M.S., Solomyak, M.Z.: Piecewise polynomial approximations of functions of the classes W_α^p. Mat. Sb. **73**, 331–355 (1967)
8. Bourdaud, G., De Cristoforis, M.L., Sickel, W.: Superposition operators and functions of bounded p-variation. Rev. Mat. Iberoamer. **22**, 455–487 (2006)
9. Dudley, R.M.: Central limit theorems for empirical measures. Ann. Probab. **6**, 899–929 (1978); Correction, ibid. **7**, 909–911 (1979)
10. Dudley, R.M.: Fréchet differentiability, p-variation and uniform Donsker classes. Ann. Probab. **20**, 1968–1982 (1992)
11. Dudley, R.M.: Uniform Central Limit Theorems. Cambridge University Press, Cambridge (1999)
12. Dudley, R.M., Koltchinskii, V.I.: Envelope moment conditions and Donsker classes. Theor. Probab. Math. Stat. **51**, 39–48 (1994)
13. Dudley, R.M., Giné, E., Zinn, J.: Uniform and universal Glivenko-Cantelli classes. J. Theor. Probab. **4**, 485–510 (1991)
14. Edmunds, D.E., Triebel, H.: Function Spaces, Entropy Numbers and Differential Operators. Cambridge University Press, Cambridge (1996)
15. Giné, E.: Invariant tests for uniformity on compact Riemannian manifolds based on Sobolev-norms. Ann. Stat. **3**, 1243–1266 (1975)
16. Giné, E., Zinn, J.: Lectures on the central limit theorem for empirical processes. In: Lecture Notes in Math., vol. 1221, pp. 50–113. Springer, New York (1986)
17. Giné, E., Zinn, J.: Empirical processes indexed by Lipschitz functions. Ann. Probab. **14**, 1329–1338 (1986)
18. Giné, E., Zinn, J.: Gaussian characterization of uniform Donsker classes. Ann. Probab. **19**, 758–782 (1991)
19. Haroske, D., Triebel, H.: Entropy numbers in weighted function spaces and eigenvalue distributions of some degenerate pseudodifferential operators I. Math. Nachr. **167**, 131–156 (1994)
20. Haroske, D., Triebel, H.: Wavelet bases and entropy numbers in weighted function spaces. Math. Nachr. **278**, 108–132 (2005)
21. Kolmogorov, A.N., Tihomirov, V.M.: ε-entropy and ε-capacity of sets in function spaces. In: AMS Translations, Series 2, vol. 17, pp. 277–364 (1961)
22. Lorentz, G.G., v.Golitschek, M., Makovoz, Y.: Constructive Approximation: Advanced Problems. Springer, Berlin (1996)
23. Marcus, D.J.: Relationships between Donsker classes and Sobolev spaces. Z. Wahrsch. Verw. Gebiete **69**, 323–330 (1985)
24. Nickl, R.: Empirical and Gaussian processes on Besov classes. In: Giné, E., Koltchinskii, V., Li, W., Zinn, J. (eds.) High Dimensional Probability IV. IMS Lecture Notes, vol. 51, pp. 185–195 (2006)
25. Ossiander, M.: A central limit theorem under metric entropy with L_2 bracketing. Ann. Probab. **15**, 897–919 (1987)
26. Sheehy, A., Wellner, J.A.: Uniform Donsker classes of functions. Ann. Probab. **20**, 1983–2030 (1992)
27. Talagrand, M.: Vapnik-Chervonenkis type conditions and uniform Donsker classes of functions. Ann. Probab. **31**, 1565–1582 (2003)
28. Triebel, H.: Theory of Function Spaces. Birkhäuser, Basel (1983)
29. Triebel, H.: Theory of Function Spaces II. Birkhäuser, Basel (1992)
30. Triebel, H.: Interpolation Theory, Function Spaces, Differential Operators. Barth, Heidelberg (1995)
31. van de Geer, S.: The entropy bound for monotone functions. Technical Report 91-10, University of Leiden (1991)
32. van de Geer, S.: Hellinger-consistency of certain nonparametric maximum likelihood estimators. Ann. Stat. **21**, 14–44 (1993)
33. van de Geer, S.: Empirical Processes in M-Estimation. Cambridge University Press, Cambridge (2000)
34. van der Vaart, A.W.: Bracketing smooth functions. Stoch. Process. Appl. **52**, 93–105 (1994)
35. van der Vaart, A.W.: New Donsker classes. Ann. Probab. **24**, 2128–2140 (1996)
36. van der Vaart, A.W., Wellner, J.A.: Weak Convergence and Empirical Processes. Springer, New York (1996)
37. Wong, W.H., Shen, X.: Probability inequalities for likelihood ratios and convergence rates of sieve MLE's. Ann. Stat. **23**, 339–362 (1995)

38. Yukich, J.E.: Rates of convergence for classes of functions: the non-i.i.d. case. J. Multivar. Anal. **20**, 175–189 (1986)
39. Yukich, J.E.: Some limit theorems for the empirical process indexed by functions. Probab. Theory Relat. Fields **74** 71–90 (1987)
40. Ziemer, W.P.: Weakly Differentiable Functions. Springer, New York (1989)