Rev. Mat. Iberoamericana **19** (2003), 235–263

Harmonic Analysis of the space BV

Albert Cohen, Wolfgang Dahmen, Ingrid Daubechies and Ronald DeVore

Abstract

We establish new results on the space BV of functions with bounded variation. While it is well known that this space admits no unconditional basis, we show that it is "almost" characterized by wavelet expansions in the following sense: if a function f is in BV, its coefficient sequence in a BV normalized wavelet basis satisfies a class of weak-ℓ^1 type estimates. These weak estimates can be employed to prove many interesting results. We use them to identify the interpolation spaces between BV and Sobolev or Besov spaces, and to derive new Gagliardo-Nirenberg-type inequalities.

1. Background and main results

Many classical function spaces —such as the Sobolev, Hölder or Besov spaces— can be characterized by harmonic analysis methods through Fourier or wavelet bases, frames, Littlewood-Paley decompositions, approximation by spline functions, etc. Such characterizations are classically useful in various contexts such as operator theory or the theoretical and numerical analysis of PDEs.

More recently, several results in data compression and statistical estimation have shown that optimal algorithms for such applications can be derived from expansions into *unconditional bases* for the function space that models the object to be compressed or estimated (see [11] and [8]). By definition a sequence $(e_n)_{n\geq 0}$ in a Banach space X is an unconditional basis if and only if

(i) It is a Schauder basis, i.e., for every $x \in X$ there exists a unique sequence $(x_n)_{n\geq 0}$ such that $\lim_{N\to+\infty} \|x - \sum_{n=0}^N x_n e_n\|_X = 0$.

(ii) There exists a constant C such that for all finite sequences $(x_n)_{n=0}^N$ and $(y_n)_{n=0}^N$ such that $|y_n| \leq |x_n|$, one has $\| \sum_n y_n e_n\|_X \leq C \| \sum_n x_n e_n\|_X$.

2000 Mathematics Subject Classification: 42C40, 46B70,26B35 , 42B25.
Keywords: Bounded variation, wavelet decompositions, weak ℓ_1, K-functionals, interpolation, Gagliardo-Nirenberg inequalities, Besov spaces.

In other words, the space X can be characterized by the *size properties* of the coefficients describing its elements in terms of the basis $(e_n)_{n\geq 0}$. This means that numerical operations, such as thresholding, attenuating, or rounding-off the coefficients, are stable in the X norm. The early development of wavelet bases (see [13]) is closely related to the question of existence of an unconditional basis for the Hardy space H_1. It is now well established that wavelet bases are unconditional bases for most classical function spaces that are known to possess one. On the other hand, certain spaces such as L_1, C_0, $W^1(L_1)$ and BV are known to possess no unconditional basis of any type.

The space BV, consisting of functions with bounded variation, is of particular interest for applications to data compression and statistical estimation. It is often chosen as a model for piecewise smooth signals such as geometric images. Recall that, if Ω is an open set of $\mathbb{R}^d$, a function $f \in L_1(\Omega)$ has bounded variation if and only if its distributional gradient ∇f is a finite measure, i.e., if its *total variation*

$$|f|_{\mathrm{BV}(\Omega)} := \sup\Big\{ \int_\Omega f \ \mathrm{div}(g) \ ; \ g \in C_c^1(\Omega, \mathbb{R}^d), \|g\|_\infty \leq 1 \Big\}, \tag{1.1}$$

is finite. Here, for $g = (g_1, \dots, g_d)$,

$$\|g\|_\infty := \Big\| \Big(\sum_{i=1}^d g_i^2 \Big)^{1/2} \Big\|_{L_\infty(\Omega)}.$$

The space of such functions is denoted as $\mathrm{BV} = \mathrm{BV}(\Omega)$. It is a Banach space when equipped with the norm

$$\|f\|_{\mathrm{BV}(\Omega)} := \|f\|_{L_1(\Omega)} + |f|_{\mathrm{BV}(\Omega)}. \tag{1.2}$$

If a function $f \in \mathrm{BV}(\Omega)$ is in the smaller Sobolev space $W^1(L_1(\Omega))$, we can apply integration by parts in (1.1) and obtain that

$$|f|_{\mathrm{BV}(\Omega)} := \int_\Omega |\nabla f|. \tag{1.3}$$

It was recently shown ([4]) that, although BV does not possess an unconditional basis, it is "almost" characterized by wavelet decompositions in terms of *weak-type conditions* imposed on wavelet coefficients. Using this information about BV, it is possible to derive optimal compression or estimation algorithms based on wavelet thresholding.

In order to describe this result, as well as the results of the present paper, we briefly discuss wavelet bases. We shall confine our discussion to the d-dimensional wavelet bases that are derived from a tensor product multiresolution analysis (see [7] or [13] for a detailed treatment) although this is not essential.

Consider first the case of orthogonal wavelet bases. Let $\psi^0 = \varphi$ be a univariate, compactly supported scaling function associated with the compactly supported, orthogonal univariate wavelet $\psi^1 = \psi$. Let $E' := \{0,1\}^d$ be the vertices of the unit cube and E denote the set of nonzero vertices. For each $e \in E'$, we define

$$\psi^e(x) = \psi^{e_1}(x_1) \cdots \psi^{e_d}(x_d). \tag{1.4}$$

Let $\mathcal{D}$ denote the set of dyadic cubes in $\mathbb{R}^d$ and let $\mathcal{D}_j$ denote those dyadic cubes that have side length 2^{-j}. For any dyadic cube $I = 2^{-j}(k + [0,1]^d)$ in $\mathcal{D}_j$, and any $e \in E'$, we define the wavelet

$$\psi_I^e(x) := 2^{j(d-1)}\psi^e(2^j x - k), \tag{1.5}$$

which is a wavelet scaled relative to I. Note that we have normalized the wavelets ψ_I^e in $\mathrm{BV}(\mathbb{R}^d)$ and not, as is more customary, in $L_2(\mathbb{R}^d)$. (For $d = 2$, the two normalizations coincide.) It follows that

$$C_1 \le |\psi_I^e|_{\mathrm{BV}} \le C_2, \tag{1.6}$$

where the constants C_1 and C_2 depend only on the $\mathrm{BV}(\mathbb{R})$ norms of the univariate functions φ and ψ. Note also that we can replace the seminorm $|\cdot|_{\mathrm{BV}}$ in (1.6) by the norm $\|\cdot\|_{\mathrm{BV}}$ as long as the size of the cubes I remains bounded. The functions

$$\psi_I^e, \quad I \in \mathcal{D},\ e \in E, \tag{1.7}$$

form a complete orthogonal system in $L_2(\mathbb{R}^d)$.

There is a similar construction of biorthogonal wavelet bases, see e.g. [7]. We start with a pair of one-dimensional compactly supported scaling functions $\psi^0 := \varphi$ and $\tilde{\psi}^0 = \tilde{\varphi}$ which are in duality:

$$\int_{\mathbb{R}} \varphi(t)\tilde{\varphi}(t-k)\, dt = \delta(k), \quad k \in \mathbb{Z}, \tag{1.8}$$

with δ the Kronecker delta, and their corresponding univariate wavelets $\psi^1 := \psi$ and $\tilde{\psi}^1 := \tilde{\psi}$. We define the functions ψ_I^e as in (1.5) and $\tilde{\psi}_I^e$ similarly except that the factor 2^j is used in place of $2^{j(d-1)}$. The collection of functions $\{\psi_I^e\}_{I\in\mathcal{D}, e\in E}$ (when renormalized so that (1.6) holds with $|\cdot|_{\mathrm{BV}}$ replaced by the L_2 norm) forms a Riesz basis for $L_2(\mathbb{R}^d)$ and (a correspondingly renormalized version of) $\{\tilde{\psi}_I^e\}_{I\in\mathcal{D}, e\in E}$ is its dual basis. The orthogonal wavelet bases given above are special cases. Even in the orthogonal case we shall keep the notation $\tilde{\psi}_I^e$ to mark the difference in normalization.

Given a tempered distribution f on $\mathbb{R}^d$, we define its wavelet coefficients by

$$f_I^e := \langle f, \tilde{\psi}_I^e \rangle \tag{1.9}$$

whenever this inner product is defined (for example, if $\tilde{\psi}_I^e$ is in C^r this will be the case for all tempered distributions of order $< r$). The wavelet decomposition of f is then formally defined as

$$f = \sum_{e \in E} \sum_{I \in \mathcal{D}} f_I^e \psi_I^e, . \tag{1.10}$$

We can simplify notation by introducing the vectors $\psi_I = (\psi_I^e)_{e \in E}$ and $f_I = (f_I^e)_{e \in E}^T$ so that we have

$$f = \sum_{I \in \mathcal{D}} f_I \psi_I. \tag{1.11}$$

We shall also consider the "non-homogeneous version" of this wavelet basis, which is obtained by taking only the scales $j \geq 0$ and by including the index $e = (0, \ldots, 0)$ when $j = 0$, i.e., a coarse "layer" of scaling functions. Denoting by $\mathcal{D}_+ := \cup_{j \geq 0} \mathcal{D}_j$ the set of dyadic cubes with scale $j \geq 0$, we write this decomposition as

$$f = \sum_{I \in \mathcal{D}_+} F_I \Psi_I, \tag{1.12}$$

where the F_I and Ψ_I coincide with f_I and ψ_I if $I \in \mathcal{D}_j$, $j > 0$, while we incorporate the index $e = (0, \ldots, 0)$ when $j = 0$. Regardless of which wavelet basis we choose, the subscript I represents the spatial localization of the wavelets ψ_I and $\tilde{\psi}_I$ (I is contained in their support), and its volume $|I| = 2^{-jd}$ indicates their scale (the size of their support is proportional with $|I|$, with a proportionality constant independent of the scale). Note that for the Haar system, i.e., when $\varphi = \tilde{\varphi} = \chi_{]0,1[}$ and $\psi = \tilde{\psi} = \chi_{]0,1/2[} - \chi_{]1/2,1[}$, the supports of ψ_I and $\tilde{\psi}_I$ coincide exactly with I.

We can now formulate the following result which was first proved in the case of the Haar system [4] and later extended to more general compactly supported wavelets [5]. In this theorem, and later, we use $|\cdot|$ to denote the Euclidean norm in $\mathbb{R}^l$. A pivotal role is played by the space $w\ell^1(\mathcal{D})$ (weak ℓ_1). It consists of those sequences $(a_I)_{I \in \mathcal{D}}$ for which

$$\|(a_I)\|_{w\ell^1} := \sup_{\epsilon > 0} [\epsilon \ \#\{I \in \mathcal{D} : \ |a_I| > \epsilon\}] \tag{1.13}$$

is finite.

Theorem 1.1 *For all $f \in \mathrm{BV}(\mathbb{R}^d)$, the coefficient sequence $(f_I)_{I\in\mathcal{D}}$ belongs to the space $w\ell^1(\mathcal{D})$. More precisely, there exists a constant $C > 0$ such that for all $f \in \mathrm{BV}(\mathbb{R}^d)$ and $\epsilon > 0$*

$$\#\{I \in \mathcal{D} : |f_I| > \epsilon\} \le C|f|_{\mathrm{BV}(\mathbb{R}^d)}\epsilon^{-1}. \tag{1.14}$$

Similarly, for the non-homogeneous basis indexed by $\mathcal{D}_+$, we have

$$\#\{I \in \mathcal{D}_+ : |F_I| > \epsilon\} \le C\|f\|_{\mathrm{BV}(\mathbb{R}^d)}\epsilon^{-1}. \tag{1.15}$$

On the other hand, from the BV normalization of the wavelets (see (1.6)), we see that whenever $(F_I)_{I\in\mathcal{D}_+} \in \ell_1(\mathcal{D}_+)$ then $f := \sum_{I\in\mathcal{D}_+} F_I\Psi_I$ belongs to BV and satisfies

$$\|f\|_{\mathrm{BV}(\mathbb{R}^d)} \le C\|(F_I)\|_{\ell_1(\mathcal{D}_+)}. \tag{1.16}$$

Therefore, we have almost characterized $\mathrm{BV}(\mathbb{R}^d)$ in the following sense. Let $\mathrm{bv}(\mathcal{D}_+)$ denote the discrete space of wavelet coefficient sequences of BV functions with

$$\|(F_I)_{I\in\mathcal{D}_+}\|_{\mathrm{bv}} := \|f\|_{\mathrm{BV}(\mathbb{R}^d)}. \tag{1.17}$$

Then, we have the continuous embeddings

$$\ell_1(\mathcal{D}_+) \subset \mathrm{bv}(\mathcal{D}_+) \subset w\ell_1(\mathcal{D}_+). \tag{1.18}$$

This result is sufficient to ensure the optimality of estimation and compression algorithms in the sense of [11] (see [4]).

Theorem 1.1 also gives a direct easy access to some fine analysis results, such as the following improved Poincaré inequality in dimension $d = 2$:

$$\|f\|^2_{L_2(\mathbb{R}^2)} \le C\|f\|_{B^{-1}_\infty(L_\infty(\mathbb{R}^2))}\|f\|_{\mathrm{BV}(\mathbb{R}^2)}, \tag{1.19}$$

where $B^{-1}_\infty(L_\infty(\mathbb{R}^2))$ is the Besov space. The classical Poincaré inequality in this case would involve the L_∞ norm instead of the Besov norm on the right side of (1.19). The importance of (1.19) is that it scales correctly for both dilation and modulation (i.e. multiplication by a character $e^{i\omega\cdot x}$) whereas the original Poincaré inequality scales correctly only for dilation. In this sense, one could say that (1.19) is the "correct" Poincaré inequality.

With Theorem 1.1 in hand, inequality (1.19) can be derived from two facts indicating a pattern of argument that will be encountered later again. First one observes the inequality

$$\|(F_I)\|^2_{\ell_2} \le \|(F_I)\|_{\ell_\infty}\|(F_I)\|_{w\ell_1}. \tag{1.20}$$

The second ingredient is that the L_2 and $B_\infty^{-1}(L_\infty(\mathbb{R}^2))$ norms of a function f are respectively equivalent to the ℓ_2 and ℓ_∞ norm of the sequence $(F_I)_{I\in\mathcal{D}_+}$. These are special cases of norm equivalences that will be described later in more detail. The proof of (1.19) also uses that for $d = 2$ the BV- and L_2-normalizations of the wavelets coincide. Note that there exists no other proof of (1.19) up to now.

The inequalities (1.19) and (1.20) can be viewed as special cases of interpolation using the real method of Lions-Peetre (see e.g. [1] for an introduction). Given a pair of linear spaces (X, Y) continuously embedded in some Hausdorff space $\mathcal{X}$, the K-functional for this pair is given by

$$K(f, t; X, Y) := \inf_{g\in Y, f-g\in X} \|f - g\|_X + t\|g\|_Y, \quad t > 0, \tag{1.21}$$

where $\|\cdot\|_X$ and $\|\cdot\|_Y$ are quasi-semi-norms for these spaces. For each $0 < \theta < 1$, $0 < q \le \infty$, the intermediate space $(X, Y)_{\theta,q}$ consists of all elements of $X + Y$ for which

$$\|f\|_{(X,Y)_{\theta,q}} := \Big(\int_0^\infty [t^{-\theta}K(f,t)]^q \frac{dt}{t}\Big)^{1/q}, \quad 0 < q < \infty, \tag{1.22}$$

is finite (with the usual change to a sup when $q = \infty$). The space $(X, Y)_{\theta,q}$ is called an interpolation space for this pair. It is an important question in analysis to characterize the interpolation spaces for a given pair (X, Y). Such characterizations are known for many (but not all) pairs of classical spaces.

In particular, the intermediate spaces for any pair (ℓ_p, ℓ_q) of sequence spaces are known to be Lorentz spaces. Also, the same conclusion holds if the spaces ℓ_p and ℓ_q are replaced by their weak counterparts. As a special case, the "framing" of bv between ℓ_1 and $w\ell_1$ gives

$$\ell_2 = (\ell_\infty, \ell_1)_{1/2,2} \subset (\ell_\infty, \mathrm{bv})_{1/2,2} \subset (\ell_\infty, w\ell_1)_{1/2,2} = \ell_2. \tag{1.23}$$

From this and the characterization of L_2 and the Besov space by wavelet coefficients, we derive

$$L_2 = (B^{-1}_{\infty,\infty}, \mathrm{BV})_{1/2,2}. \tag{1.24}$$

This method of determining interpolation spaces for a pair of smoothness spaces by identifying them with sequence spaces, via a boundedly invertible linear mapping, is called the method of retracts. In the case of (1.24), this result (given in [4]) was new.

Given any pair (X, Y), one always has the interpolation inequality

$$\|f\|_{(X,Y)_{\theta,q}} \leq \|f\|_X^{1-\theta} \|f\|_Y^{\theta} \tag{1.25}$$

(see [1], p. 49). Thus, given (1.24), (1.19) and (1.20) in turn follow from this general principle (although (1.20) can be proved directly in a simple way as well).

Despite the above success, Theorem 1.1 is not sufficient to answer other fine questions in analysis. In fact, the present paper was motivated by questions raised by Yves Meyer concerning the correct form of Gagliardo-Nirenberg-type inequalities. Improving these inequalities in a similar way to (1.19) is equivalent to establishing new results on interpolation between BV and other Sobolev and Besov spaces. The difficulty in accomplishing this is that general Sobolev and Besov spaces are described by applying *weighted* $\ell_p(w)$ norms to wavelet coefficient sequences. The weights w take the form $|I|^s$ where we denote as before by $|I| := \mathrm{vol}(I)$ the volume of I. Theorem 1.1 is no longer tailored to this context, since the interpolation spaces between such a weighted ℓ_p space and $w\ell_1$ no longer yields the desired sequence space.

Fortunately, there is a possible way around this which was first utilized in [10]. The key is to incorporate weights both in renormalizing the coefficients and in the weak ℓ_1 space. To describe this, we introduce the following sequence spaces.

Definition 1.2 *Let $\gamma \in \mathbb{R}$. For $0 < p < \infty$, the space $\ell_p^\gamma(\mathcal{D})$ consists of those sequences $(c_I)_{I\in\mathcal{D}}$ such that $(|I|^{-\gamma} c_I)_{I\in\mathcal{D}} \in \ell_p(\mathcal{D}, |I|^\gamma)$, i.e.,*

$$\|(c_I)_{I\in\mathcal{D}}\|^p_{\ell_p^\gamma(\mathcal{D})} := \Big(\sum_{I\in\mathcal{D}} |I|^{(1-p)\gamma} |c_I|^p\Big)^{1/p} < \infty. \tag{1.26}$$

The space $w\ell_p^\gamma(\mathcal{D})$ consists of those sequences $(c_I)_{I\in\mathcal{D}}$ such that $(|I|^{-\gamma} c_I)_{I\in\mathcal{D}} \in w\ell_p(\mathcal{D}, |I|^\gamma)$, i.e.,

$$\|(c_I)\|_{w\ell_p^\gamma(\mathcal{D})} := \sup_{\epsilon>0} \epsilon^p \sum_{|c_I|>\epsilon|I|^\gamma} |I|^\gamma < \infty. \tag{1.27}$$

For $p = \infty$, the space $\ell_\infty^\gamma(\mathcal{D})$ consists of those sequences $(c_I)_{I\in\mathcal{D}}$ such that $(|I|^{-\gamma} c_I)_{I\in\mathcal{D}} \in \ell_\infty(\mathcal{D})$, i.e. $|c_I| \leq C|I|^\gamma$. The spaces $\ell_p^\gamma(\mathcal{D}_+)$ and $w\ell_p^\gamma(\mathcal{D}_+)$ are defined analogously.

Note that when $\gamma = 0$ this corresponds to the classical ℓ_p and $w\ell_p$ spaces. Note also that ℓ_1^γ coincides with ℓ_1 for all γ, while $w\ell_1^\gamma$ differs from $w\ell_1$. In fact there is no natural ordering of the spaces $w\ell_1^\gamma$ as γ varies.

Introducing the spaces $w\ell_p^\gamma$ helps us answer some questions concerning interpolation of smoothness spaces. In the present context of BV, it is easy to reduce the questions of Meyer to the following:

For which values of γ do we have the embedding of bv *into $w\ell_1^\gamma(\mathcal{D})$ or equivalently the weak-type estimate*

$$\sum_{|f_I|>\epsilon|I|^\gamma} |I|^\gamma \le C|f|_{\mathrm{BV}(\mathbb{R}^d)}\epsilon^{-1}\ ? \tag{1.28}$$

The main result of this paper is to give a precise answer to this question in the following theorem.

Theorem 1.3 *Inequality* (1.28) *holds if and only if $\gamma > 1$ or $\gamma < 1 - 1/d$. The same conclusion holds if in* (1.28) *we replace $(f_I)_{I\in\mathcal{D}}$ by $(F_I)_{I\in\mathcal{D}_+}$ and $|f|_{\mathrm{BV}(\mathbb{R}^d)}$ by $\|f\|_{\mathrm{BV}(\mathbb{R}^d)}$.*

Although Theorem 1.3 includes Theorem 1.1 as a particular case ($\gamma = 0$), the spirit of our proof is quite different from [4].

The proof of Theorem 1.3 is given in the following sections. We use the remainder of the present section to formulate and prove applications of Theorem 1.3 to interpolation and Gagliardo-Nirenberg-type inequalities. We first discuss interpolation between BV and the classical Besov-Sobolev spaces.

The Besov spaces $B_p^s(L_p(\mathbb{R}^d))$ are typically defined using Littlewood-Paley decompositions or moduli of smoothness. However, they have an equivalent formulation in terms of wavelet decompositions (see [13] or [2]) that we shall use here for their definition. Let the univariate scaling function φ and its associated wavelet ψ be in C^r and similarly let $\tilde\varphi$ and $\tilde\psi$ be in $C^{\tilde r}$. Then, for each $-\tilde r < s < r$, we define the Besov space $B_p^s(L_p(\mathbb{R}^d))$, $1 < p \le \infty$, as the set of all tempered distributions f such that

$$\|f\|_{B_p^s(L_p(\mathbb{R}^d))} := \|(|F_I|)\|_{\ell_p^\gamma(\mathcal{D}_+)}, \quad \gamma := 1 + (s-1)p^*/d \tag{1.29}$$

is finite, where p^* denotes the conjugate index to p.

This definition is in agreement with the characterization of Besov spaces by wavelet decomposition but it looks a little strange because we have used the ℓ_p^γ norms. The usual definition uses the L_p normalized wavelets

$$\Psi_{I,p} = |I|^{1/p^*-1/d}\ \Psi_I$$

and their corresponding coefficients $F_{I,p}$. Then it takes the form

$$\|f\|_{B_p^s(L_p(\mathbb{R}^d))} = \|(|I|^{-s/d}|F_{I,p}|)\|_{\ell_p(\mathcal{D}_+)} \tag{1.30}$$

which is identical with (1.29) because $F_{I,p} = |I|^{1/d-1/p^*}\, F_I$.

There is a similar wavelet description of the homogeneous Besov spaces $\dot{B}^s_p(L_p(\mathbb{R}^d))$ which were originally defined using Littlewood-Paley decompositions (see [13]). One can define the space $B^s_p(L_p(\mathbb{R}^d))$ as the set of all tempered distributions f such that

$$\|f\|_{\dot{B}^s_p(L_p(\mathbb{R}^d))} := \|(|f_I|)\|_{\ell^\gamma_p(\mathcal{D})}, \quad \gamma := 1 + (s-1)p^*/d \tag{1.31}$$

is finite.

We shall use the well known fact that for any $\gamma \in \mathbb{R}$ and $1 < p \le \infty$, we have (see e.g. Theorem 5.3.1, p. 113 in [1])

$$\ell^\gamma_q = (\ell^\gamma_p, \ell^\gamma_1)_{\theta,q} = (\ell^\gamma_p, w\ell^\gamma_1)_{\theta,q}, \tag{1.32}$$

whenever $0 < \theta < 1$ and

$$\frac{1}{q} = \frac{1-\theta}{p} + \theta. \tag{1.33}$$

We should note that γ is the same for all the spaces in (1.32). Since the appropriate value of γ is fixed by the Besov space that we wish to pair with BV, we have no flexibility in its choice and therefore cannot just simply apply Theorem 1.1 which corresponds to $\gamma = 0$.

Clearly from (1.32) and (1.25) it follows that we

$$\|(a_I)\|_{\ell^\gamma_q} \le C\|(a_I)\|^{1-\theta}_{\ell^\gamma_p}\|(a_I)\|^\theta_{w\ell^\gamma_1}. \tag{1.34}$$

Theorem 1.4 *Assume that $\gamma > 1$ or $\gamma < 1 - 1/d$, and let (s,p) satisfy $(s-1)p^*/d = \gamma - 1$ for some $1 < p \le \infty$. Then, for any $0 < \theta < 1$, we have*

$$(B^s_p(L_p(\mathbb{R}^d)), \mathrm{BV}(\mathbb{R}^d))_{\theta,q} = B^t_q(L_q(\mathbb{R}^d)) \tag{1.35}$$

with equivalent norms and with

$$\frac{1}{q} = \frac{1-\theta}{p} + \theta, \quad t = (1-\theta)s + \theta. \tag{1.36}$$

Similarly, we have

$$(\dot{B}^s_p(L_p(\mathbb{R}^d)), \dot{\mathrm{BV}}(\mathbb{R}^d))_{\theta,q} = \dot{B}^t_q(L_q(\mathbb{R}^d)) \tag{1.37}$$

with the same restrictions on p, q, t. Here $\dot{\mathrm{BV}}$ means that in the definition of the K-functional (1.21) we use the seminorm $|\cdot|_{\mathrm{BV}}$ rather than $\|\cdot\|_{\mathrm{BV}}$.

Proof. Consider the wavelet transform which linearly maps f into $(F_I)_{I\in\mathcal{D}_+}$. In view of (1.29), it is an isometry between $B^s_p(L_p(\mathbb{R}^d))$ and $\ell^\gamma_p(\mathcal{D}_+)$. We also know from (1.18) that the image $\mathrm{bv}(\mathcal{D}_+)$ of $\mathrm{BV}(\mathbb{R}^d)$ is framed by $\ell^\gamma_1(\mathcal{D}_+)$ and $w\ell^\gamma_1(\mathcal{D}_+)$. Hence, using (1.32), we deduce that a distribution f is in $(B^s_p(L_p(\mathbb{R}^d)),\mathrm{BV}(\mathbb{R}^d))_{\theta,q}$ if and only if $(F_I)_{I\in\mathcal{D}_+}\in\ell^\gamma_q(\mathcal{D}_+)$ with equivalent norms.

Now observe that for q and t as in (1.36) one has $(s-1)p^*=(t-1)q^*$. Thus one also has $\gamma=1+(t-1)q^*/d$ and invoking the definition of Besov spaces (1.29), the proof is completed. In the homogeneous case, we use the mapping of f into $(f_I)_{I\in\mathcal{D}}$ to arrive at (1.37). ∎

Combining Theorem 1.4 with (1.25) we immediately obtain the following Gagliardo-Nirenberg-type inequalities.

Theorem 1.5 *Under the same assumptions and using the same notation as in Theorem 1.4 we have*

$$\|f\|_{B^t_q(L_q(\mathbb{R}^d))}\le C\|f\|^{1-\theta}_{B^s_p(L_p(\mathbb{R}^d))}\|f\|^\theta_{\mathrm{BV}(\mathbb{R}^d)}. \tag{1.38}$$

and

$$\|f\|_{\dot B^t_q(L_q(\mathbb{R}^d))}\le C\|f\|^{1-\theta}_{\dot B^s_p(L_p(\mathbb{R}^d))}|f|^\theta_{\mathrm{BV}(\mathbb{R}^d)}. \tag{1.39}$$

In particular, for the Sobolev space $H^s(\mathbb{R}^d)=W^s(L_2(\mathbb{R}^d))$, we have

$$\|f\|^2_{H^s(\mathbb{R}^d)}\le C\|f\|_{B^{2s-1}_\infty(L_\infty(\mathbb{R}^d))}\|f\|_{\mathrm{BV}(\mathbb{R}^d)}, \tag{1.40}$$

provided $s<1/2$ or $s>1$. For $s=0$, this establishes (1.19) in any dimension.

The remainder of this paper will be devoted to the proof of Theorem 1.3. We begin in Section 2 by gathering some known results about functions in BV. In Section 3, we first deal with the case $\gamma<0$ or $\gamma>1$. The case $0\le\gamma<1-\frac{1}{d}$, which is by far more difficult, is handled in Section 4. Section 5 links this technical result with wavelet expansions and completes the proof of Theorem 1.3. We conclude this section with indicating some implications concerning restricted nonlinear approximation. Finally, Section 6 illustrates that the restriction on γ is sharp by providing counter-examples for $1-\frac{1}{d}\le\gamma\le 1$.

In all our arguments, and unless stated otherwise, C denotes a generic constant, the value of which may vary even within the same proof.

2. Some properties of BV functions

For a detailed treatment of BV functions including the proofs of the following fundamental results, we refer the reader to [14] or [12].

Although we shall not use it in the sequel, we first recall the alternate (and equivalent) definition of BV by finite differences: if Ω is an open set of $\mathbb{R}^d$, $f \in L_1(\Omega)$ has bounded variation if and only if the quantity

$$\sup_{|h|\le 1} \frac{\|f - f(\cdot + h)\|_{L_1(\Omega_h)}}{|h|}, \tag{2.1}$$

is finite where $\Omega_h := \{x \in \Omega : x + th \in \Omega \text{ for } t \in [0,1]\}$. Moreover for a fixed Ω, this quantity is equivalent to the total variation $|f|_{\mathrm{BV}(\Omega)}$. We also recall that the space $\mathrm{BV}(\Omega)$ is (non-compactly) embedded in $L_{d^*}(\Omega)$ with $d^* = \frac{d}{d-1}$ and that we have the embedding inequality (see [14], p. 81)

$$\|f\|_{L_{d^*}(\Omega)} \le C(\Omega)\|f\|_{\mathrm{BV}(\Omega)}. \tag{2.2}$$

We shall use the possibility of approximating the functions of $\mathrm{BV}(\Omega)$ by smooth functions in the following sense (see e.g. [12], p. 172 or [14], p. 225).

Theorem 2.1 *Let $f \in \mathrm{BV}(\Omega)$. Then there exists a sequence $\{f_k\}_{k\ge 0}$ in $\mathrm{BV}(\Omega) \cap C^\infty(\Omega)$ such that*

$$\lim_{k\to+\infty} \|f - f_k\|_{L_1(\Omega)} = 0 \quad \text{and} \quad \lim_{k\to+\infty} |f_k|_{\mathrm{BV}(\Omega)} = |f|_{\mathrm{BV}(\Omega)}. \tag{2.3}$$

This result will allow us to reduce the proof of our weak-type estimates to smooth functions for which we have $|f|_{\mathrm{BV}(\Omega)} = \int_\Omega |\nabla f|$.

Characteristic functions of sets are particular instances of BV functions which will play an important role in our analysis. If E is a bounded open set with smooth boundary, then it is easy to check from the definition that $\chi_E \in \mathrm{BV}(\Omega)$ and that

$$|\chi_E|_{\mathrm{BV}(\Omega)} = \mathcal{H}^{d-1}(\partial E \cap \Omega), \tag{2.4}$$

where here and later $\mathcal{H}^s$ denotes the s-dimensional Hausdorff measure. The above equality is not true for more general open sets with finite perimeter but no Lipschitz boundary (take e.g. $E := \{(x,y)\ |x| < 1, 0 < |y| < 1\}$), for which we have only the inequality

$$|\chi_E|_{\mathrm{BV}(\Omega)} \le \mathcal{H}^{d-1}(\partial E \cap \Omega). \tag{2.5}$$

The importance of characteristic functions in the description of BV is emphasized by the *co-area formula* which has the following classical form for sufficiently smooth functions. If $f \in \mathrm{BV}(\Omega) \cap C^1(\Omega)$, then one has

$$\int_\Omega |\nabla f| = \int_{\mathbb{R}} \mathcal{H}^{d-1}(\Omega \cap f^{-1}(\{t\}))dt, \tag{2.6}$$

(see [12], p. 112 or [14], p. 76). To extend this relation to general BV functions one introduces the level sets $E_t = E_t(f)$ defined by $E_t = \{x \in \Omega : f(x) \geq t\}$. The above formula then takes the following form (see [12], p. 185):

Theorem 2.2 *For $f \in \mathrm{BV}(\Omega)$ one has*

$$|f|_{\mathrm{BV}(\Omega)} = \int_{\mathbb{R}} |\chi_{E_t}|_{\mathrm{BV}(\Omega)} dt. \tag{2.7}$$

Such level sets might not have a C^1 boundary for almost every t and therefore one cannot substitute $\mathcal{H}^{d-1}(\partial E_t \cap \Omega)$ in place of $|\chi_{E_t}|_{\mathrm{BV}}$. The co-area formula (2.7) reveals that BV admits an *atomic decomposition* in terms of characteristic functions since we have

$$f(x) = \lim_{z \to -\infty} z + \int_z^{+\infty} \chi_{E_t}(x) dt. \tag{2.8}$$

Such a decomposition can be particularly useful when proving properties of the type $\Phi(f) \leq C|f|_{\mathrm{BV}}$ where Φ is a convex functional, since it reduces the proof to the case where f is a single atom χ_E.

We shall also need a version of the isoperimetric inequality which we prove here by applying the embedding of BV into L^{d^*} to characteristic functions.

Theorem 2.3 *Let Q be an open cube of $\mathbb{R}^d$ and let E be a domain with a smooth boundary. Define $E_Q := E \cap Q$ and its complement $\tilde{E}_Q := Q \setminus E_Q$. Then there exists a constant C independent of E and of Q such that*

$$\min\{|E_Q|, |\tilde{E}_Q|\} \leq C[\mathcal{H}^{d-1}(\partial E \cap Q)]^{d^*}. \tag{2.9}$$

Proof. Let E_Q^* denote the set of minimal measure among E_Q and $\tilde{E}_Q$ and define

$$a_Q(f) := |Q|^{-1} \int_Q f,$$

We clearly have $|\chi_E - a_Q(\chi_E)| \geq 1/2$ on E_Q^* and therefore

$$\int_Q |\chi_E - a_Q(\chi_E)|^{d^*} \geq 2^{-d^*} \min\{|E_Q|, |\tilde{E}_Q|\}. \tag{2.10}$$

In view of the formula (2.4), (2.9) follows as soon as we can estimate the left hand side of (2.10) by $|\chi_E|^{d^*}_{\mathrm{BV}(\Omega)}$. This in turn is a consequence of the following Poincaré inequality for general BV functions

$$\|f - a_Q(f)\|_{L_{d^*}(Q)} \leq C|f|_{\mathrm{BV}(Q)}. \tag{2.11}$$

This could be derived directly but we will use here an argument that will be needed later anyway. To this end, note that

$$\begin{aligned} \|f - a_Q(f)\|_{L_1(Q)} &= |Q|^{-1} \int_Q \left| \int_Q (f(x) - f(y))dy \right| dx \\ &\leq |Q|^{-1} \int_{Q\times Q} |f(x) - f(y)|dxdy, \end{aligned} \tag{2.12}$$

and assume first that $f \in W^1(L_1(\Omega))$. For each $x = (x_1, \dots, x_d)$ and $y = (y_1, \dots, y_d)$ we can define the segments

$$S_i(x, y) := Q \cap \{(x_1, \dots, x_{i-1}, t, y_{i+1}, \dots, y_d)\ ;\ t \in \mathbb{R}\},\quad i = 1, \dots, d.$$

With such a definition, we can connect x and y by a path $S(x, y) \subset \cup_i S_i(x, y)$ so that

$$|f(x) - f(y)| \leq \sum_{i=1}^{d} \int_{S_i(x,y)} \left|\frac{\partial f}{\partial x_i}\right|.$$

Integrating with respect to x and y we can estimate the right hand side of (2.12) by

$$|Q|^{-1} \sum_{i=1}^{d} \int_{Q\times Q} \int_{S_i(x,y)} \left|\frac{\partial f}{\partial x_i}\right| dx\, dy \leq |Q|^{1/d} \sum_{i=1}^{d} \int_Q \left|\frac{\partial f}{\partial x_i}\right| \leq \sqrt{d}|Q|^{1/d} \int_I |\nabla f|.$$

Now (1.3) and Theorem 2.1 imply that for any $f \in \mathrm{BV}(Q)$

$$\begin{aligned} \|f - a_Q(f)\|_{L_1(Q)} &\leq |Q|^{-1} \int_{Q\times Q} |f(x) - f(y)|dxdy \\ &\leq \sqrt{d}\ |Q|^{1/d}\, |f|_{\mathrm{BV}(Q)}. \end{aligned} \tag{2.13}$$

In particular, for the unit d-cube $\square$ the estimate $\|f - a_\square(f)\|_{\mathrm{BV}(\square)} \leq (1 + \sqrt{d})|f|_{\mathrm{BV}(\square)}$ follows. The embedding (2.2) of $\mathrm{BV}(\square)$ into $L_{d^*}(\square)$ yields $\|f - a_Q(f)\|_{L_{d^*}(\square)} \leq C(\square)\|f - a_Q(f)\|_{\mathrm{BV}(\square)}$ and thus $\|f - a_\square(f)\|_{L_{d^*}(\square)} \leq C(\square)|f|_{\mathrm{BV}(\square)}$. One easily checks that this latter estimate remains invariant under rescaling which confirms (2.11) and completes the proof. ∎

In the sequel we shall assume $\Omega = \mathbb{R}^d$, and the space BV will always refer to $\mathrm{BV}(\mathbb{R}^d)$. Let g be a function in L_∞ supported on $]0,1[^d$ and such that $\int g = 0$. For $I := 2^{-j}(]0,1[^d + k)$ a dyadic cube in $\mathcal{D}$, we define

$$g_I := 2^j g(2^j \cdot - k). \tag{2.14}$$

For $I \in \mathcal{D}$, we introduce four quantities which measure in some sense the oscillation of a function f on I. The first one is the size of the inner product with g_I, i.e.

$$c_I := c_I(f) := |\langle f, g_I \rangle|. \tag{2.15}$$

The second one is the renormalized error of approximation by constants

$$r_I := r_I(f) := |I|^{-1/d} \|f - a_I(f)\|_{L_1(I)}. \tag{2.16}$$

The third one is the renormalized averaged modulus of continuity

$$w_I := w_I(f) := |I|^{-1-1/d} \int_{I \times I} |f(x) - f(y)| dx\, dy. \tag{2.17}$$

The above three quantities are well defined whenever $f \in L_1(I)$. When $f \in W^1(L_1(I))$, we define the fourth one as the variation of f on I

$$v_I := v_I(f) := \int_I |\nabla f| \ ; \tag{2.18}$$

for general BV functions, we set $v_I(f) := |f|_{\mathrm{BV}(I)}$.

Lemma 2.4 *We have for all $f \in L_1(I)$*

$$c_I(f) \leq \|g\|_{L_\infty(\mathbb{R}^d)} r_I(f), \tag{2.19}$$

and

$$r_I(f) \leq w_I(f). \tag{2.20}$$

For all $f \in \mathrm{BV}(I)$ we have

$$w_I(f) \leq \sqrt{d}\, v_I(f). \tag{2.21}$$

Proof. Observing that g_I is orthogonal to constants, we obtain

$$\begin{aligned} c_I(f) &= |\langle f, g_I \rangle| = |\langle f - a_I(f), g_I \rangle| \\ &\leq \|f - a_I(f)\|_{L_1(I)} \|g_I\|_{L_\infty} = \|g\|_{L_\infty} |I|^{-1/d} \|f - a_I(f)\|_{L_1(I)}, \end{aligned}$$

which is (2.19). The second inequality (2.20) follows from (2.12). Finally (2.21) immediately follows from (2.13). ∎

This list of inequalities will be used to prove the following result.

Theorem 2.5 *Let $f \in \mathrm{BV}(\mathbb{R}^d)$. Then $(c_I(f))_{I\in\mathcal{D}} \in w\ell_1^\gamma(\mathcal{D})$ for all $\gamma < 1-1/d$ or $\gamma > 1$. More precisely, there exists a constant C depending only on γ such that for all $f \in BV(\mathbb{R}^d)$ and $\epsilon > 0$ we have*

$$\sum_{I\in\mathcal{D};\ c_I(f)>\epsilon|I|^\gamma} |I|^\gamma \le C|f|_{\mathrm{BV}(\mathbb{R}^d)}\epsilon^{-1}. \tag{2.22}$$

The proof of this result is the object of the next two sections. This theorem will then be used in §5 to prove Theorem 1.3. For the proof of Theorem 2.5, we shall restrict ourselves to $f \in \mathrm{BV}(\mathbb{R}^d) \cap C^\infty(\mathbb{R}^d)$. The result for a general $f \in \mathrm{BV}(\mathbb{R}^d)$ is then proved by using the approximation sequence $(f_k)_{k\ge0}$ of Theorem 2.1 and noting that $\langle f_k, g_I\rangle$ tends to $\langle f, g_I\rangle$ for all I. If follows that if (2.22) holds for all the f_k, then for each finite subset

$$\Lambda_\epsilon^* \subset \Lambda_\epsilon := \{I \in \mathcal{D} :\ |\langle f, g_I\rangle| > \epsilon|I|^\gamma\}, \tag{2.23}$$

we have the property

$$\sum_{I\in\Lambda_\epsilon^*} |I|^\gamma \le C|f_k|_{\mathrm{BV}(\mathbb{R}^d)}\epsilon^{-1}, \tag{2.24}$$

provided k is sufficiently large. Letting k go to infinity, we conclude that (2.22) also holds for f.

3. The case $\gamma < 0$ or $\gamma > 1$

We begin with the cases $\gamma < 0$ or $\gamma > 1$ which have simple proofs. In these cases, it is sufficient to use the estimate $c_I(f) \le Cv_I(f)$ of Lemma 2.4 with $C = \sqrt{d}\|g\|_{L^\infty}$.

Theorem 3.1 *Assume that $\gamma > 1$ or $\gamma < 0$. Then $f \in \mathrm{BV}(\mathbb{R}^d) \cap C^\infty(\mathbb{R}^d)$ implies that $(v_I(f)) \in w\ell_1^\gamma(\mathcal{D})$. More precisely, there exists $C = C(\gamma)$ such that for all such f and each $\epsilon > 0$,*

$$\sum_{v_I(f)>\epsilon|I|^\gamma} |I|^\gamma \le C\epsilon^{-1}\int_{\mathbb{R}^d} |\nabla f|. \tag{3.1}$$

Proof. For $f \in \mathrm{BV}(\mathbb{R}^d)\cap C^\infty(\mathbb{R}^d)$ and $\epsilon > 0$, we want to estimate $\sum_{I\in\Lambda_\epsilon} |I|^\gamma$ where

$$\Lambda_\epsilon := \Lambda_\epsilon(f) := \{I \in \mathcal{D} :\ v_I(f) > \epsilon|I|^\gamma\}. \tag{3.2}$$

We first treat the case $\gamma > 1$. We define $\Lambda_\epsilon^{\max}$ as the subset of *maximal cubes* of Λ_ϵ, i.e. those $I \in \Lambda_\epsilon$ such that for all $J \in \Lambda_\epsilon$, $I \subseteq J$ implies $I = J$.

Since $v_I(f) \le |f|_{\rm BV}$ and since $\gamma > 0$, there exists a constant $A > 0$ depending of f and ϵ such that for $|I| \ge A$, we always have $v_I(f) \le \epsilon |I|^\gamma$, i.e. $I \notin \Lambda_\epsilon$. It follows that any cube $J \in \Lambda_\epsilon$ is always contained in some maximal cube $I \in \Lambda_\epsilon^{\max}$. Consequently, we have the estimate

$$\begin{aligned}\sum_{I\in\Lambda_\epsilon} |I|^\gamma &\le \sum_{I\in\Lambda_\epsilon^{\max}} \sum_{J\subseteq I} |J|^\gamma = \sum_{I\in\Lambda_\epsilon^{\max}} \sum_{j\ge 0} \sum_{J\subseteq I, |J|=2^{-jd}|I|} |J|^\gamma \\ &= \sum_{I\in\Lambda_\epsilon^{\max}} |I|^\gamma \sum_{j\ge 0} 2^{(1-\gamma)dj} \le C \sum_{I\in\Lambda_\epsilon^{\max}} |I|^\gamma \le C\epsilon^{-1} \sum_{I\in\Lambda_\epsilon^{\max}} v_I(f).\end{aligned}$$

Since the maximal cubes of $\Lambda_\epsilon^{\max}$ are necessarily pairwise disjoint we conclude that $\sum_{I\in\Lambda_\epsilon^{\max}} v_I(f) \le \int_{\mathbb{R}^d} |\nabla f|$ which proves (3.1).

In the case $\gamma < 0$, we define $\Lambda_\epsilon^{\min}$ as the subset of *minimal cubes* of Λ_ϵ, i.e., those $I \in \Lambda_\epsilon$ such that for all $J \in \Lambda_\epsilon$, $J \subseteq I$ implies $I = J$. Since $v_I(f) = \int_I |\nabla f| \le \|\nabla f\|_{L_\infty(J)} |I|$ for all $I \subseteq J$, and since $\gamma < 0$, for any fixed dyadic cube J there exists $a > 0$ depending of f and ϵ such that if $I \subset J$ and $|I| \le a$, we have $v_I(f) \le \epsilon |I|^\gamma$, i.e., $I \notin \Lambda_\epsilon$. It follows that any $J \in \Lambda_\epsilon$ contains only a finite number of $I \in \Lambda_\epsilon$, and in turn always contains a minimal cube $I \in \Lambda_\epsilon^{\min}$. Using also the fact that each $I \in \Lambda_\epsilon^{\min}$ is contained in at most one dyadic cube $J \in \mathcal{D}_j$ for any j, we have the estimate

$$\begin{aligned}\sum_{I\in\Lambda_\epsilon} |I|^\gamma &\le \sum_{I\in\Lambda_\epsilon^{\min}} \sum_{J\supseteq I} |J|^\gamma = \sum_{I\in\Lambda_\epsilon^{\min}} |I|^\gamma \sum_{j\ge 0} 2^{\gamma dj} \\ &\le C \sum_{I\in\Lambda_\epsilon^{\min}} |I|^\gamma \le C\epsilon^{-1} \sum_{I\in\Lambda_\epsilon^{\min}} v_I(f).\end{aligned}$$

We conclude the proof in a similar manner as above, noting that the minimal cubes of $\Lambda_\epsilon^{\min}$ are necessarily pairwise disjoint. ■

In view of the remarks at the end of §2, Theorem 3.1 implies Theorem 2.5 in the cases $\gamma < 0$ or $\gamma > 1$.

4. The case $0 \le \gamma < 1 - 1/d$

In this case, the estimate $c_I(f) \le C v_I(f)$ is not sufficient to prove Theorem 2.5 because the sequence $(v_I(f))$ does not satisfy the weak-type estimate (3.1) when $0 \le \gamma \le 1$. For instance, take $\gamma = 0$ and consider a non trivial smooth function f with compact support in $]0,1[^d$; observe that there exists an infinite number of dyadic cubes I containing $]0,1[^d$ for which we have $v_I(f) = C > 0$.

Instead we shall use the finer estimate $c_I(f) \le C w_I(f)$, with $C = \|g\|_{L^\infty}$, combined with the following result.

Theorem 4.1 *Let $\gamma < 1 - 1/d$. Then $f \in \mathrm{BV}(\mathbb{R}^d) \cap C^\infty(\mathbb{R}^d)$ implies that $(w_I(f))_{I\in\mathcal{D}} \in w\ell_1^\gamma$. More precisely, there exists a constant $C = C(\gamma)$ such that for all such f and each $\epsilon > 0$,*

$$\sum_{w_I(f)>\epsilon|I|^\gamma} |I|^\gamma \le C\epsilon^{-1} \int_{\mathbb{R}^d} |\nabla f|. \tag{4.1}$$

The proof of this result will involve some intermediate lemmas. Define the set

$$\Lambda_\epsilon := \Lambda_\epsilon(f) := \{I \in \mathcal{D} : w_I(f) > \epsilon|I|^\gamma\}. \tag{4.2}$$

Our goal is to show that

$$\epsilon \sum_{I\in\Lambda_\epsilon} |I|^\gamma \le C|f|_{\mathrm{BV}(\mathbb{R}^d)}. \tag{4.3}$$

We first fix some α such that $\gamma < \alpha < 1 - 1/d$ and establish a distinction between two types of cubes in Λ_ϵ.

Definition 4.2 *We say a cube $I \in \Lambda_\epsilon$ is good if for each collection $\mathcal{P} \subset \Lambda_\epsilon$ of pairwise disjoint cubes strictly contained in I, we have*

$$\sum_{J\in\mathcal{P}} |J|^\alpha \le |I|^\alpha,$$

or if I is minimal in Λ_ϵ, i.e., there is no $J \in \Lambda_\epsilon$ strictly contained in I. If $I \in \Lambda_\epsilon$ is not good we say it is bad. We denote the set of good cubes in Λ_ϵ by $\mathcal{G}$ and the set of bad cubes by $\mathcal{B}$.

Clearly $\mathcal{G}$ and $\mathcal{B}$ depend on f, ϵ, γ and α. Our next lemma shows that it is sufficient to prove (4.3) with $\mathcal{G}$ in place of Λ_ϵ.

Lemma 4.3 *We have*

$$\sum_{I\in\mathcal{B}} |I|^\gamma \le C \sum_{I\in\mathcal{G}} |I|^\gamma, \tag{4.4}$$

where the constant $C > 0$ depends only on α and γ.

Proof. Since $w_J(f) \le C\|\nabla f\|_{L_\infty(I)}|J|$ for all $J \subseteq I$ and since $\gamma < 1$, there exists a constant $a > 0$ such that if $J \subset I$ and $|J| \le a$, we always have $w_J(f) \le \epsilon|J|^\gamma$, i.e., $J \notin \Lambda_\epsilon$. It follows that any $I \in \Lambda_\epsilon$ contains only a finite number of $J \in \Lambda_\epsilon$.

For $I \in \mathcal{B}$, we denote by $\mathcal{G}(I)$ the set of all cubes $J \subset I$ such that $J \in \mathcal{G}$. Clearly this set is also finite. We shall first prove that

$$|I|^\alpha \le \sum_{J \in \mathcal{G}(I)} |J|^\alpha. \tag{4.5}$$

From the definition of bad cubes, there is a set $\mathcal{P}(I) \subset \Lambda_\epsilon$ of disjoint cubes contained in I such that

$$|I|^\alpha \le \sum_{J \in \mathcal{P}(I)} |J|^\alpha \le \sum_{J \in \mathcal{P}(I) \cap \mathcal{G}} |J|^\alpha + \sum_{J \in \mathcal{P}(I) \cap \mathcal{B}} |J|^\alpha = \Sigma_1 + \Sigma_2.$$

The terms in Σ_1 are not processed further and become part of the right side of (4.5). The terms in Σ_2 are processed further. Namely, for each J appearing in $\sum_2$ there is a set $\mathcal{P}(J) \subset \Lambda_\epsilon$ such that

$$|J|^\alpha \le \sum_{K \in \mathcal{P}(J)} |K|^\alpha \le \sum_{K \in \mathcal{P}(J) \cap \mathcal{G}} |K|^\alpha + \sum_{K \in \mathcal{P}(J) \cap \mathcal{B}} |K|^\alpha.$$

Again, the terms in the first sum are not processed further and become part of the right side of (4.5), remarking that $\mathcal{P}(J) \cap \mathcal{G}$ is necessarily disjoint from $\mathcal{P}(I) \cap \mathcal{G}$. The terms in the second sum are processed further. Continuing in this way, we arrive at (4.5) in a finite number of steps since $\mathcal{G}(I)$ is finite, and since the minimal cubes of Λ_ϵ are by definition contained in $\mathcal{G}$.

It follows from (4.5) that if $I \in \mathcal{B}$, then

$$|I|^\gamma \le |I|^{\gamma - \alpha} \sum_{J \in \mathcal{G}(I)} |J|^\alpha = \sum_{J \in \mathcal{G}(I)} |J|^\gamma 2^{-\delta d(I,J)},$$

where $\delta := (\alpha - \gamma)d > 0$ and $d(I, J)$ is the number of levels between J and I, i.e.

$$d(I, J) := \frac{|\log(|I|/|J|)|}{d \log 2}. \tag{4.6}$$

Therefore, the left side of (4.4) does not exceed

$$\sum_{I \in \mathcal{B}} \sum_{J \in \mathcal{G}(I)} |J|^\gamma 2^{-\delta d(I,J)} \le \sum_{J \in \mathcal{G}} |J|^\gamma \sum_{I \in \mathcal{B}, I \supset J} 2^{-\delta d(I,J)}.$$

For $J \in \mathcal{G}$ and $k > 0$, there is at most one $I \supset J$ with $d(I, J) = k$, and therefore

$$\sum_{I \in \mathcal{B}, I \supset J} 2^{-\delta d(I,J)} \le \sum_{k=1}^{\infty} 2^{-k\delta} = C.$$

This proves the lemma. ∎

It follows from Lemma 4.3 that we need to estimate only $\sum_{I\in\mathcal{G}} |I|^\gamma$ in order to prove Theorem 4.1. We shall actually prove that the subsequence $(w_I(f))_{I\in\mathcal{G}}$ satisfies a *strong* ℓ_1 property. For this purpose, we introduce the following definition.

Definition 4.4 *A subset $\mathcal{R} \subset \mathcal{D}$ is called α-sparse if and only if for all $I \in \mathcal{R}$ and any set $\mathcal{P} \subset \mathcal{R}$ of disjoint dyadic cubes contained in I, we have*

$$\sum_{J\in\mathcal{P}} |J|^\alpha \leq |I|^\alpha. \tag{4.7}$$

Clearly $\mathcal{G}$ is an example of an α-sparse set.

With this definition we have the following theorem.

Theorem 4.5 *There exists a constant C such that for any α-sparse set $\mathcal{R}$, we have*

$$\sum_{I\in\mathcal{R}} w_I(f) \leq C|f|_{\mathrm{BV}(\mathbb{R}^d)}. \tag{4.8}$$

Proof. Since we want to prove a *strong* ℓ^1 estimate, we can use the co-area formula to reduce the proof to the case where f is of the type

$$f = \chi_E, \tag{4.9}$$

where E is a set with smooth boundary of finite $d-1$-dimensional measure $\mathcal{H}^{d-1}(\partial E)$. Indeed, assume for a moment that (4.8) holds for such characteristic functions. If $f \in \mathrm{BV}(\mathbb{R}^d) \cap C^\infty(\mathbb{R}^d)$, then, for almost every $t \in \mathbb{R}$, the level sets $E_t = \{x \in \Omega : \ f(x) \geq t\}$ have a smooth boundary and thus satisfy

$$\sum_{I\in\mathcal{R}} w_I(\chi_{E_t}) \leq C|\chi_{E_t}|_{\mathrm{BV}(\mathbb{R}^d)}. \tag{4.10}$$

For all I, (2.8) yields

$$w_I(f) \leq \int_{-\infty}^{+\infty} w_I(\chi_{E_t})dt. \tag{4.11}$$

Combining these inequalities with (2.7), we conclude that for all finite subsets $\Lambda \subset \mathcal{R}$ we have

$$\sum_{I\in\Lambda} w_I(f) \leq C|f|_{\mathrm{BV}(\mathbb{R}^d)}, \tag{4.12}$$

and therefore (4.8) also holds for f.

Also note that $|E_t| < \infty$ if $t > 0$ and $|\mathbb{R}^d \setminus E_t| < \infty$ if $t < 0$ (since $f \in L^1$). Thus, it suffices to establish (4.8) for $f = \chi_E$, where E is a set such that either $|E| < \infty$ or $|\mathbb{R}^d \setminus E| < \infty$, which we will assume for the rest of the proof. For $I \in \mathcal{R}$, we define by $E_I := I \cap E$ and its complement $\tilde{E}_I = I \setminus E_I$. From its definition, we see that w_I satisfies the estimate

$$w_I(f) \le |I|^{-1-1/d}|I| \min\{|E_I|, |\tilde{E}_I|\} = |I|^{1-1/d} \min\{|E_I|, |\tilde{E}_I|\}/|I|. \tag{4.13}$$

Clearly,

$$\sum_{I\in\mathcal{R}} w_I(f) = \sum_{I\in\mathcal{R}^*} w_I(f), \tag{4.14}$$

where $\mathcal{R}^* := \{I \in \mathcal{R} : \ \min\{|E_I|, |\tilde{E}_I|\} > 0\}$ is the set of cubes whose interior intersects the boundary ∂E. For each $k > 0$, denote by $\mathcal{R}_k$ the set of cubes in $\mathcal{R}^*$ such that

$$2^{-k-1} < \min\{|E_I|, |\tilde{E}_I|\}/|I| \le 2^{-k}. \tag{4.15}$$

The sets $\mathcal{R}_k$ are pairwise disjoint and $\mathcal{R}^* = \cup_{k>0}\mathcal{R}_k$. We thus have

$$\sum_{I\in\mathcal{R}} w_I(f) = \sum_{k>0} \sum_{I\in\mathcal{R}_k} w_I(f). \tag{4.16}$$

We denote by $\mathcal{R}_{k,0}$ the *maximal cubes* of $\mathcal{R}_k$, i.e., the set of those $I \in \mathcal{R}_k$ such that there exists no $J \in \mathcal{R}_k$ strictly containing I. Note that since $|E| < \infty$ or $|\mathbb{R}^d \setminus E| < \infty$, there exists $A > 0$ such that $|I| \ge A$ implies

$$\min\{|E_I|, |\tilde{E}_I|\}/|I| \le \min\{|E|, |\mathbb{R}^d \setminus E|\}/|I| \le 2^{-k-1},$$

i.e. $I \notin \mathcal{R}_k$. Therefore, any in $\mathcal{R}_k$ is contained in some maximal cube of $\mathcal{R}_{k,0}$. If I is in $\mathcal{R}_{k,0}$ and $J \in \mathcal{R}_k$ is contained in I, we define the *generation* of J as the number of different cubes $K \ne I$ in $\mathcal{R}_k$ such that $J \subseteq K \subset I$. In particular, all cubes in $\mathcal{R}_{k,0}$ have generation 0. We denote by $\mathcal{R}_{k,j}$ the collection of cubes in $\mathcal{R}_k$ of generation j. Note that the cubes in $\mathcal{R}_{k,j}$ are pairwise disjoint, and that $\mathcal{R}_k = \cup_{j\ge0}\mathcal{R}_{k,j}$, so that we have

$$\sum_{I\in\mathcal{R}} w_I(f) = \sum_{k>0} \sum_{j\ge0} \sum_{I\in\mathcal{R}_{k,j}} w_I(f). \tag{4.17}$$

For a fixed $k > 0$, we can write in view of (4.15) and (4.13)

$$\sum_{j\ge0} \sum_{I\in\mathcal{R}_{k,j}} w_I(f) \le 2^{-k} \sum_{j\ge0} \sum_{I\in\mathcal{R}_{k,j}} |I|^{1-1/d}. \tag{4.18}$$

We next define $\eta = 1 - 1/d - \alpha > 0$ and remark that for $j \geq 0$ each $J \in \mathcal{R}_{k,j+1}$ must have a parent in $\mathcal{R}_{k,j}$. From the definition of α-sparse sets and the fact that the $\mathcal{R}_{k,j}$, $j \geq 0$, are sets of disjoint cubes, we infer that

$$\begin{aligned}\sum_{I\in\mathcal{R}_{k,j-1}} |I|^{1-1/d} &= \sum_{I\in\mathcal{R}_{k,j-1}} |I|^{\eta}|I|^{\alpha} \geq \sum_{I\in\mathcal{R}_{k,j-1}} |I|^{\eta} \sum_{J\in\mathcal{R}_{k,j},J\subset I} |J|^{\alpha} \\ &= \sum_{I\in\mathcal{R}_{k,j-1}} \sum_{J\in\mathcal{R}_{k,j},J\subset I} (|I|/|J|)^{\eta}|J|^{1-1/d} \geq 2^{d\eta} \sum_{J\in\mathcal{R}_{k,j}} |J|^{1-1/d},\end{aligned}$$

where the first inequality used the definition (4.7) of α-sparse. Therefore, we obtain by induction that

$$\sum_{I\in\mathcal{R}_{k,j}} |I|^{1-1/d} \leq 2^{-d\eta j} \sum_{I\in\mathcal{R}_{k,0}} |I|^{1-1/d}, \tag{4.19}$$

so that the summation over j, for fixed k, can be bounded by

$$\sum_{j\geq 0} \sum_{I\in\mathcal{R}_{k,j}} w_I(f) \leq C2^{-k} \sum_{I\in\mathcal{R}_{k,0}} |I|^{1-1/d}. \tag{4.20}$$

Now if $I \in \mathcal{R}_{k,0}$, we see from (4.15) that $|I| \leq 2^{k+1}\min\{|E_I|, |\tilde{E}_I|\}$ and therefore

$$\sum_{I\in\mathcal{R}_{k,0}} |I|^{1-1/d} \leq 2^{(k+1)(1-1/d)} \sum_{I\in\mathcal{R}_{k,0}} \Big(\min\{|E_I|, |\tilde{E}_I|\}\Big)^{1-1/d}. \tag{4.21}$$

From the isoperimetric inequality of Theorem 2.3 we obtain

$$\Big(\min\{|E_I|, |\tilde{E}_I|\}\Big)^{1-1/d} \leq C\mathcal{H}^{d-1}(\partial E \cap I). \tag{4.22}$$

Since the maximal cubes of $\mathcal{R}_{k,0}$ are pairwise disjoint, it follows from (4.21) and (4.22) that

$$\sum_{I\in\mathcal{R}_{k,0}} |I|^{1-1/d} \leq C2^{k(1-1/d)}\mathcal{H}^{d-1}(\partial E). \tag{4.23}$$

Combining this estimate with (4.20) and (4.18), we obtain

$$\sum_{j\geq 0} \sum_{I\in\mathcal{R}_{k,j}} w_I(f) \leq C2^{-k/d}\mathcal{H}^{d-1}(\partial E).$$

Summing over $k > 0$ according to (4.17), we finally arrive at the estimate

$$\sum_{I\in\mathcal{R}} w_I(f) \leq C\mathcal{H}^{d-1}(\partial E) = C|\chi_E|_{\mathrm{BV}(\mathbb{R}^d)}, \tag{4.24}$$

where the last equality is (2.4). This concludes the proof of the theorem. ■

We now apply Theorem 4.5 to $\mathcal{R} = \mathcal{G}$. Since $\ell_1 = \ell_1^\gamma \subset w\ell_1^\gamma$, the theorem implies the weak estimate $\sum_{I \in \mathcal{G}} |I|^\gamma \le C|f|_{\mathrm{BV}(\mathbb{R}^d)}\epsilon^{-1}$. From Lemma 4.3 we see that the proof of Theorem 4.1 is now complete. Combined with the results of §3, this also completes the proof of Theorem 2.5.

5. Application to wavelet decompositions

The results of the two previous sections prove Theorem 2.5. We cannot simply replace the coefficients $c_I(f)$ by the wavelet coefficients f_I in this theorem, because, in contrast to the functions g_I, the compactly supported dual wavelets $\tilde{\psi}_I$ generally have a support strictly larger than I, so that $|f_I| \le C w_I(f)$ need not be true. In order to circumvent this problem we use a technique proposed by Meyer, already applied in [5].

Theorem 5.1 *Let $f \in \mathrm{BV}(\mathbb{R}^d)$. Then $(f_I)_{I\in\mathcal{D}} \in w\ell_1^\gamma(\mathcal{D})$ for all $\gamma < 1 - 1/d$ or $\gamma > 1$. More precisely, there exists a constant C only depending on γ such that for all $f \in \mathrm{BV}(\mathbb{R}^d)$ and $\epsilon > 0$ we have*

$$\sum_{|f_I| > \epsilon |I|^\gamma} \le C|f|_{\mathrm{BV}(\mathbb{R}^d)}\epsilon^{-1}. \tag{5.1}$$

Proof. We first remark that up to a shift of spatial indices, we can always assume that the generators $\tilde{\psi}^e$ of the dual wavelets are supported in $]0,p[^d$ where p is a sufficiently large *prime* integer. We fix any $e \in E$ and define $g := \tilde{\psi}^e(p\cdot)$, which is now supported in $]0,1[^d$. For an arbitrary but fixed $r \in P := \{0, 1, \ldots, p-1\}^d$, we define $\tilde{f}_r(x) := p^{d-1} f(px + r)$. Theorem 2.5 implies that for $\gamma < 1 - 1/d$ or $\gamma > 1$, we have the weak-type estimate

$$\sum_{c_I^r(f) > \epsilon |I|^\gamma} |I|^\gamma \le C|\tilde{f}_r|_{\mathrm{BV}(\mathbb{R}^d)}\epsilon^{-1} = C|f|_{\mathrm{BV}(\mathbb{R}^d)}\epsilon^{-1}, \tag{5.2}$$

with $c_I^r(f) := |\langle \tilde{f}_r, g_I\rangle|$. Now for $I := 2^{-j}(]0,1[^d + k)$, we have

$$\begin{aligned} \langle \tilde{f}_r, g_I \rangle &= 2^j \int_{\mathbb{R}^d} \tilde{f}_r(x) g(2^j x - k)dx = 2^j \int_{\mathbb{R}^d} p^{d-1} f(px + r) g(2^j x - k) dx \\ &= p^{-1} 2^j \int_{\mathbb{R}^d} f(x) \tilde{\psi}^e(2^j x - l) dx = p^{-1} f_J^e, \end{aligned}$$

where $l := 2^j r + pk$ and $J = 2^{-j}(]0,1[^d + l)$.

We next observe that for $j \in \mathbb{N}$, the mapping

$$(k, r) \mapsto 2^j r + pk \tag{5.3}$$

is a bijection from $\mathbb{Z}^d \times P$ onto $\mathbb{Z}^d$, due to the fact that $x \mapsto 2^j x$ is a bijection from $(\mathbb{Z}/p\mathbb{Z})^d$ onto itself. In other words, each coefficient f_I^e appears as one of the $c_I^r(f)$, $r \in P$. Since the sets P and E are finite, it follows from these observations that

$$\sum_{|f_I|>\epsilon|I|^\gamma, I\in\mathcal{D}_+} |I|^\gamma \leq C|f|_{\mathrm{BV}(\mathbb{R}^d)}\epsilon^{-1}. \tag{5.4}$$

We can remove the restriction that $I \in \mathcal{D}_+$ in (5.4) as follows. We apply this estimate to $f_q := 2^{q(1-1/d)} f(2^q\cdot)$, $q \in \mathbb{N}$, and observe that, by a change of variable,

$$\sum_{|(f_q)_{\tilde{I}}|>\tilde{\epsilon}|\tilde{I}|^\gamma, \tilde{I}\in\mathcal{D}_+} |\tilde{I}|^\gamma \leq C|f_q|_{\mathrm{BV}(\mathbb{R}^d)}\tilde{\epsilon}^{-1},$$

becomes

$$\sum_{|f_I|>\epsilon|I|^\gamma, |I|\leq 2^{dq}} |I|^\gamma \leq C|f|_{\mathrm{BV}(\mathbb{R}^d)}\epsilon^{-1},$$

where $|I| = 2^{qd}|\tilde{I}|$ and $\tilde{\epsilon} = \epsilon 2^{q(1-d)(1-1/d)}$; C is still the same constant, independent of q and ϵ. By letting q go to $+\infty$, we arrive at (5.1). ■

A similar result can be derived for the non-homogeneous basis associated with the decomposition (1.12).

Theorem 5.2 *Let $f \in \mathrm{BV}(\mathbb{R}^d)$. Then $(F_I)_{I\in\mathcal{D}_+} \in w\ell_1^\gamma(\mathcal{D}_+)$ for all $\gamma < 1 - 1/d$ or $\gamma > 1$. More precisely, there exists a constant C depending only on γ such that for all $f \in \mathrm{BV}(\mathbb{R}^d)$ and $\epsilon > 0$ we have*

$$\sum_{|F_I|>\epsilon|I|^\gamma} |I|^\gamma \leq C\|f\|_{\mathrm{BV}(\mathbb{R}^d)}\epsilon^{-1}. \tag{5.5}$$

Proof. By Theorem 5.1, we already have the weak type estimate

$$\sum_{|F_I|>\epsilon|I|^\gamma, |I|<1} |I|^\gamma \leq C\|f\|_{BV}\epsilon^{-1}, \tag{5.6}$$

since $F_I = f_I$ if $|I| < 1$. For $|I| = 1$, we have a strong estimate

$$\sum_{I\in\mathcal{D}_0} |F_I| \leq C \sum_{I\in\mathcal{D}_0} \int_{\mathrm{supp}(\tilde{\psi}_I)} |f| \leq CA^d\|f\|_{L_1(\mathbb{R}^d)}, \tag{5.7}$$

where A is such that $\mathrm{supp}(\tilde{\psi}^e) \subset [0, A]^d$ for all $e \in E'$. Combining these estimates, we obtain (5.5). ■

The above results can easily be adapted to most constructions of wavelets defined on simple bounded domains such as a cube (see e.g. [2] or [6] for examples of such constructions).

Let us finally mention that the weak type estimates of Theorem 5.1 and Theorem 5.2 have equivalent formulations in terms of the approximation performance of thresholding procedures studied in [3]. For $0 < r \leq \infty$, consider the L_r-thresholding operator $\mathcal{T}_\epsilon^r$ defined by

$$\mathcal{T}_\epsilon^r f := \sum_{\|f_I \psi_I\|_{L_r} > \epsilon} f_I \psi_I \tag{5.8}$$

The results of [3] show that for $0 < p < \infty$, the rate of decay of $\|f - \mathcal{T}_\epsilon^r f\|_{H_p}$, as ϵ goes to zero, is determined by weighted weak-type estimates on the renormalized coefficient sequence $(\|f_I \psi_I\|_{L_r})_{I \in \mathcal{D}}$. Here H_p denotes the Hardy space which coincides with L_p when $p > 1$. More precisely, we have by Theorem 7.1 of [3] that for $\mu < p$,

$$\|f - \mathcal{T}_\epsilon^r f\|_{H_p} < C^{\mu/p} \epsilon^{1-\mu/p}, \tag{5.9}$$

if and only if the sequence $(\|f_I \psi_I\|_{L_r})_{I \in \mathcal{D}}$ belongs to the space $w\ell_\mu(\mathcal{D}, |I|^\gamma)$ with $\gamma := 1 - p/r$; the smallest C satisfying (5.9) is then equivalent to $\|(\|f_I \psi_I\|_{L_r})_{I \in \mathcal{D}}\|_{w\ell_\mu(\mathcal{D}, |I|^\gamma)}$. Note that when $p = r$, i.e. when we use the same metric for thresholding as for measuring the approximation error, we find the standard $w\ell_\mu$ spaces. However, there are situations in which one prefers to use different metrics for thresholding and measuring the approximation error, such as in statistical estimation, where one may be interested in estimating a noisy function in some arbitrary L_p norm, but where the structure of the white noise imposes the L_2 metric for thresholding. For this particular situation, the case $\gamma < 1 - 1/d$ in Theorem 5.1, combined with (5.9) implies the following result:

Theorem 5.3 *Let $f \in \mathrm{BV}(\mathbb{R}^d)$. Then for $0 < r < \infty$ and $p = 1 + r/d$, we have the thresholding estimate*

$$\|f - \mathcal{T}_\epsilon^r f\|_{L_p} \leq C |f|_{BV(\mathbb{R}^d)} \epsilon^{1-1/p}. \tag{5.10}$$

6. Counter-examples

The purpose of this last section is to prove that Theorem 5.1 and Theorem 5.2 are no longer true for the range $1 - 1/d \leq \gamma \leq 1$. We shall exhibit counter-examples in the case where the wavelets are given by the Haar system, i.e. $\varphi = \tilde{\varphi} := \chi_{]0,1[}$ and $\psi = \tilde{\psi} := \chi_{]0,1/2[} - \chi_{]1/2,1[}$. We first consider the one-dimensional setting, corresponding to the range $0 \leq \gamma \leq 1$. If I is a dyadic interval, then the BV-normalized wavelet coefficient f_I of a function f is given by

$$f_I = \langle f, \tilde{\psi}_I \rangle = \frac{1}{2}(a_{I_l}(f) - a_{I_r}(f)), \tag{6.1}$$

where $a_{I_l}(f)$ and $a_{I_r}(f)$ are respectively the averages of f on I_l and I_r, the left and right half-intervals of I. The counter examples that we shall build are functions supported on $]0,1[$ and we shall consider their wavelet coefficients only for $|I| \leq 1$. We shall treat separately the cases $\gamma = 0$, $\gamma = 1$ and $0 < \gamma < 1$.

In the case $\gamma = 1$, we consider the function

$$f(x) = x\chi_{]0,1[}(x),$$

which is clearly in $\mathrm{BV}(\mathbb{R})$. If $I \subset]0,1[$, a straightforward computation shows that

$$f_I = -|I|/4.$$

Therefore, taking $\epsilon = 1/5$, we obtain that

$$\sum_{|f_I| > \epsilon |I|} |I| \geq \sum_{I \subset]0,1[} |I| = +\infty,$$

which shows that the weak estimate does not hold when $\gamma = 1$.

In the case $\gamma = 0$ we consider the function

$$f = \chi_{[0,1/3]},$$

which is clearly in $\mathrm{BV}(\mathbb{R})$. For each $j \geq 0$, there exist one dyadic interval I_j containing the jump point $1/3$ and such that $|I_j| = 2^{-j}$. This jump point is always located at either the $1/3$ or $2/3$ position of I_j, since $1/3 = \sum_{j=1}^{\infty} 2^{-2j}$. It follows that

$$f_{I_j} = 1/3,$$

for all $j \geq 0$. Therefore taking $\epsilon = 1/4$, we obtain that

$$\#\{I : |f_I| > \epsilon\} = \infty,$$

which shows that the weak estimate does not hold when $\gamma = 0$.

We now consider the case $0 < \gamma < 1$. Here, we set $\alpha := 1/\gamma > 1$ and we define the sequence $(j_k)_{k \geq 0}$ of integers by

$$2^{-j_k - 1} < 2^{-\alpha k} \leq 2^{-j_k}. \tag{6.2}$$

Note that $j_0 = 0$ and that the sequence j_k is strictly increasing because $\alpha > 1$. More precisely, we can write

$$j_{k+1} = j_k + m_k,$$

with $m_k > 0$ for all k and $m_k > 1$ for infinitely many k.

We now construct a family of piecewise constant functions $(f_n)_{n>0}$ as follows. For each $n > 0$, the distributional derivative of f_n is a sum of Dirac masses:

$$f'_n = 2^{-n}\Big[\sum_{m=0}^{2^n-1} \delta_{x^n_m}\Big] - \delta_1,$$

where the 2^n jump points $x^n_m \in]0,1[$ will be specified in a moment. Since we subtract δ_1, these functions are supported on $[0,1]$. Clearly $|f_n|_{\mathrm{BV}(\mathbb{R})} = 2$ and $\|f_n\|_{\mathrm{BV}(\mathbb{R})} \leq 3$, independently of n.

If I is a dyadic interval, the wavelet coefficient c_I of f_n is given by

$$c_I := \langle f'_n, h_I\rangle = 2^{-n} \sum_{m \text{ s.t. } x^n_m \in I} h_I(x^n_m), \tag{6.3}$$

where h_I is the primitive function of $\tilde{\psi}_I$, i.e., the hat function $h(x) = (1 - |x|)_+$ rescaled to I.

We have not yet specified where to position the points x^n_m in $]0,1[$. We wish to place them so that the right sum in (6.3) is large for many choices of I. For each $k = 0, \dots, n$, we shall inductively construct 2^k pairwise disjoint dyadic intervals $I_{k,l}$, $l = 0, \dots, 2^k - 1$ of size $|I_{k,l}| = 2^{-j_k}$, and position the points x^n_m so that

$$S_{k,l} := \{x^n_m\}_{m=2^{n-k}l}^{2^{n-k}(l+1)-1} \subset I_{k,l}. \tag{6.4}$$

We start the construction with $I_{0,0} =]0,1[$, and for $k = 1, \dots, n-1$ the construction is continued using the following iteration: for a given $I_{k,l}$, we define $I_{k+1,2l}$ and $I_{k+1,2l+1}$ as the two adjacent dyadic intervals of size $2^{-j_{k+1}}$ which respectively admit the center of $I_{k,l}$ as their right and left endpoints. Iterating this construction, it suffices to choose each point x^n_m in the corresponding interval $I_{n,m}$. In the case where $m_k > 1$, we thus notice that all the points in $S_{k,l}$ are concentrated in a central region of $I_{k,l}$ on which $h_{I_{k,l}}(x) \geq 1/2$, so that according to (6.3) and (6.2) we then have

$$c_{I_{k,l}} \geq 2^{-n-1}\#(S_{k,l}) = 2^{-k-1} \geq 2^{-1-\gamma}|I_{k,l}|^\gamma.$$

Therefore, if we fix $\epsilon = 1/4 < 2^{-1-\gamma}$, we see that for $k < n$ such that $m_k > 1$, we have

$$\sum_{c_I \geq \epsilon |I|^\gamma, |I| = 2^{-j_k}} |I|^\gamma = 2^k 2^{-\gamma j_k} \geq 1,$$

and thus

$$\sum_{c_I \geq \epsilon |I|^\gamma} |I|^\gamma \geq \#\{k \text{ s.t. } k < n \text{ and } m_k > 1\} = K(n).$$

Since $\lim_{n\to+\infty} K(n) = +\infty$, this shows that the weak estimate does not hold for $0 < \gamma < 1$.

We shall now generalize the above counterexamples to the multidimensional case for $1 - 1/d \le \gamma \le 1$ by using the following observations. If f is a one-dimensional $\mathrm{BV}(\mathbb{R})$ function supported in $[0, 1]$, then the multidimensional function

$$F(x_1, \dots, x_d) := f(x_1)\chi_{[0,1]^d}(x_1, \dots, x_d)$$

is in $BV(\mathbb{R}^d)$ with $\|F\|_{\mathrm{BV}(\mathbb{R}^d)} \le C(d)\|f\|_{\mathrm{BV}(\mathbb{R})}$. Moreover if

$$I = I_1 \times \cdots \times I_d$$

is a dyadic cube contained in $]0, 1[^d$, and if $e = (1, 0, \cdots, 0)$, we have

$$\tilde{\psi}_I^e(x_1, \dots, x_d) = \tilde{\psi}_{I_1}(x_1)\chi_{I_2\times\cdots\times I_d}(x_2, \dots, x_d).$$

and therefore the wavelet coefficients $c_I^e(F)$ of F satisfy

$$c_I^e(F) = \langle F, \tilde{\psi}_I^e\rangle = |I|^{1-1/d}\langle f, \tilde{\psi}_{I_1}\rangle = |I|^{1-1/d} c_{I_1}(f).$$

It follows that

$$\begin{aligned}\sum_{|c_I^e(F)|\ge\epsilon|I|^\gamma} |I|^\gamma &= \sum_{\substack{I_1 \text{ such that} \\ |c_{I_1}(f)| \ge \epsilon|I|^{\gamma+1/d-1}}} \ \sum_{\substack{(I_2,\dots,I_d) \text{ such that} \\ I = I_1\times\cdots\times I_d}} |I|^\gamma \\ &= \sum_{I_1 \text{ s.t. } |c_{I_1}(f)|\ge\epsilon|I_1|^{\tilde{\gamma}}} |I_1|^{\tilde{\gamma}},\end{aligned}$$

with $\gamma = 1 - 1/d + \tilde{\gamma}/d$. Applying these observations to the above one-dimensional counter-examples for $0 \le \tilde{\gamma} \le 1$, we thus obtain our multidimensional counter-examples for $1 - 1/d \le \gamma \le 1$.

Acknowledgments: Large parts of the work for this paper were done during visits of various subsets of the authors to the Laboratoire d'Analyse Numérique at the Université Pierre et Marie Curie and to the Industrial Mathematics Institute at the University of South Carolina. We are grateful to these institutions for their hospitality and support. In addition, I.D. would also like to thank the Institute for Advanced Study in Princeton, where she spent a sabbatical semester while working on this project.

References

[1] Bergh, J. and Löfström, J.: *Interpolation spaces*. Springer Verlag, 1976.

[2] Cohen, A.: *Wavelet methods in numerical analysis*, in the Handbook of Numerical Analysis, vol. VII, P.-G. Ciarlet et J.-L. Lions eds., Elsevier, Amsterdam, 2000.

[3] Cohen, A., DeVore, R. and Hochmuth, R.: Restricted Nonlinear Approximation. *Constructive Approximation* **16** (2000), 85–113.

[4] Cohen, A., DeVore, R., Petrushev, P. and Xu, H.: Non linear approximation and the space $BV(\mathbb{R}^2)$. *Amer. J. Math.* **121** (1999), 587–628.

[5] Cohen, A., Meyer, Y. and Oru, F.: *Improved Sobolev inequalities*. Proceedings séminaires X-EDP, Ecole Polytechnique, Palaiseau, 1998.

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

[7] Daubechies, I.: *Ten Lectures on Wavelets*. SIAM, 1992.

[8] Donoho, D., Vetterli, M., DeVore, R. and Daubechies, I.: Harmonic analysis and signal processing. *IEEE Trans. Inf. Theory* **44** (1998), 2435–2476.

[9] DeVore, R.: *Nonlinear Approximation*. Acta Numerica **7**, Cambridge University Press, 1998, 51–150.

[10] DeVore, R. and Petrova, G.: The averaging lemma. *J. Amer. Math. Soc.* **14** (2000), 279–296.

[11] Donoho, D.: Unconditional bases are optimal for data compression and statistical estimation. *Appl. Comp. Harm. Anal.* **1** (1993), 100–105.

[12] Evans, L. and Gariepy, R.: *Measure theory and fine properties of functions*. CRC Press, New York, 1992.

[13] Meyer, Y.: *Ondelettes et Opérateurs*. Hermann, Paris, 1990.

[14] Ziemer, W.: *Weakly differentiable functions*. Springer Verlag, New York, 1989.

Recibido: 15 de febrero de 2002

Albert Cohen
Laboratoire d'Analyse Numérique
Université Pierre et Marie Curie
4 Place Jussieu, 75252 Paris cedex 05
France
cohen@ann.jussieu.fr
http://www.ann.jussieu.fr/∼cohen/

Wolfgang Dahmen
Institut für Geometrie und Praktische Mathematik
RWTH Aachen
Templergraben 55
52056 Aachen, Germany
dahmen@igpm.rwth-aachen.de
http://www.igpm.rwth-aachen.de/~dahmen/

Ingrid Daubechies
Department of Mathematics and Program in
Applied and Computational Mathematics
Princeton University
Fine Hall, Washington Road
Princeton, NJ 08544-1000, USA
ingrid@math.princeton.edu
http://www.math.princeton.edu/~icd/

Ronald DeVore
Department of Mathematics
University of South Carolina
Columbia, SC 29208, USA
devore@math.sc.edu
http://www.math.sc.edu/~devore/