Greedy Wavelet Projections are Bounded on BV *

Paweł Bechler, Ronald DeVore, Anna Kamont, Guergana Petrova, Przemysław Wojtaszczyk

October 21, 2003

Abstract

Let $\mathrm{BV} = \mathrm{BV}(\mathbb{R}^d)$ be the space of functions of bounded variation on $\mathbb{R}^d$ with $d \geq 2$. Let ψ_λ, $\lambda \in \Delta$, be a wavelet basis of compactly supported functions normalized in BV, i.e. $|\psi_\lambda|_{\mathrm{BV}(\mathbb{R}^d)} = 1$, $\lambda \in \Delta$. Each $f \in \mathrm{BV}$ has a unique wavelet expansion $\sum_{\lambda \in \Delta} c_\lambda(f)\psi_\lambda$ with convergence in $L_1(\mathbb{R}^d)$. If $\Lambda_N(f)$ is the set of N indicies $\lambda \in \Delta$ for which $|c_\lambda(f)|$ are largest (with ties handled in an arbitrary way), then $\mathcal{G}_N(f) := \sum_{\lambda \in \Lambda_N(f)} c_\lambda(f)\psi_\lambda$ is called a greedy approximation to f. It is shown that $|\mathcal{G}_N(f)|_{\mathrm{BV}(\mathbb{R}^d)} \leq C|f|_{\mathrm{BV}(\mathbb{R}^d)}$ with C a constant independent of f. This answers in the affirmative a conjecture of Meyer [15] (see p. 79).

AMS subject classification: 42C40, 46B70, 26B35, 42B25.
Key Words: N-term approximation, greedy approximation, functions of bounded variation, thresholding, bounded projections.

1 Introduction

The space $\mathrm{BV} := \mathrm{BV}(\Omega)$ of functions of bounded variation on a domain $\Omega \subset \mathbb{R}^d$ is important in mathematics (geometric measure theory, differential geometry) and applications (image processing, nonlinear PDEs). The structure of BV is complicated by the fact that neither it nor the closely related Sobolev space $W^1(L_1(\Omega))$ have an unconditional basis (BV does not even have a basis). Wavelet decompositions of BV functions, while not characterizing this space, give fine information (see [4, 19, 2]) about its structure and these decompositions can be used to solve various extremal problems.

Consider, for example, the extremal problem

$$K(f,t) := K(f,t; L_2(\Omega), \mathrm{BV}(\Omega)) := \inf_{g \in \mathrm{BV}(\Omega)} \|f - g\|_{L_2(\Omega)} + t|g|_{\mathrm{BV}(\Omega)}, \tag{1.1}$$

where $\Omega = [0,1]^2$ and $t > 0$ is a parameter. The expression (1.1) is called a K-functional in interpolation of linear operators. It is used to describe interpolation spaces between $L_2(\Omega)$

*This work has been supported in part by the NRC New Investigators Twinning Program 2003-2004 as well as the Office of Naval Research Contract N00014-03-1-0051, the Air Force of Scientific Research Contracts UFEIES0302005USC and DAAD 19-02-1-0028, the Foundation for Polish Science and KBN grant 5P03A 03620 located at the Institute of Mathematics of the Polish Academy of Sciences.

and BV(Ω). This and related functionals also occur in image processing in such problems as denoising and deblurring. The rate of decay of $K(f,t)$ as $t \to 0$ gives information about the smoothness of f relative to $L_2(\Omega)$ and BV(Ω). A function $g = g_t$ is a called a near minimizer (with constant C) to (1.1) if

$$\|f - g_t\|_{L_2(\Omega)} + t|g_t|_{\mathrm{BV}(\Omega)} \leq CK(f,t).$$

One would like simple constructive methods for finding minimizers or near minimizers to (1.1).

In [4], it is shown that thresholding the Haar decomposition of f provides a near minimizer to (1.1). Namely, if H_λ, $\lambda \in \Delta$, is the Haar basis on $[0,1]^2$, then given $f \in L_2([0,1]^2)$, we can write

$$f = \sum_{\lambda \in \Delta} c_\lambda(f) H_\lambda$$

with the H_λ normalized in $L_2([0,1]^2)$ (which is equivalent to normalizing in $\mathrm{BV}([0,1]^2)$). For each $t > 0$, a near minimizer g_t is given by thresholding the Haar series:

$$g_t := T_{t^2} f := \sum_{\lambda \in \Lambda(f,t^2)} c_\lambda(f) H_\lambda,$$

where for any $t > 0$

$$\Lambda(f,t) := \{\lambda : |c_\lambda(f)| > t\}.$$

The proof that thresholding is a near minimizer relies on three basic results concerning Haar decompositions and BV. To describe these, we introduce the concept of N-term approximation using the Haar basis. We define Σ_N^w as the collection of all functions $S = \sum_{\lambda \in \Lambda} c_\lambda H_\lambda$, where $\Lambda \subset \Delta$ is any index set with cardinality $\#(\Lambda) \leq N$. Given $f \in L_2([0,1]^2)$, we consider the approximation of f using the elements of Σ_N^w:

$$\sigma_N(f)_{L_2([0,1]^2)} := \inf_{S \in \Sigma_N^w} \|f - S\|_{L_2([0,1]^2)}.$$

The first of these basic results is the following direct estimate (see [4]) for the approximation error:

$$\sigma_N(f)_{L_2([0,1]^2)} \leq C_0 N^{-1/2} |f|_{\mathrm{BV}([0,1]^2)}, \quad N = 1, 2, \ldots.$$

This inequality is called an inequality of Jackson type (corresponding to analogous inequalities in approximation by algebraic polynomials). The Jackson inequality is proved by showing that the Haar coefficients of a $\mathrm{BV}([0,1]^2)$ function are in weak ℓ_1. That is

$$\#(\Lambda(f,\epsilon)) \leq C_0 \epsilon^{-1}, \quad \epsilon > 0.$$

This weak ℓ_1 property was shown in [6] to hold in the more general setting of wavelet expansions of functions in $\mathrm{BV}(\mathbb{R}^d)$ using compactly supported orthogonal wavelets. This allows the generalization of the Jackson inequality to arbitrary space dimensions and arbitrary compactly supported orthogonal wavelet systems (see Lemma 4.2).

The second basic result (see [4]) is the Bernstein inequality which (in the case of $[0,1]^2$) says that

$$|S|_{\mathrm{BV}([0,1]^2)} \le C_0 N^{1/2} \|S\|_{L_2([0,1]^2)}, \quad S \in \Sigma_N^w,\ N = 1,2,\dots.$$

We shall show in §5 that this inequality also generalizes to $\mathbb{R}^d$ and general compactly supported orthogonal wavelet systems.

The Jackson and Bernstein inequalities are not enough to show that thresholding the Haar expansion is an approximate minimizer for (1.1). One also needs the stability of thresholding in $\mathrm{BV}([0,1]^2)$:

$$|T_\epsilon(f)|_{\mathrm{BV}([0,1]^2)} \le C_0 |f|_{\mathrm{BV}([0,1]^2)}, \quad f \in \mathrm{BV}([0,1]^2).$$

This remarkable property says that projecting onto any sum involving the N largest wavelet coefficients of the Haar series of a function in $\mathrm{BV}([0,1]^2)$ results in a function with controllable $\mathrm{BV}([0,1]^2)$ norm. Note that this property does not hold for projecting onto an arbitrary N-term sum of the Haar series nor does it hold in $\mathbb{R}^1$ (see §7). This stability result for Haar expansions was generalized to space dimensions $d > 2$ in [19]. Yves Meyer [15] (see p. 79) has conjectured that this property holds for any compactly supported wavelet system. The main result of this paper is to prove this conjecture.

Theorem 1.1 *Let φ be a compactly supported univariate scaling function in* $\mathrm{BV}(\mathbb{R}^1)$ *which generates the compactly supported orthogonal wavelet ψ. For $d \ge 2$, we consider the multivariate orthogonal wavelet system $(\psi_\lambda)_{\lambda\in\Delta}$ obtained from φ and ψ, and normalized in* $\mathrm{BV}(\mathbb{R}^d)$*. Then this wavelet system has the following* BV *stability property. If $f \in$* $\mathrm{BV}(\mathbb{R}^d)$*, $d \ge 2$, let*

$$f = \sum_{\lambda\in\Delta} c_\lambda(f)\psi_\lambda$$

be the wavelet expansion of f. Let for any N, $\Lambda_N(f)$ be the set of N indices $\lambda \in \Delta$ for which $|c_\lambda(f)|$ are largest. Then the nonlinear operator

$$\mathcal{G}_N(f) := \sum_{\lambda\in\Lambda_N(f)} c_\lambda(f)\psi_\lambda$$

satisfies

$$|\mathcal{G}_N(f)|_{\mathrm{BV}(\mathbb{R}^d)} \le C(\varphi,d)|f|_{\mathrm{BV}(\mathbb{R}^d)}.$$

As a consequence of this theorem we shall also show that $\mathcal{G}_N(f)$ realizes the K-functional for the pair $(L_{d^*}(\mathbb{R}^d), \mathrm{BV}(\mathbb{R}^d))$.

Theorem 1.2 *Let φ be a compactly supported univariate scaling function in* $\mathrm{BV}(\mathbb{R}^1)$ *which generates the compactly supported orthogonal wavelet ψ. For $d \ge 2$, we consider the multivariate orthogonal wavelet system $(\psi_\lambda)_{\lambda\in\Delta}$ obtained from φ and ψ, and normalized in* $\mathrm{BV}(\mathbb{R}^d)$*. Then the greedy operator*

$$\mathcal{G}_N(f) := \sum_{\lambda\in\Lambda_N(f)} c_\lambda(f)\psi_\lambda,$$

with $\Lambda_N(f)$ the set of N indices $\lambda \in \Delta$ for which $|c_\lambda(f)|$ are largest, satisfies

$$\|f - \mathcal{G}_N(f)\|_{L_{d^*}(\mathbb{R}^d)} + N^{-1/d}|\mathcal{G}_N(f)|_{\mathrm{BV}(\mathbb{R}^d)} \le C(\varphi,d)K(f, N^{-1/2}; L_{d^*}(\mathbb{R}^d), \mathrm{BV}(\mathbb{R}^d)). \quad (1.2)$$

2 The space BV

There are several treatments of the space BV. We mention two valuable references [15, 20] which contain all of the properties of BV functions that we shall need. There are several equivalent definitions of BV. The approach we take below is simply the most direct and convenient for our setting.

Let Ω be an open set in $I\!R^d$. We begin with the Sobolev space $W^1(L_1(\Omega))$ which is the collection of all functions in $L_1(\Omega)$ such that the distributional gradient ∇f is also in $L_1(\Omega)$. The semi-norm on this space is

$$|f|_{W^1(L_1(\Omega))} := \|\nabla f\|_{L_1(\Omega)},$$

and the norm for this space is obtained by adding the $L_1(\Omega)$ norm:

$$\|f\|_{W^1(L_1(\Omega))} := |f|_{W^1(L_1(\Omega))} + \|f\|_{L_1(\Omega)}.$$

The space $\mathrm{BV}(\Omega)$ can now be defined as the set of all $f \in L_1(\Omega)$ for which there is a sequence (f_n) satisfying

$$\|f - f_n\|_{L_1(\Omega)} \to 0, \quad \sup_n |f_n|_{W^1(L_1(\Omega))} < \infty. \tag{2.1}$$

The semi-norm on BV is then defined as

$$\inf_{(f_n)} \liminf_{n\to\infty} |f_n|_{W^1(L_1(\Omega))}, \tag{2.2}$$

where the infimum is taken over all sequences satisfying (2.1). To see that this definition is equivalent to other definitions of BV the reader should consult Theorems 5.2.1 and 5.3.3 in [20].

We mention a couple of properties of the BV norm that we shall use in this paper.

Remark 2.1 *In the case $\Omega = I\!R^d$, the functions f_n appearing in* (2.1) *and* (2.2) *can be taken to be in $C^\infty(I\!R^d)$ with compact support.*

Remark 2.2 *Let I_0 be a dyadic cube in $I\!R^d$ and I_j, $j = 1, \dots, m$, be a finite collection of disjoint dyadic cubes each of which is contained in I_0. Let χ_{I_k} be the characteristic function of I_k, $k = 0, \dots, m$. Then the function $f = \chi_{I_0} - \sum_{j=1}^m \chi_{I_j}$ has* BV *semi-norm*

$$|f|_{\mathrm{BV}(I\!R^d)} \le \sum_{j=0}^m \mathrm{meas}_{d-1}(\partial I_j), \tag{2.3}$$

where $\partial\Omega$ denotes the boundary of a set $\Omega \subset I\!R^d$ and meas_{d-1} *is the $(d-1)$ dimensional surface measure.*

The second result can be proved directly or derived from the well known co-area formula for BV functions (see [20], p. 231). We have equality in (2.3) if the boundaries of the I_j, $j = 0, 1, \dots, m$, are disjoint.

Remark 2.3 *If $\Omega_j \subset I\!R^d$, $j = 1, \dots, m$, is a partition of Ω, then*

$$\sum_{j=1}^m |f|_{\mathrm{BV}(\Omega_j)} \le |f|_{\mathrm{BV}(\Omega)}. \tag{2.4}$$

This follows from the set additivity of the L_1 semi-norm in the case $f \in W^1(L_1(\Omega))$ and by taking limits in the general case $f \in \mathrm{BV}(\Omega)$.

3 Wavelet decompositions

We shall limit our analysis to the case of compactly supported orthogonal wavelets on $I\!R^d$. The results we put forward in this paper hold equally well for biorthogonal compactly supported wavelets with the same proofs but somewhat more cumbersome notation.

Let φ be a compactly supported univariate scaling function with orthogonal shifts which satisfies the two scale relation

$$\varphi(x) = \sum_k \alpha_k \varphi(2x - k),$$

where only a finite number of the α_k are nonzero. We shall assume throughout this paper that φ is in $\mathrm{BV}(I\!R^1)$. Let ψ be the univariate wavelet function with compact support which is obtained from φ by multiresolution. Examples of such wavelets and scaling functions were given by Daubechies [8].

We use the standard construction of multidimensional wavelet bases. Let E' denote the set of vertices of the cube $[0,1]^d$ and E denote the set of nonzero vertices. We shall use the notation $\psi^0 := \varphi$ and $\psi^1 := \psi$. For each $e \in E'$, we define

$$\psi^e(x_1, \ldots, x_d) := \psi^{e_1}(x_1) \cdots \psi^{e_d}(x_d).$$

Let $\mathcal{D}$ denote the set of dyadic cubes in $I\!R^d$ and let $\mathcal{D}_k$ denote those dyadic cubes which have sidelength 2^{-k} and $\mathcal{D}_+ := \cup_{k \geq 0} \mathcal{D}_k$. For any dyadic cube $I = 2^{-k}(j + [0,1]^d) \in \mathcal{D}_k$, $k \in Z\!\!\!Z$, $j \in Z\!\!\!Z^d$, we define the functions

$$\psi_I^e(x) := \gamma(I,e)\psi^e(2^k x - j), e \in E',$$

with the $\gamma(I,e) > 0$ chosen so that

$$|\psi_I^e|_{\mathrm{BV}(I\!R^d)} = 1, \quad I \in \mathcal{D},\ e \in E'.$$

These functions are scaled to I. It follows that the constants $\gamma(I,e) = |I|^{-1/d^*}\gamma(e)$[1] with $d^* := \frac{d}{d-1}$ and therefore we have

$$c_1 \leq \|\psi_I^e\|_{L_{d^*}(I\!R^d)} \leq c_2, \quad I \in \mathcal{D}, e \in E',$$

with constants c_1, c_2 depending only on φ and d. In other words, normalization in BV is equivalent to normalization in L_{d^*}.

To simplify the notation that follows, we introduce the indexing set Δ which consists of all pairs $\lambda = (I,e)$ with $I \in \mathcal{D}_+$ and $e \in E$ ($e \in E'$ if $I \in \mathcal{D}_0$). We define $|\lambda| := k$ when $I \in \mathcal{D}_k$. The set of functions $\{\psi_\lambda\}_{\lambda \in \Delta}$ is a complete orthogonal system. Any locally integrable function f on $I\!R^d$ has a formal wavelet series

$$f = \sum_{\lambda \in \Delta} c_\lambda(f)\psi_\lambda,$$

where the wavelet coefficients $c_\lambda(f)$ are given by

$$c_\lambda(f) := c_I^e(f) := \langle f(\cdot), \gamma'(e,I)\psi^e(2^k \cdot -j)\rangle, \quad \lambda = (I,e) \in \Delta,\ I = 2^{-k}(j + [0,1]^d),$$

[1] Througout this paper, we shall use the notation $|A|$ to denote the Lebesgue measure of a set $A \subset I\!R^d$.

where the normalization factors $\gamma'(e,I)$ of the dual wavelet scale are like $\gamma'(I,e) \sim |I|^{-1/d}$.

The set of functions $\{\psi_\lambda\}_{\lambda\in\Delta}$ is a basis for many function spaces. For example, they are an orthogonal basis for $L_2(I\!R^d)$. They are an unconditional basis for the L_p spaces $1 < p < \infty$ and for the Besov spaces whenever they admit an unconditional basis. They are a basis for $W^1(L_1(\Omega))$, but not unconditional (this space does not admit an unconditional basis).

We shall use the abbreviated notation $\phi := \psi^{(0,\dots,0)}$ for the function which is a tensor product of scaling functions. Similarly, we write

$$\phi_I(x) := |I|^{-\frac{1}{d^*}}\phi(2^k x - j), \quad I = 2^{-k}(j + [0,1]^d),$$

to index the scaling functions at level k. The shift invariant space $\mathcal{S}_k := \mathcal{S}_k(\phi)$ is the span of the functions ϕ_I, $I \in \mathcal{D}_k$. Each space $\mathcal{S}_k$ is a dilate of the space $\mathcal{S}_0$. At each dyadic level k, the shifts ϕ_I, $I \in \mathcal{D}_k$ sum to a constant:

$$\sum_{I\in\mathcal{D}_k} \phi_I = c|I|^{-\frac{1}{d^*}} \tag{3.1}$$

with c a constant. Any wavelet ψ_λ or scaling function at a dyadic level $j < k$ (i.e. $|\lambda| = j$) is an element in $\mathcal{S}_k$ and can be written as a finite linear combination of the ϕ_I, $I \in \mathcal{D}_k$.

4 Approximation by piecewise constants

We shall use in the course of our proofs some results on approximation of BV functions by piecewise constant functions. Throughout this and the next section, we assume that $d \geq 2$. The results we shall need are for the most part proved in two earlier works [4] (for the case $d = 2$) and [19] (for the case $d > 2$).

We shall discuss three types of approximation by piecewise constants. The first of these is N-term approximation using Haar functions. In this case, we can be more general and treat N-term approximation using compactly supported wavelets. So let $(\psi_\lambda)_{\lambda\in\Delta}$ be one of the wavelet bases introduced in the previous section. We take the basis functions ψ_λ to be normalized in BV.

We define the nonlinear space

$$\Sigma_N^w := \{\sum_{\lambda\in\Lambda} c_\lambda\psi_\lambda : \ \#(\Lambda) \leq N\}.$$

Thus, each element in Σ_N^w is a linear combination of at most N wavelets which can occur at arbitrary positions or scales.

We define the error in approximating $f \in L_p(I\!R^d)$ by the elements of Σ_N^w by

$$\sigma_N^w(f)_{L_p(I\!R^d)} := \inf_{S\in\Sigma_N^w} \|f - S\|_{L_p(I\!R^d)}. \tag{4.1}$$

A fundamental result in wavelet approximation [17] is that the approximation error $\sigma_N^w(f)_{L_p(I\!R^d)}$ can be obtained up to a constant $C(d,\varphi)$ by greedy approximation. We describe this result only in the case $p = d^*$ although it holds for all $1 < p < \infty$ when one uses wavelets normalized in $L_p(I\!R^d)$. For each $N = 1, 2, \ldots$, we define the greedy approximant

$$\mathcal{G}_N(f) := \sum_{\lambda \in \Lambda_N(f)} c_\lambda(f)\psi_\lambda,$$

where $\Lambda_N(f)$ is the set of the N indices of the largest coefficients $c_\lambda(f)$, $\lambda \in \Delta$, in absolute value (ties in the size of these coefficients can be handled in an arbitrary way). Then, we have

Proposition 4.1 *For any $f \in L_{d^*}(I\!R^d)$, we have*

$$\|f - \mathcal{G}_N(f)\|_{L_{d^*}(I\!R^d)} \leq C(d, \varphi)\sigma_N(f)^w_{L_{d^*}(I\!R^d)}. \tag{4.2}$$

Proof: This result can be derived easily from a result of [17] where it is shown that (4.2) holds when $\mathcal{G}_N(f)$ is replaced by $\mathcal{G}_N^{L_{d^*}}(f)$. Here, $\mathcal{G}_N^{L_{d^*}}(f)$ is defined as above except that one starts with the wavelet coefficients normalized in L_{d^*} instead of BV.

Let $\{\psi_\lambda\}_{\lambda\in\Delta}$ be as usual the wavelet basis normalized for BV: $|\psi_\lambda|_{\mathrm{BV}} = 1$. For each $\lambda \in \Delta$, we choose ξ_λ such that $\|\xi_\lambda\psi_\lambda\|_{L_{d^*}(I\!R^d)} = 1$. The equivalence of the BV and L_{d^*} normalizations gives that $c_1 \leq \|\psi_\lambda\|_{L_{d^*}(I\!R^d)} \leq c_2$, with $c_1, c_2 > 0$ independent of λ. Because of the unconditionality of the wavelet basis for $L_{d^*}(I\!R^d)$, there are $C_1, C_2 > 0$ such that for any sequence of coefficients $\{a_\lambda\}_{\lambda\in\Delta}$,

$$C_1\|\sum_{\lambda\in\Delta} a_\lambda\psi_\lambda\|_{L_{d^*}(I\!R^d)} \leq \|\sum_{\lambda\in\Delta} a_\lambda\xi_\lambda\psi_\lambda\|_{L_{d^*}(I\!R^d)} \leq C_2\|\sum_{\lambda\in\Delta} a_\lambda\psi_\lambda\|_{L_{d^*}(I\!R^d)}. \tag{4.3}$$

Given any function $f = \sum_{\lambda\in\Delta} c_\lambda(f)\psi_\lambda$ in $L_{d^*}(I\!R^d)$, we let $g := \sum_{\lambda\in\Delta} c_\lambda(f)\xi_\lambda\psi_\lambda$ which by (4.3) is also in $L_{d^*}(I\!R^d)$. If $\mathcal{G}_N(f) = \sum_{\lambda\in\Delta} c_\lambda(f)\psi_\lambda$ then $\mathcal{G}_N^{L_{d^*}}(g) = \sum_{\lambda\in\Delta} c_\lambda(f)\xi_\lambda\psi_\lambda$. Hence, using (4.3), we have

$$C_1\|f - \mathcal{G}_N(f)\|_{L_{d^*}(I\!R^d)} \leq \|g - \mathcal{G}_N^{L_{d^*}}(g)\|_{L_{d^*}(I\!R^d)} \leq C(d, \varphi)\sigma_N^w(g)_{L_{d^*}(I\!R^d)}. \tag{4.4}$$

On the other hand, if $S = \sum_{\lambda\in\Lambda} a_\lambda\psi_\lambda$ is a best N-term approximation to f in $L_{d^*}(I\!R^d)$ then, using (4.3) again, we have

$$\sigma_n^w(g)_{L_{d^*}(I\!R^d)} \leq \|g - \sum_{\lambda\in\Lambda} a_\lambda\xi_\lambda\psi_\lambda\|_{L_{d^*}(I\!R^d)} \leq C_2\|f - S\|_{L_{d^*}(I\!R^d)} = C_2\sigma_N(f)_{L_{d^*}(I\!R^d)}. \tag{4.5}$$

The estimates (4.4) and (4.5) combine to prove the proposition. □

We are interested in quantitative estimates for the approximation error $\sigma_N(f)_{L_{d^*}(I\!R^d)}$ whenever $f \in \mathrm{BV}(I\!R^d)$. This will be provided by the following lemma.

Lemma 4.2 *For any function $f \in \mathrm{BV}(I\!R^d)$ we have the estimate*

$$\sigma_N^w(f)_{L_{d^*}(I\!R^d)} \leq C(d, \varphi)N^{-1/d}|f|_{\mathrm{BV}(I\!R^d)}, \quad N = 1, 2, \ldots.$$

Proof: The set $\mathcal{A}^{\alpha}_{\infty}(L_p(I\!\!R^d))$ of functions $f \in L_p(I\!\!R^d)$ which satisfy

$$\sigma_N^w(f)_{L_p(I\!\!R^d)} \leq CN^{-\alpha} \tag{4.6}$$

is called an approximation space. The norm $\|f\|_{\mathcal{A}^{\alpha}_{\infty}(L_p(I\!\!R^d))}$ in this space is the smallest $C > 0$ for which (4.6) is valid. For $1 < p < \infty$, and $\alpha > 0$, it was proved in [5] that $f \in \mathcal{A}^{\alpha}_{\infty}(L_p(I\!\!R^d))$ if and only if the sequence $(\|(c_\lambda(f)\psi_\lambda\|_{L_p(I\!\!R^d)})_{\lambda\in\Delta}$ is in the space weak ℓ_τ (denoted by $w\ell_\tau$) with $\frac{1}{\tau} = \alpha + \frac{1}{p}$. Moreover, $\|f\|_{\mathcal{A}^{\alpha}_{\infty}(L_p(I\!\!R^d))}$ is equivalent to $\|(\|(c_\lambda(f)\psi_\lambda\|_{L_p(I\!\!R^d)})_{\lambda\in\Delta}\|_{w\ell_\tau}$. In the case of interest to us, we have $p = d^* = \frac{d}{d-1}$ and $\alpha = 1/d$ so that $\tau = 1$. It was shown in [4] (for the case of Haar wavelets) and in [6] (for general wavelets) that the wavelet coefficients of a BV function f are in weak ℓ_1 and satisfy

$$\|(\|(c_\lambda(f)\psi_\lambda)_{\lambda\in\Delta}\|_{L_{d^*}(I\!\!R^d)})\|_{w\ell_1} \leq C(\varphi, d)|f|_{\mathrm{BV}(I\!\!R^d)}.$$

Therefore, the lemma follows. $\square$

In the case of $\varphi = \chi_{[0,1]}$, the wavelets ψ_λ, $\lambda \in \Delta$, are the Haar wavelets and the elements in Σ_N^w are piecewise constant functions which take at most CN values. We shall now consider two other types of nonlinear approximation using piecewise constants which will be important for us later. For the first of these, let

$$\Sigma_N^c := \{\sum_{I\in\Lambda} c_I\chi_I : \ \#(\Lambda) \leq N\},$$

where $\Lambda \subset \mathcal{D}$ is a set of dyadic cubes and for each set S in $I\!\!R^d$, χ_S denotes the characteristic function of S. Note that we do not require that the cubes in Λ are disjoint. In analogy with (4.1), we define

$$\sigma_N^c(f)_{L_p(I\!\!R^d)} := \inf_{S\in\Sigma_N^c} \|f - S\|_{L_p(I\!\!R^d)}.$$

Since each Haar wavelet H_λ is a linear combination of at most 2^d characteristic functions of dyadic cubes, it follows that $\Sigma_N^w \subset \Sigma_{2^dN}^c$ and hence from Lemma 4.2, we have

$$\sigma_N^c(f)_{L_d^*(I\!\!R^d)} \leq C(d,\varphi)N^{-1/d}|f|_{\mathrm{BV}(I\!\!R^d)}, \quad N = 1, 2, \ldots.$$

Lastly, we shall consider approximation by dyadic rings. If I and $J \subset I$ are two distinct dyadic cubes (J maybe the empty set), then we define the dyadic ring $R = R(I,J)$ to be the set $R = I \setminus J$. Consider the nonlinear space

$$\Sigma_N^r := \{\sum_{R\in\mathcal{P}} c_R\chi_R : \ \#(\mathcal{P}) \leq N\},$$

where $\mathcal{P}$ is a family of disjoint rings (i.e. any two R in $\mathcal{P}$ are disjoint). Note that $\chi_R = \chi_I - \chi_J$, and therefore

$$\Sigma_N^r \subset \Sigma_{2N}^c.$$

In analogy with the approximation errors defined above for N-term approximation by wavelets and constants, we define

$$\sigma_N^r(f)_{L_p(\mathbb{R}^d)} := \inf_{S \in \Sigma_N^r} \|f - S\|_{L_p(\mathbb{R}^d)}.$$

The following lemma concerning approximation by the elements of Σ_N^r was proved in [4] for the case $d = 2$ and by [19] for the case of general d in [19] (see Proposition 18).

Lemma 4.3 *For any function $f \in \mathrm{BV}(\mathbb{R}^d)$ we have the estimate*

$$\sigma_N^r(f)_{L_{d^*}(\mathbb{R}^d)} \leq C(d, \varphi) N^{-1/d} |f|_{\mathrm{BV}(\mathbb{R}^d)}, \quad N = 1, 2, \dots. \tag{4.7}$$

This result was proved in [4, 19] for functions in $\mathrm{BV}([0,1]^d)$. However, we can deduce it for general functions in $\mathrm{BV}(\mathbb{R}^d)$ using the following argument which we will also apply later in similar settings. First, it is enough to prove this result for functions with compact support since it then follows for general f by a limiting argument (see the definition of $\mathrm{BV}(\mathbb{R}^d)$ given in (2.1) and Remark 2.1). Suppose then that f is supported on $Q_k := [-2^{k-1}, 2^{k-1}]^d$ for some $k \geq 1$. We consider the mapping $\eta(x) := 2^k(x - e/2)$ where $e := (1, 1, \dots, 1) \in \mathbb{Z}^d$. Then η, which is composed of a shift (by $e/2$) and then a dyadic dilation (by 2^k), maps $[0,1]^d$ onto Q_k. Moreover, η maps any dyadic cube properly contained in $[0,1]^d$ into a dyadic cube contained in Q_k. Now let $g := f(\eta)$ and apply the analogue of (4.7) for $[0,1]^d$ to g. This result gives a partition $\mathcal{P}$ with $\#(\mathcal{P}) \leq N$ and a function $S = \sum_{R \in \mathcal{P}} c_R \chi_R$, where the $R \in \mathcal{P}$ are all of the form $R = I - J$ with I, J dyadic subcubes of $[0,1]^d$. If one of these R has $I = [0,1]^d$ then we can replace this R by at most 2^d rings corresponding to each of the children of $[0,1]^d$ and in this way we can assume that any ring in $\mathcal{P}$ involves dyadic cubes with sidelength < 1. The function $S(\eta^{-1})$ is in Σ_{cN}^r. From the fact that S approximates g in $L_{d^*}([0,1]^d)$ to the accuracy $C(d, \varphi) N^{-1/d} |g|_{\mathrm{BV}([0,1]^d)}$ we deduce that $S(\eta^{-1})$ approximates f to the accuracy $C(d, \varphi) N^{-1/d} |f|_{\mathrm{BV}([0,1]^d)}$ (recall that the L_{d^*} and BV norms scale the same under dilation. This the gives (4.7).

Let us make one last observation about approximation using the elements of Σ_N^r. Given a locally integrable function f, for each measurable set $\Omega \subset \mathbb{R}^d$, we denote by f_Ω the average of f over Ω:

$$f_\Omega := \frac{1}{|\Omega|} \int_\Omega f(x)\, dx.$$

Lemma 4.4 *If $f \in \mathrm{BV}(\mathbb{R}^d)$, there is a collection $\mathcal{P}$ of disjoint rings R, such that $\#(\mathcal{P}) \leq N$ and the function*

$$\mathcal{R}_N(f) := \sum_{R \in \mathcal{P}} f_R \chi_R$$

satisfies

$$\|f - \mathcal{R}_N(f)\|_{L_{d^*}(\mathbb{R}^d)} \leq C(\varphi, d) N^{-1/d} |f|_{\mathrm{BV}(\mathbb{R}^d)}. \tag{4.8}$$

Proof: Let $S \in \Sigma_N^r$ satisfy

$$\|f - S\|_{L_{d^*}(\mathbb{R}^d)} \le C(\varphi, d) N^{-1/d} |f|_{\mathrm{BV}(\mathbb{R}^d)}. \tag{4.9}$$

The existence of such a function is guaranteed by (4.7). We can write $S = \sum_{R \in \mathcal{P}} c_R \chi_R$, where $\mathcal{P}$ is a collection of at most N disjoint rings. From the disjointness of the rings in $\mathcal{P}$, we have

$$\|f - \mathcal{R}_n(f)\|_{L_{d^*}(\mathbb{R}^d)}^{d^*} = \sum_{R \in \mathcal{P}} \|f - f_R\|_{L_{d^*}(R)}^{d^*} + \|f\|_{L_{d^*}(\mathcal{P}^c)}^{d^*}, \quad \mathcal{P}^c := \mathbb{R}^d \setminus (\cup_{R \in \mathcal{P}} R), \tag{4.10}$$

and

$$\|f - S\|_{L_{d^*}(\mathbb{R}^d)}^{d^*} = \sum_{R \in \mathcal{P}} \|f - c_R\|_{L_{d^*}(R)}^{d^*} + \|f\|_{L_{d^*}(\mathcal{P}^c)}^{d^*}. \tag{4.11}$$

On the other hand,

$$\|f - f_R\|_{L_{d^*}(R)} \le 2 \inf_{c \in \mathbb{R}} \|f - c\|_{L_{d^*}(R)} \le 2\|f - c_R\|_{L_{d^*}(R)}.$$

This follows from the fact that the mapping $f \to f_R$ is a norm one projector on $L_{d^*}(\mathbb{R}^d)$. When this is used in (4.10), then (4.11) and (4.9) prove (4.8). $\square$

5 Inverse inequalities

There are certain inequalities (called Bernstein inequalities) which are companion to the Jackson inequalities. It was shown in [4] (for the case $d = 2$) and [19] (for the case $d > 2$) that any $S \in \Sigma_N^c$ satisfies

$$|S|_{\mathrm{BV}(\mathbb{R}^d)} \le C(d) N^{1/d} \|S\|_{L_{d^*}(\mathbb{R}^d)}. \tag{5.1}$$

This inequality was proved when S was supported on $[0,1]^d$ in the above references. If $S \in \Sigma_N^c$, we can assume that *supp* $S \subset [-K, K]^d$ for some K and by dilation and shifts we can map $[-K, K]^d \to [0,1]^d$ and deduce the general case (5.1) from that for $[0,1]^d$.

¿From (5.1), it follows that the same Bernstein inequality holds when $S \in \Sigma_N^r$ or $S \in \Sigma_N^w$ when the wavelet is the Haar wavelet. It will follow from the results of this section that the Bernstein inequality also holds for Σ_N^w for general compactly supported wavelets. However, our more general goal is to prove a Bernstein inequality for functions that are a sum of elements from both Σ_N^w and Σ_N^c.

We begin with a local Bernstein inequality between $\mathrm{BV}(\mathbb{R}^d)$ and $L_{d^*}(\mathbb{R}^d)$ with $d^* = \frac{d}{d-1}$. For any $I \in \mathcal{D}_k$, we denote by I' a general set of the form $I \setminus \cup_{j=1}^{2^d} J_j$ where each J_j is a (possibly empty) subcube of the children I_j, $j = 1, \dots, 2^d$, of I.

Lemma 5.1 *For each $f \in \mathcal{S}_k$ and for each $I \in \mathcal{D}_k$ and any of the sets I' we have*

$$|f\chi_{I'}|_{\mathrm{BV}(\mathbb{R}^d)} \le C(\varphi, d) \|f\chi_{I'}\|_{L_{d^*}(\mathbb{R}^d)}.$$

Proof: First of all, by dilation and translation, we can assume $k = 0$ and that $I = [0,1]^d$. We fix one of the children I_j of I and denote by $I'_j := I_j \setminus J_j$. It is enough to show

$$|f\chi_{I'_j}|_{\mathrm{BV}(\mathbb{R}^d)} \leq c_0 \|f\chi_{I'_j}\|_{L_1(\mathbb{R}^d)}, \quad j = 1, \ldots, 2^d, \tag{5.2}$$

with c_0 a constant depending only on φ, d. Indeed, we have $\chi_{I'} = \sum_{j=1}^{2^d} \chi_{I'_j}$ and from (5.2)

$$\begin{aligned} |f\chi_{I'}|_{\mathrm{BV}(\mathbb{R}^d)} &\leq \sum_{j=1}^{2^d} |f\chi_{I'_j}|_{\mathrm{BV}(\mathbb{R}^d)} \leq c_0 \sum_{j=1}^{2^d} \|f\chi_{I'_j}\|_{L_1(\mathbb{R}^d)} \\ &= c_0 \|f\chi_{I'}\|_{L_1(\mathbb{R}^d)} \leq c_0 \|f\chi_{I'}\|_{L_{d^*}(\mathbb{R}^d)}, \end{aligned}$$

where the last inequality uses Hölder's inequality and $|I'| \leq 1$.

To prove (5.2), we let J be one of the I_j, fix J, and let $J' = J \setminus J_j$. The result for the other I_j will follow by translation. We first observe that since the space $\mathcal{S}_0$ has dimension $\leq C(\varphi, d)$ on J, we have (by equivalence of norms on a finite dimensional space) that

$$\|f\chi_J\|_{L_\infty(\mathbb{R}^d)} \leq c_1 \|f\chi_J\|_{L_1(\mathbb{R}^d)}, \quad f \in \mathcal{S}_0, \tag{5.3}$$

and

$$|f\chi_J|_{\mathrm{BV}(\mathbb{R}^d)} \leq \|f\chi_J\|_{\mathrm{BV}(\mathbb{R}^d)} \leq c_1 \|f\chi_J\|_{L_1(\mathbb{R}^d)}, \quad f \in \mathcal{S}_0, \tag{5.4}$$

with c_1 depending only on φ and d.

We consider two cases. The first is that J' is obtained from J by removing a cube with measure $\leq \delta$, where δ will be specified in a moment. In this case, we note that

$$\begin{aligned} \|f\chi_J\|_{L_1(\mathbb{R}^d)} &\leq \|f\chi_J\|_{L_1(J')} + |J \setminus J'| \cdot \|f\chi_J\|_{L_\infty(J)} \\ &\leq \|f\chi_J\|_{L_1(J')} + c_1 |J \setminus J'| \cdot \|f\chi_J\|_{L_1(\mathbb{R}^d)}. \end{aligned} \tag{5.5}$$

Now, we select $\delta := \frac{1}{2c_1}$. Then, whenever $|J \setminus J'| \leq \delta$, we have $c_1 |J \setminus J'| \leq \frac{1}{2}$, and therefore (5.5) gives

$$\|f\chi_J\|_{L_1(\mathbb{R}^d)} \leq 2\|f\chi_{J'}\|_{L_1(\mathbb{R}^d)}. \tag{5.6}$$

Next, we note that

$$\begin{aligned} |f\chi_{J'}|_{\mathrm{BV}(\mathbb{R}^d)} &\leq |f\chi_J|_{\mathrm{BV}(\mathbb{R}^d)} + \mathrm{meas}_{d-1}(J \setminus J') \|f\chi_J\|_{L_\infty(\mathbb{R}^d)} \\ &\leq c_1 (1 + \mathrm{meas}_{d-1}(J \setminus J')) \|f\chi_J\|_{L_1(\mathbb{R}^d)} \\ &\leq c_1 (1 + 2^{d+1}) \|f\chi_J\|_{L_1(\mathbb{R}^d)} \leq 2c_1 (1 + 2^{d+1}) \|f\chi_{J'}\|_{L_1(\mathbb{R}^d)}. \end{aligned}$$

In the above inequalities, we have used relations (5.3), (5.4), (5.6), and the fact that $\mathrm{meas}_{d-1}(J \setminus J') \leq 2^{d+1}$. Thus, we have proved (5.2) in the case $|J \setminus J'| \leq \delta$.

To complete the proof, we consider the case when the dyadic cube $J \setminus J'$ has measure $> \delta$. For each such J' we have (by equivalence of norms on the finite dimensional space $\mathcal{S}_0|_{J'}$, see for comparison (5.4))

$$\|f\chi_{J'}\|_{BV(\mathbb{R}^d)} \leq c(J', \varphi, d) \|f\chi_{J'}\|_{L_1(\mathbb{R}^d)}.$$

There is a finite number of such sets J' and therefore by enlarging the constant from the first case (if necessary), we obtain (5.2) for all J' in the second case as well. This proves (5.2) and as noted earlier proves the Lemma. $\square$

We shall utilize a construction given in [10]. Let Λ be any finite collection of dyadic cubes. Given $I \in \Lambda$, we define the set $B(I) = B(I, \Lambda)$ of maximal cubes in I:

$$B(I,\Lambda) := \{J \in \Lambda : J \subset I, J \neq I \text{ and if } J' \in \Lambda \text{ with } J' \subset I, J' \neq I, J' \cap J \neq, \emptyset \text{ then } J' \subseteq J\}.$$

The following Lemma was proved in [10].

Lemma 5.2 *If $\Lambda \subset \mathcal{D}$ is any finite collection of dyadic cubes, then there exists a set of dyadic cubes $\tilde{\Lambda}$ such that*

(i) $\Lambda \subset \tilde{\Lambda}$ *and* $\#(\tilde{\Lambda}) \leq 2^d \#(\Lambda)$,

(ii) *For each cube* $I \in \tilde{\Lambda}$, $\#(B(I,\tilde{\Lambda})) \leq 2^d$, *where the* $B(I,\tilde{\Lambda})$ *are defined relative to* $\tilde{\Lambda}$.

(iii) *each child of I contains at most one cube from $B(I,\tilde{\Lambda})$.*

Let us note that in [10] this lemma was proved for the case when the cubes in Λ are contained in $[0,1]^d$. However, we can deduce the lemma as stated above from this by using the following reasoning. It follows by shifts of dyadic cubes that the lemma is true if all of the dyadic cubes of Λ are contained in a single dyadic cube of sidelength one. In the general case given in the above lemma, we can by dilating (if necessary) assume that all dyadic cubes in Λ are contained in $[-1,1]^d$. We can then partition $\Lambda = \cup_{j=1}^{2^d} \Lambda_j$, where Λ_j, $j = 1, \ldots, 2^d$, is the set of cubes in Λ that are contained in I_j, where I_j is one of the 2^d dyadic cubes of sidelength one that make up $[-1,1]^d$. We apply the lemma (as stated in [10]) to each Λ_j to receive $\tilde{\Lambda}_j$. Then $\tilde{\Lambda} := \cup_{j=1}^{2^d} \tilde{\Lambda}_j$ satisfies the above lemma.

We introduce one final notation before stating the main result of this section. Given a dyadic cube $I \in \mathcal{D}$, let

$$S(I) := \{J \in \mathcal{D} : |J| = |I|, \ \operatorname{supp} \phi_I \cap J \neq \emptyset\}.$$

The cubes in $S(I)$ are called the *support cubes* of ϕ_I. It is clear that $\#(S(I)) \leq C(\varphi, d)$ because φ has compact support.

The following theorem is the main result of this section. It establishes a Bernstein inequality for hybrid linear combinations of scaling functions and characteristic functions of dyadic cubes.

Theorem 5.3 *If $\Lambda_1, \Lambda_2 \subset \mathcal{D}$ each has cardinality at most N (i.e. $\#(\Lambda_1), \#(\Lambda_2) \leq N$), then any function*

$$f = \sum_{K \in \Lambda_1} a_K \phi_K + \sum_{K \in \Lambda_2} b_K \chi_K, \tag{5.7}$$

satisfies

$$|f|_{BV(\mathbb{R}^d)} \leq C(\varphi, d) N^{1/d} \|f\|_{L_{d^*}(\mathbb{R}^d)}.$$

Proof: By dilating f (if necessary), we can assume that each of the functions ϕ_K and χ_K appearing in (5.7) are supported in $[-1,1]^d$. (Recall again that the BV and L_{d^*} norms scale the same under dilation.) Let I_j, $j=1,\ldots,2^d$, be the dyadic cubes of sidelength one that make up $[-1,1]^d$. We define

$$\Lambda := \left(\bigcup_{K\in\Lambda_1} S(K)\right) \cup \Lambda_2 \cup \{I_j : j=1,\ldots,2^d\}.$$

We now apply Lemma 5.2 and receive the set $\tilde{\Lambda}$ with $\#(\tilde{\Lambda}) \le C(\varphi,d)N$.

For each $I\in\tilde{\Lambda}$ we now define $I' := I \setminus \bigcup_{J\in B(I)} J$, where $B(I) = B(I,\tilde{\Lambda})$. We have $\bigcup_{I\in\tilde{\Lambda}} I' = [-1,1]^d$ and the sets I' are pairwise disjoint. Therefore

$$f = \sum_{I\in\tilde{\Lambda}} f\chi_{I'}. \tag{5.8}$$

Claim: *For any $I\in\tilde{\Lambda}$ with $I\in\mathcal{D}_k$, each summand appearing in the representation* (5.7) *is in $\mathcal{S}_k$ on I'.*

To prove this claim, we first consider any ϕ_K, $K\in\Lambda_1$, appearing in the first sum. If $|K|\ge|I|$ then $\phi_K\in\mathcal{S}_k$ and we have our claim for this term. In the case $|K|<|I|$ let J be any support cube of ϕ_K with $J\cap I\ne\emptyset$. Then $|J|=|K|$ and hence J is contained in one of the cubes of $B(I)$ and hence $J\cap I'=\emptyset$. Thus such a ϕ_K is zero on I'. Thus we have established our claim for terms appearing in the first summand.

We now consider an arbitrary term χ_K appearing in the second summand for which $K\cap I\ne\emptyset$. If $|K|<|I|$ then K is contained in one of the cubes in $B(I)$ which in turn means that χ_K is zero on I'. If $|K|\ge|I|$ then ϕ_K is identically one on I'. Since the constant functions are in $\mathcal{S}_k$, we have proved our claim for the terms in the second summand as well.

We can now complete the proof of the theorem by returning to (5.8). Because of the claim, we can apply Lemma 5.1 to each term in (5.8) and thereby obtain

$$|f|_{BV(\mathbb{R}^d)} \le \sum_{I\in\tilde{\Lambda}} |f\chi_{I'}|_{BV(\mathbb{R}^d)} \le C(\varphi,d)\sum_{I\in\tilde{\Lambda}} \|f\|_{L_{d^*}(\mathbb{R}^d)}.$$

On the other hand, from the Hölder inequality,

$$\begin{aligned}\sum_{I\in\tilde{\Lambda}} \|f\chi_{I'}\|_{L_{d^*}(\mathbb{R}^d)} &\le (\#(\tilde{\Lambda}))^{1/d}\left(\sum_{I\in\tilde{\Lambda}} \|f\chi_{I'}\|^{d^*}_{L_{d^*}(\mathbb{R}^d)}\right)^{1/d^*} \\ &= (\#(\tilde{\Lambda}))^{1/d}\|f\|_{L_{d^*}(\mathbb{R}^d)} \le C(\varphi,d)N^{1/d}\|f\|_{L_{d^*}(\mathbb{R}^d)},\end{aligned}$$

because the sets $I'\in\tilde{\Lambda}$ are disjoint. $\square$

Theorem 5.3 contains many Bernstein inequalities as a special case. These are summarized in the following Corollary

Corollary 5.4 *The Bernstein inequality*

$$|f|_{BV(\mathbb{R}^d)} \le C(\varphi,d)N^{1/d}\,\|f\|_{L_{d^*}(\mathbb{R}^d)}$$

is valid whenever
(i) $f\in\Sigma_N^w$, $f\in\Sigma_N^c$, $f\in\Sigma_N^r$.
(ii) $f\in\Sigma_N^w\oplus\Sigma_N^r$.

Proof: Indeed, in each of these situations f can be rewritten in the form (5.7) with each of the two sums in (5.7) having at most $C(\varphi, d)N$ terms. $\square$

6 Proof of Theorem 1.1

and Theorem 1.2 We can now prove Theorem 1.1. Given $f \in \mathrm{BV}(I\!R^d)$, let $\mathcal{R}_N(f) \in \Sigma_N^r$ be the function in Σ_N^r satisfying Lemma 4.4. We have

$$\begin{aligned} |\mathcal{G}_N(f)|_{BV(I\!R^d)} &\leq |\mathcal{G}_N(f) - \mathcal{R}_N(f)|_{BV(I\!R^d)} + |\mathcal{R}_N(f)|_{BV(I\!R^d)} \\ &\leq CN^{1/d} \|\mathcal{G}_N(f) - \mathcal{R}_N(f)\|_{L_{d^*}(I\!R^d)} + |\mathcal{R}_N(f)|_{BV(I\!R^d)}. \end{aligned} \tag{6.1}$$

In the last inequality we have used (ii) of Corollary 5.4 for the function $(\mathcal{G}_N(f) - \mathcal{R}_N(f))$. We estimate now the first term in (6.1) by

$$\begin{aligned} N^{1/d} \|\mathcal{G}_N(f) - \mathcal{R}_N(f)\|_{L_{d^*}(I\!R^d)} &\leq N^{1/d} \left(\|\mathcal{G}_N(f) - f\|_{L_{d^*}(I\!R^d)} + \|f - \mathcal{R}_N(f)\|_{L_{d^*}(I\!R^d)} \right) \\ &\leq C|f|_{BV(I\!R^d)}, \end{aligned} \tag{6.2}$$

where in the last inequality we have used (4.2) and Lemma 4.2 to estimate the first term and Lemma 4.4 to estimate the second term. It follows from Corollary 12 of [19] (see also [4] for the case $d = 2$) that

$$|\mathcal{R}_N(f)|_{BV(I\!R^d)} \leq C|f|_{BV(I\!R^d)}. \tag{6.3}$$

Here we have used our general arguments of dilation and shifts to deduce (6.3). Using the estimates (6.2) and (6.3) in (6.1) gives the desired estimate. $\square$

Theorem 1.2 can be proved exactly as Theorem 12 in [19].

7 Further discussion

We briefly discuss some further issues which will help put our results into perspective

7.1 The case $d = 1$

Theorem 1.1 does not hold in the case $d = 1$. Consider, for example, the function $f = \chi_{[0,1/3]}$ which is in $\mathrm{BV}([0,1])$. We take the Haar basis H_λ, $\lambda \in \Delta$, normalized in $\mathrm{BV}([0,1])$: $|\mathcal{H}_\lambda|_{\mathrm{BV}([0,1])} = 1$. This is the same as normalizing this basis in $L_\infty([0,1])$. For each dyadic level $k = 0, 1, \ldots$, there is exactly one Haar coefficient that is nonzero (it corresponds to the dyadic interval $I \in \mathcal{D}_k$ which contains $1/3$). This coefficient $c_\lambda(f)$ has absolute value $1/3$ so that $c_\lambda(f)H_\lambda(1/3) = \pm 1/3$. For any given N, we can take N of these intervals so that all of the numbers $c_\lambda(f)H_\lambda(1/3)$ have the same sign. Then, the function $\mathcal{G}_N(f)$ obtained by retaining exactly these N terms of the Haar expansion of f will have $\mathrm{BV}([0,1])$ norm $\geq N/3$. If one wants to avoid the question of choosing arbitrarily in the case of ties, then one can perturb these coefficients slightly.

7.2 Quasi-greedy bases

Let X be a Banach space and $\{b_\lambda\}_{\lambda\in\Delta}$ be a (Schauder) basis for X with $\|b_\lambda\|_X = 1$, for all $\lambda \in \Delta$. Each $f \in X$ has a unique basis expansion $f = \sum_{\lambda\in\Delta} c_\lambda(f)b_\lambda$. We define the greedy approximant $\mathcal{G}_N(f)$ as before:

$$\mathcal{G}_N(f) := \sum_{\lambda\in\Lambda_N(f)} c_\lambda(f)b_\lambda,$$

where $\Lambda_N(f)$ is the set of indicies corresponding to the N largest coefficients in absolute value (with ties handled in an arbitrary way).

The basis $\{b_\lambda\}$ for the space X is said to be *quasi-greedy* if $\|f - \mathcal{G}_N(f)\|_X \to 0$, $N \to \infty$. It is known that the Haar basis is not quasi-greedy for $L_1(\mathbb{R}^d)$ (see [13]). On the other hand, it follows from what we have proved in this paper, that the wavelet bases are quasi-greedy in $W^1(L_1(\mathbb{R}^d))$. Indeed, it was proved in [18] that a basis is quasi-greedy for X if and only if

$$\|\mathcal{G}_N(f)\|_X \le C\|f\|_X, \quad f \in X,$$

with $C > 0$ an absolute constant. From Theorem 1.1, we know that the wavelet bases satisfy

$$|\mathcal{G}_N(f)|_{W^1(L_1(\mathbb{R}^d))} \le C(\varphi, d)|f|_{W^1(L_1(\mathbb{R}^d))}. \tag{7.1}$$

We want to change from semi-norm to norm in (7.1) which we can accomplish as follows. Since the basis $\{\psi_\lambda\}_{\lambda\in\Delta}$ is normalized in $\mathrm{BV}(\mathbb{R}^d)$, it follows that

$$\|\psi_\lambda\|_{L_1(\mathbb{R}^d)} \le C2^{-k}, \quad |\lambda| = k.$$

Secondly, we have the embedding $W^1(L_1(\mathbb{R}^d)) \subset B^1_\infty(L_1(\mathbb{R}^d))$ and

$$\|f\|_{B^1_\infty(L_1(\mathbb{R}^d))} \le C(d)\|f\|_{W^1(L_1(\mathbb{R}^d))},$$

where $B^1_\infty(L_1(\mathbb{R}^d))$ is the Besov space whose norm is given by

$$\|f\|_{B^1_\infty(L_1(\mathbb{R}^d))} := \sup_{k\ge 0} \sum_{\lambda\in\mathcal{D}_k} |c_\lambda(f)|.$$

Therefore, taking any index set Λ (not necessarily a greedy selection), we have

$$\|\sum_{\lambda\in\Lambda} c_\lambda(f)\psi_\lambda\|_{L_1(\mathbb{R}^d)} \le \sum_{k=0}^{\infty} 2^{-k} \sum_{\lambda\in\Lambda\cap\mathcal{D}_k} |c_\lambda(f)| \le \|f\|_{W^1(L_1(\mathbb{R}^d))}.$$

Hence, we can add the $L_1(\mathbb{R}^d)$-norm of $\mathcal{G}_N(f)$ to the left side of (7.1) and replace the $W^1(L_1(\mathbb{R}^d))$ semi-norm of f by the $W^1(L_1(\mathbb{R}^d))$ norm and obtain that

$$\|\mathcal{G}_N(f)\|_{W^1(L_1(\mathbb{R}^d))} \le C(\varphi, d)\|f\|_{W^1(L_1(\mathbb{R}^d))}.$$

7.3 Thresholding

Continuing with our setting of a wavelet basis $\{\psi_\lambda\}_{\lambda\in\Delta}$ normalized for $\mathrm{BV}(I\!\!R^d)$, for each $\epsilon > 0$, we define the hard thresholding operator

$$T_\epsilon(f) := \sum_{\lambda\in\Lambda(f,\epsilon)} c_\lambda(f)\psi_\lambda,$$

where $\Lambda(f,\epsilon) := \{\lambda : |c_\lambda(f)| > \epsilon\}$. It follows from Theorem 1.1 that this operator is bounded on $\mathrm{BV}(I\!\!R^d)$:

$$|T_\epsilon(f)|_{\mathrm{BV}(I\!\!R^d)} \le C(\varphi,d)|f|_{\mathrm{BV}(I\!\!R^d)}, \quad f\in \mathrm{BV}(I\!\!R^d).$$

There is another version of thresholding (called *soft thresholding*) which is preferred in some problems of statistical optimization. To describe soft thresholding, we fix a function $\eta(t)$ defined on $[0,\infty)$ such that η is increasing and

$$\begin{aligned} 0\le \eta(t) \le 1 \quad &\text{for all} \quad t,\\ \eta(t)=0 \quad &\text{for} \quad 0\le t\le 1/2,\\ \eta(t)=1 \quad &\text{for} \quad t\ge 1. \end{aligned}$$

Given $\epsilon > 0$ and $f\in BV(I\!\!R^d)$, we define the soft thresholding operator T_ϵ^η by

$$T_\epsilon^\eta(f) := \sum_{\lambda\in\Delta} \eta(|c_\lambda(f)|/\epsilon)c_\lambda(f)\psi_\lambda.$$

Claim: *For each $\epsilon>0$, we have*

$$|T_\epsilon^\eta(f)|_{BV(I\!\!R^d)} \le C(\varphi,d)|f|_{BV(I\!\!R^d)}, \quad f\in \mathrm{BV}(I\!\!R^d).$$

Proof of Claim: We order the coefficients $c_\lambda := c_\lambda(f)$ of f in decreasing order as $|c_{\lambda_1}| \ge |c_{\lambda_2}| \ge \dots$. We fix integers $N_0 < N_1 < \ldots < N_s$ and numbers $1 =: \beta_0 > \beta_1 > \cdots > \beta_s > \beta_{s+1} := 1/2$ in such a way that

$$\begin{aligned} |c_{\lambda_j}| \ge \epsilon = \beta_0\epsilon \quad &\text{for} \quad j\le N_0,\\ |c_{\lambda_j}| \le \epsilon/2 = \beta_{s+1}\epsilon \quad &\text{for} \quad j> N_s,\\ |c_{\lambda_j}| = \beta_{i+1}\epsilon \quad &\text{for} \quad N_i<j\le N_{i+1},\ i=0,\ldots,s-1. \end{aligned}$$

One checks that

$$T_\epsilon^\eta(f) = \sum_{i=0}^s [\eta(\beta_i)-\eta(\beta_{i+1})]\mathcal{G}_{N_i}(f),$$

so from the triangle inequality we get

$$\begin{aligned} |T_\epsilon^\eta(f)|_{\mathrm{BV}(I\!\!R^d)} &\le \sum_{i=0}^s [\eta(\beta_i)-\eta(\beta_{i+1})]|\mathcal{G}_{N_i}(f)|_{\mathrm{BV}(I\!\!R^d)}\\ &\le C(\varphi,d)|f|_{\mathrm{BV}(I\!\!R^d)}\sum_{i=0}^s [\eta(\beta_i)-\eta(\beta_{i+1})]\\ &= C(\varphi,d)|f|_{\mathrm{BV}(I\!\!R^d)}. \end{aligned}$$

Note that the above claim and its proof, although stated for the space $\mathrm{BV}(I\!\!R^d)$, hold for any Banach space X (used in place of $\mathrm{BV}(I\!\!R^d)$) which has a quasi-greedy basis (used in place of $\{\psi_\lambda\}$).

7.4 The case of domains $\Omega \subset I\!R^d$

Versions of Theorem 1.1 remain valid for $\mathrm{BV}(\Omega)$ with Ω certain domains in $I\!R^d$. We briefly mention two of the typical settings.

For certain domains $\Omega \subset I\!R^d$, one can construct wavelet bases $\{\psi_\lambda\}_{\lambda\in\Delta}$ such that $\mathrm{supp}(\psi_\lambda) \subset \Omega$ for each $\lambda \in \Delta$. The ψ_λ whose support is sufficiently inside the interior of the domain are the usual wavelets on $I\!R^d$. Near the boundary, the ψ_λ have a different structure. The first examples of such constructions were made in [3] for an interval on $I\!R$. These constructions were then extended to certain multidimensional domains (such as polyhedral domains) (see [7]) and then ultimately to quite general domains in [1]. These constructed bases have the three main properties we need to prove Theorem 1.1. They are of compact support. The scaling functions on a given dyadic level form a partition of unity. The scaling functions and wavelets on a dyadic level k can be written as a linear combination of a fixed number of scaling functions at level $k+1$. Thus, an analogue of Theorem 1.1 is valid for such basis where now the $\mathrm{BV}(I\!R^d)$ norm is replaced by the $\mathrm{BV}(\Omega)$ norm.

The second setting applies to quite general domains $\Omega \subset I\!R^d$. For example, it is sufficient that Ω is a Lipschitz graph domain (a minimally smooth domain in the sense of Stein (see [16], p.180)). Any function in $\mathrm{BV}(\Omega)$ can be extended to a function Ef in $\mathrm{BV}(I\!R^d)$ satisfying

$$\|Ef\|_{\mathrm{BV}(I\!R^d)} \le C(\Omega)\|f\|_{\mathrm{BV}(\Omega)}.$$

Such extension theorems are typically proved for the space $W^1(L_1(\Omega))$ and then follow for $\mathrm{BV}(\Omega)$ by a limiting argument. We can expand Ef in a wavelet expansion

$$Ef = \sum_{\lambda\in\Delta} c_\lambda(Ef)\psi_\lambda.$$

This decompositions serves as a wavelet representation for f on Ω:

$$f = \sum_{\lambda\in\Delta(\Omega)} c_\lambda(Ef)\psi_\lambda,$$

where $\Delta(\Omega)$ is the set of all indicies $\lambda \in \Delta$ for which ψ_λ does not vanish identically on Ω.

Consider now the thresholding operator T_ϵ applied to f and Ef. Since $T_\epsilon(f) = T_\epsilon(Ef)$ on Ω, we deduce that

$$\begin{aligned} |T_\epsilon(f)|_{\mathrm{BV}(\Omega)} &= |T_\epsilon(Ef)|_{\mathrm{BV}(\Omega)} \le |T_\epsilon(Ef)|_{\mathrm{BV}(I\!R^d)} \\ &\le C(\varphi,d)|Ef|_{\mathrm{BV}(I\!R^d)} \le C(\varphi,d,\Omega)\|f\|_{\mathrm{BV}(\Omega)}. \end{aligned}$$

In general, we cannot replace the norm on the right by the semi-norm.

References

[1] A. Cohen, W. Dahmen, R. DeVore, *Multiscale decompositions on bounded domains in L_p, $(0 < p < \infty)$*, Trans. Amer. Math. Soc., **352** (2000), 3651-3685.

[2] A. Cohen, W. Dahmen, I. Daubechies, R. DeVore, *Harmonic analysis of the space BV*, Rev. Mat. Iberoamericana, **19** (2003), 235-263.

[3] A. Cohen, I. Daubechies, P. Vial, *Wavelets and fast wavelet transforms on an interval*, Appl. Comput. Harmon. Anal., **1** 1(1993), 54–81.

[4] A. Cohen, R. DeVore, P. Petrushev, H. Xu, *Nonlinear approximation and the space $BV(I\!R^2)$*, Amer. J. Math., **121** 3(1999), 587–628.

[5] A. Cohen, R. DeVore, R. Hochmuth, *Restricted nonlinear approximation*, Constr. Approx., **16** (2000), 85–113.

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

[7] S. Dahlke, W. Dahmen, R. Hochmuth, R. Schneider, *Stable Multiscale Bases and Local Error Estimation for Elliptic Problems*, Appl. Numer. Math., **23** 1(1997), 21-48.

[8] I. Daubechies, Ten Lectures on Wavelets, CBMS-NSF Regional Coference Series in Applied Mathematics, SIAM, Philadelphia, 1992.

[9] R. DeVore, *Nonlinear approximation*, Acta Numer., **7** (1998), 51–150.

[10] R. DeVore, V. Popov, *Interpolation of Besov spaces*, Trans, Amer. Math. Soc., **305** 1(1988), 397–414.

[11] R. DeVore, B. Jawerth, B. Lucier, *Image compression through transform coding*, IEEE Proc. Inform. Theory, **38** (1992), 719–746.

[12] R. DeVore, G. Lorentz, *Conctructive Approximation*, Springer–Verlag, Berlin-New York, 1993.

[13] S. Konyagin, V. Temlyakov, *Greedy approximation with regard to bases and general minimal systems*, Serdica Math. J., **28** (2002), 305-328.

[14] Y. Meyer, *Ondelettes et Opérateurs*, Hemann, Paris, 1990.

[15] Y. Meyer, *Oscillating patterns in image processing and in some nonlinear evolution equations*, The Fifteenth Dean Jacqueline B. Lewis Memorial Lectures, Rutgers University, 2001.

[16] E. Stein, *Singular Integrals and Differentiability Properties of Functions*, Princeton University Press, Princeton, N.J., 1970.

[17] V. Temlyakov, *The best m-term approximation and greedy algorithms*, Adv. in Comput. Math., **8** 2(1998), 249–265.

[18] P. Wojtaszczyk, *Greedy algorithm for general biorthogonal systems*, J. Approx. Theory, **107** 2(2000), 293–314.

[19] P. Wojtaszczyk, *Projections and nonlinear approximation in the space $BV(I\!R^d)$*, Proc. London Math. Soc., **87** 3(2003), 471–497.

[20] W.P Ziemer, *Weakly Differentiable Functions*, Springer–Verlag, New York, 1989.

Paweł Bechler, Institute of Mathematics, Polish Academy of Science, ul. Sniadeckich 8, 00-950 Warsaw, Poland, E-mail: pbechler@impan.gov.pl

Ronald DeVore, Department of Mathematics, University of South Carolina, Columbia, SC 29208, USA, E-mail: devore@math.sc.edu

Anna Kamont, Institute of Mathematics, Polish Academy of Sciences, Branch in Gdansk, ul. Abrahama 18, 81-825 Sopot, Poland, E-mail: A.Kamont@impan.gda.pl

Guergana Petrova, Department of Mathematics, Texas A& M University College Station, TX 77843, USA, E-mail: gpetrova@math.tamu.edu

Przemysław Wojtaszczyk, Institute of Applied Mathematics and Mechanics, Warsaw University, ul. Banacha 2, 02-097 Warsaw, Poland, E-mail: pwojt@mimuw.edu.pl