SPARSE EVALUATION OF COMPOSITIONS OF FUNCTIONS USING MULTISCALE EXPANSIONS*

ALBERT COHEN[†], WOLFGANG DAHMEN[‡], AND RONALD DEVORE[§]

Abstract. This paper is concerned with the estimation and evaluation of wavelet coefficients of the *composition* $\mathcal{F} \circ u$ of two functions $\mathcal{F}$ and u from the wavelet coefficients of u. Our main objective is to show that certain sequence spaces that can be used to measure the sparsity of the arrays of wavelet coefficients are stable under a class of nonlinear mappings $\mathcal{F}$ that occur naturally, e.g., in nonlinear PDEs. We indicate how these results can be used to facilitate the sparse evaluation of arrays of wavelet coefficients of compositions at asymptotically optimal computational cost. Furthermore, the basic requirements are verified for several concrete choices of nonlinear mappings. These results are generalized to compositions by a multivariate map $\mathcal{F}$ of several functions $u_1, \ldots, u_n$ and their derivatives, i.e., $\mathcal{F}(D^{\alpha_1} u_1, \ldots, D^{\alpha_n} u_n)$.

Key words. nonlinear mappings, thresholding, tree structures, adaptive evaluation of nonlinear operators

AMS subject classifications. 35A15, 35A22, 41A60, 46B15, 46A45, 47H30

DOI. 10.1137/S0036141002412070

1. Introduction. This paper is concerned with the estimation and evaluation of the wavelet coefficients of a *composition* of two functions $\mathcal{F}$ and u, where u is given in terms of a wavelet expansion. Our interest in this subject stems from recent developments of *adaptive wavelet schemes* for the numerical solution of several types of initial or boundary value problems for PDEs. Such schemes typically rely on the sparsity of the wavelet representation of the solution allowing for data compression, as well as the ability to perform accurate numerical computations in the compressed representation. For initial value problems, *dynamically adaptive* schemes introduced in [20] require a reliable prediction of significant wavelet coefficients from the current state when progressing to the next time level. In the case of hyperbolic conservation laws, this question was first addressed in [19] and further discussed in [12]. Here one has to estimate the action of the nonlinear terms defining the convective fluxes on the current approximation in its multiscale representation. Another related example is the wavelet analysis of turbulent incompressible flows where such estimates are related to the energy transfer between different scales; see, e.g., [18] and [17]. For boundary value problems, adaptive wavelet schemes also require the tracking of the significant coefficients as the iterative solution process progresses; see, e.g., [1], [5], [9], and [10].

In all these examples, we are interested in the following general question: does composition with $\mathcal{F}$ preserve the *sparsity* of the wavelet coefficients of the function u? By the sparsity, we mean that only a quantifiable relatively small set of these

---

†Laboratoire d'Analyse Numerique, Universite Pierre et Marie Curie, 175 Rue du Chevaleret, 75013 Paris, France (cohen@ann.jussieu.fr, http://www.ann.jussieu.fr/~cohen/).

‡Institut für Geometrie und Praktische Mathematik, RWTH Aachen, 52056 Aachen, Germany (dahmen@igpm.rwth-aachen.de, http://www.igpm.rwth-aachen.de/~dahmen/).

§Department of Mathematics, University of South Carolina, Columbia, SC 29208 (devore@math.sc.edu, http://www.math.sc.edu/~devore/).

coefficients is needed to recover the underlying function (with accuracy measured in a given norm) to within some target accuracy. It is well known that sparsity of wavelet coefficients in this sense is closely related (in fact equivalent) to the *regularity* of the function with respect to certain scales of Besov spaces; see, e.g., [16]. Hence the above issue is closely connected with the question, how is the regularity of a given function u affected by the composition with some nonlinear function $\mathcal{F}$, or, more generally, given some regularity spaces $\mathcal{R}_i$, $i = 1, \ldots, m$, what is the image of $\prod_{i=1}^{m} \mathcal{R}_i$ under the mapping

$$(u_1(\cdot), \ldots, u_m(\cdot)) \to \mathcal{F}(\cdot, u_1(\cdot), \ldots, u_m(\cdot))?$$

This mapping is often referred to as a Nemytskij operator. The mapping properties of Nemytskij operators between Besov spaces were treated by several authors, and the reader is referred, e.g., to [3], [4], [22], and, for a detailed treatment, to the book by Runst and Sickel [21]. Sharp results are indeed available on the amount of smoothness which can be expected for $\mathcal{F}(u)$ given the smoothness of u, under fairly general assumptions on $\mathcal{F}$. Thus, in principle, in all cases covered by these results the sparsity of the wavelet coefficients of compositions can be predicted fairly well. However, these results tell us neither *which* coefficients of compositions $\mathcal{F}(u)$ are significant, based on knowledge about u, nor how to calculate them efficiently once they have been identified, which is a crucial issue in the perspective of numerical computations. The objective of the present paper is therefore also to develop concepts and tools for treating this latter problem.

Our paper is organized as follows. We present the problem formulation in section 2, which involves the wavelet discretization $\mathbf{F}$ of the mapping $\mathcal{F}$ as well as a notion of *tree structure* in the organization of wavelet coefficients. We prove in section 3 that this mapping preserves sparsity, under some general assumptions describing the stability and local action of $\mathbf{F}$ in the space-scale domain. We also present specific algorithms that construct sparse approximants with a prescribed accuracy ε at asymptotically optimal cost. This type of scheme is needed for the adaptive solution process of nonlinear operator equations; see [11]. We shall prove in section 4 the validity of the required assumptions for general local nonlinear mappings of subcritical type. Finally, the generalization of these results to compositions of the form $\mathcal{F}(D^{\alpha_1} u_1, \ldots, D^{\alpha_n} u_n)$ between a multivariate map $\mathcal{F}$ and the derivatives of several functions $u_1, \ldots, u_n$ is discussed in section 5.

2. Problem formulation.

2.1. Background and wavelet prerequisites.
To explain the relevant features of the problem it suffices to describe the following (simple) example in a little more detail. Consider the nonlinear boundary value problem of the form

$$(2.1) \qquad -\Delta u + \mathcal{F}(u) = f \ \text{ in } \ \Omega, \ \ u = 0 \ \text{ on } \ \partial\Omega,$$

where $\Omega \subset \mathbb{R}^d$ is some open bounded domain. The variational formulation of (2.1) in the space $H = H_0^1(\Omega)$ reads as follows: find $u \in H_0^1(\Omega)$ such that

$$(2.2) \qquad \int_\Omega \nabla u \cdot \nabla v + \int_\Omega \mathcal{F}(u)v = \int_\Omega fv$$

for all $v \in H_0^1(\Omega)$. Here $H_0^1(\Omega)$ is the usual Sobolev space of distributions with first order weak derivatives in $L_2(\Omega)$ vanishing on the boundary $\partial\Omega$ in the sense of

traces. (Of course, other boundary conditions may also be considered.) For (2.2) to be meaningful $\mathcal{F}$ should map $H_0^1(\Omega)$ into its dual $H^{-1}(\Omega)$. This is perhaps the simplest instance of a variational problem inducing a bijective mapping from a Hilbert space H onto its dual H'.

For more general problems, H is a product of closed subspaces H^t of Sobolev spaces determined, e.g., by homogeneous boundary conditions on part of the domain boundary; see, e.g., [10] for examples. For simplicity we will confine the subsequent discussion to the case of a single model space $H = H^t$ for some $t > 0$.

2.2. Wavelet discretization. As already explained, we are motivated by adaptive numerical methods based on discretizing the variational formulation (2.2) in a wavelet basis $\Psi = \{\psi_\lambda : \lambda \in \mathcal{J}\}$. The indices λ encode scale, spatial location, and the type of the wavelet ψ_λ. We will denote by $|\lambda|$ the *scale* associated with ψ_λ. We shall consider only compactly supported wavelets, i.e., the supports of the wavelets scale, as follows:

$$(2.3) \qquad S_\lambda := \operatorname{supp} \psi_\lambda, \quad c_0 2^{-|\lambda|} \leq \operatorname{diam} S_\lambda \leq C_0 2^{-|\lambda|},$$

with $c_0, C_0 > 0$ absolute constants. The index set $\mathcal{J}$ has the following structure $\mathcal{J} = \mathcal{J}_\phi \cup \mathcal{J}_\psi$, where $\mathcal{J}_\phi$ is finite and indexes the scaling functions on a fixed coarsest level j_0. $\mathcal{J}_\psi$ indexes the "true wavelets" ψ_λ with $|\lambda| > j_0$. From compactness of the supports we know that at each level, the set $\mathcal{J}_j := \{\lambda \in \mathcal{J} : |\lambda| = j\}$ is finite. In fact, one has $\#\mathcal{J}_j \sim 2^{jd}$ with constants depending on the underlying bounded domain.

As already explained in the introduction, our evaluation algorithms will rely on a tree structure associated to the set of wavelet indices. In the simplest case of a one-dimensional basis $\psi_\lambda = \psi_{j,k} = 2^{j/2}\psi(2^j \cdot -k)$, this structure is obvious: each index (j, k) has two children $(j + 1, 2k)$ and $(j + 1, 2k + 1)$. A similar tree structure can be associated to all available constructions of wavelet bases on a multidimensional domain: each index λ then has $m(\lambda) \geq 2$ children μ such that $|\mu| = |\lambda| + 1$, where $m(\lambda)$ might vary from one index to another but is uniformly bounded by some fixed K. We shall use the notation $\mu \prec \lambda$ in order to express that μ is a descendent of λ in the tree. Moreover, $\mu \preceq \lambda$ means that μ either is a descendent of λ or equals λ. We also have the property

$$(2.4) \qquad \mu \prec \lambda \Rightarrow S_\mu \subset S_\lambda.$$

One key feature is that Ψ is a *Riesz basis* of the relevant space $H = H^t$. This means that every $v \in H$ has a unique expansion $v = \sum v_\lambda \psi_\lambda$ and that there exist some constants c, C independent of v such that

$$(2.5) \qquad c\|(v_\lambda)_{\lambda \in \mathcal{J}}\| \leq \left\|\sum_{\lambda \in \mathcal{J}} v_\lambda \psi_\lambda\right\|_H \leq C\|(v_\lambda)_{\lambda \in \mathcal{J}}\|,$$

where $\|(v_\lambda)_{\lambda \in \mathcal{J}}\|^2 = \sum_{\lambda \in \mathcal{J}} |v_\lambda|^2$ denotes the $\ell_2(\mathcal{J})$-norm. In particular, the wavelets will always be assumed to be normalized in H, i.e., $\|\psi_\lambda\|_H = 1$. We abbreviate by

$$\mathbf{v} = (v_\lambda)_{\lambda \in \mathcal{J}}$$

the corresponding sequence of wavelet coefficients. Details on the construction of wavelet bases for Sobolev spaces of general domains can be found in [6], [7], [14].

Note that, by duality, (2.5) is equivalent to

$$(2.6) \qquad C^{-1}\|(\langle w, \psi_\lambda \rangle)_{\lambda \in \mathcal{J}}\| \leq \|w\|_{H'} \leq c^{-1}\|(\langle w, \psi_\lambda \rangle)_{\lambda \in \mathcal{J}}\|, \quad w \in H',$$

where $\langle \cdot, \cdot \rangle$ denotes the duality pairing between H and H'. Clearly the quantities $\langle w, \psi_\lambda \rangle$ are the coordinates of $w \in H'$ with respect to the *dual* Riesz basis $\tilde{\Psi}$ to Ψ.

Since, as pointed out above, the nonlinearity $\mathcal{F}$ is supposed to map H into H' we shall therefore describe $w = \mathcal{F}(u)$ by its inner product sequence $\mathbf{w} = (w_\lambda)_{\lambda \in \mathcal{J}}$ with

$$(2.7) \qquad\qquad w_\lambda = \langle w, \psi_\lambda \rangle, \quad \lambda \in \mathcal{J}.$$

We shall denote by $\mathbf{F}$ the corresponding *discrete* nonlinear map

$$(2.8) \qquad\qquad \mathbf{u} \mapsto \mathbf{w} = \mathbf{F}(\mathbf{u}) = (\langle \mathcal{F}(u), \psi_\lambda \rangle)_{\lambda \in \mathcal{J}}.$$

A key issue in the applications mentioned above can roughly be described as follows. Suppose that $u \in H$ can be approximated in the energy norm $\| \cdot \|_H$ within a tolerance ε by a linear combination of $N(\varepsilon, u)$ wavelets ψ_λ. What is the number $N(\varepsilon, \mathcal{F}(u))$ of dual wavelets needed to recover $\mathcal{F}(u)$ within tolerance ε? Note that, due to the *norm equivalences* (2.5) and (2.6), this can be restated as follows: Supposing that the wavelet coefficients $\mathbf{u}$ of $u \in H$ can be approximated in $\ell_2(\mathcal{J})$ with accuracy ε by a finitely supported vector involving only $N(\varepsilon, u)$ nonzero terms, how many entries of the sequence $\mathbf{F}(\mathbf{u})$ are needed to approximate $\mathbf{F}(\mathbf{u})$ in $\ell_2(\mathcal{J})$? Thus in the wavelet coordinate domain *all* approximations take place in $\ell_2(\mathcal{J})$. In brief, when does sparse approximability of $\mathbf{u}$ imply sparse approximability of $\mathbf{F}(\mathbf{u})$?

Questions of the above type are by now well understood for *linear* operators and their wavelet representations, as we shall now describe. In this context, the level of sparsity of $\mathbf{u}$ is measured by the *smallest* $\tau \leq 2$ such that $\mathbf{u} \in \ell_\tau^w(\mathcal{J})$. Here $\ell_\tau^w(\mathcal{J})$ is the collection of all $\mathbf{u} \in \ell_2(\mathcal{J})$ which satisfy

$$(2.9) \qquad\qquad \#\{\lambda \in \mathcal{J} : |u_\lambda| > \eta\} \leq C\eta^{-\tau}, \quad \eta > 0.$$

In fact, $\ell_\tau^w(\mathcal{J})$ is a (quasi-)normed linear space endowed with the norm

$$(2.10) \qquad\qquad \|\mathbf{u}\|_{\ell_\tau^w(\mathcal{J})} := \sup_{\eta > 0} \eta [\#\{\lambda \in \mathcal{J} : |u_\lambda| > \eta\}]^{1/\tau}.$$

An equivalent norm is given by the quantity

$$(2.11) \qquad\qquad \sup_{n > 0} n^{1/\tau} u_n^*,$$

where $(u_n^*)_{n > 0}$ is a nonincreasing rearrangement of $(|u_\lambda|)_{\lambda \in \mathcal{J}}$. Note that if $\tau < 2$, we have[1]

$$(2.12) \qquad\qquad \|\mathbf{u}\| \lesssim \|\mathbf{u}\|_{\ell_\tau^w(\mathcal{J})}.$$

Moreover, defining the error of *best N-term approximation* in $\ell_2(\mathcal{J})$

$$(2.13) \qquad\qquad \sigma_N(\mathbf{u}) := \inf_{\#\mathrm{supp}\,\mathbf{v} \leq N} \|\mathbf{u} - \mathbf{v}\| = \left(\sum_{n > N} |u_n^*|^2 \right)^{1/2},$$

one has the following characterization [9].

---

[1] Here and later we use the notation $a \lesssim b$ if $a \leq Cb$ with an absolute constant C independent of all parameters on which a, b depend.

PROPOSITION 2.1. *For* $\mathbf{u} \in \ell_2(\mathcal{J})$ *and* $s > 0$, *one has* $\sigma_N(\mathbf{u}) \lesssim N^{-s}$ *if and only if* $\mathbf{u} \in \ell_\tau^w(\mathcal{J})$ *with*

$$\frac{1}{\tau} = s + \frac{1}{2}. \tag{2.14}$$

Moreover,

$$\sigma_N(\mathbf{u}) \lesssim N^{-s} \|\mathbf{u}\|_{\ell_\tau^w(\mathcal{J})}. \tag{2.15}$$

Thus the smaller τ is, the fewer terms are needed to achieve a desired target accuracy for $\mathbf{u} \in \ell_\tau^w(\mathcal{J})$. In the case where $\mathbf{F}(\mathbf{u}) = \mathbf{A}\mathbf{u}$ is a linear operator bounded in $\ell_2(\mathcal{J})$, it is shown in [9] that this operator maps $\ell_\tau^w(\mathcal{J})$ into itself provided that it can be approximated by sparse matrices $\mathbf{A}_N$ with N entries per rows and columns at the rate $\|\mathbf{A} - \mathbf{A}_N\|_{\ell_2(\mathcal{J})} \lesssim N^{-r}$ for some $r > \frac{1}{\tau} - \frac{1}{2}$. Moreover, it is also shown how to practically build N-term approximations $\mathbf{w}_N$ of $\mathbf{w} = \mathbf{A}\mathbf{u}$, which fulfill the optimal rate $\|\mathbf{w}_N - \mathbf{w}\|_{\ell_2(\mathcal{J})} \lesssim N^{-s}$, from similar approximations of $\mathbf{u}$ at $\mathcal{O}(N)$ computational cost.

2.3. Tree structures and weak spaces. When dealing with nonlinear mappings, the following slight modification of these notions turns out to be appropriate. The approximants will be constrained by imposing a *tree structure* to the set of indices identifying the active coefficients. We shall say that a set $\mathcal{T} \subset \mathcal{J}$ is a *tree* if $\lambda \in \mathcal{T}$ implies $\mu \in \mathcal{T}$ whenever $\lambda \prec \mu$.

If the tree $\mathcal{T} \subset \mathcal{J}$ is finite, we define the set $\mathcal{L} = \mathcal{L}(\mathcal{T})$ of *outer leaves* as the set of those indices outside the tree such that their parent belongs to the tree

$$\mathcal{L} := \{\lambda \in \mathcal{J} : \lambda \notin \mathcal{T}, \ \lambda \prec \mu \implies \mu \in \mathcal{T}\}. \tag{2.16}$$

We shall make use of the following easily verifiable equivalence:

$$\#\mathcal{T} \sim \#\mathcal{L}, \tag{2.17}$$

where the constants depend only on K. Defining

$$\Gamma_\lambda := \{\mu \in \mathcal{J} : \mu \preceq \lambda\}, \tag{2.18}$$

the tree with root node λ, one easily verifies that

$$\mathcal{J} \setminus \mathcal{T} = \bigcup_{\lambda \in \mathcal{L}(\mathcal{T})} \Gamma_\lambda. \tag{2.19}$$

We are now interested in the approximation of $\mathbf{u}$ by an N-term approximation $\mathbf{v}$, where the support of $\mathbf{v}$ is assumed in addition to have a tree structure. A natural counterpart to classical best N-term approximation error, discussed in the previous section, is therefore given by redefining σ_N according to

$$\sigma_N(\mathbf{u}) := \inf\{\|u - v\| \ : \ \#(\text{supp}(\mathbf{v})) \leq N \text{ and } \text{supp}(\mathbf{v}) \text{ is a tree}\}. \tag{2.20}$$

We define $\mathcal{A}^s$ as the class of vectors $\mathbf{u}$ such that

$$\sigma_N(\mathbf{u}) \lesssim N^{-s} \tag{2.21}$$

and the corresponding quasi norm

$$\|\mathbf{u}\|_{\mathcal{A}^s} := \sup_{N>0} N^s \sigma_N(\mathbf{u}). \tag{2.22}$$

In contrast to best N-term approximation, the practical determination of a best N-term tree approximant is not a simple task. In particular, when $\mathbf{u}$ is a finite vector, the main difficulty is to build such an approximation without searching through all possible subtrees, which would result in exponential complexity in N. In [2], two algorithms have been proposed which construct near best trees in linear time, based on the evaluation of the local residuals

$$(2.23) \qquad \tilde{u}_\lambda := \left(\sum_{\mu \in \Gamma_\lambda} |u_\mu|^2 \right)^{1/2}.$$

Note that

$$(2.24) \qquad \|\mathbf{u} - \mathbf{u}|_{\mathcal{T}}\|^2 = \sum_{\lambda \in \mathcal{L}(\mathcal{T})} \tilde{u}_\lambda^2.$$

More precisely, given a tolerance ε, the algorithms proposed in [2] allow us to build a tree $\mathcal{T} = \mathcal{T}(\varepsilon, \mathbf{u})$ such that

$$(2.25) \qquad \|\mathbf{u} - \mathbf{u}|_{\mathcal{T}}\| \le \varepsilon$$

with the following property: whenever a tree $\tilde{\mathcal{T}}$ satisfies $\|\mathbf{u} - \mathbf{u}|_{\tilde{\mathcal{T}}}\| \le c\varepsilon$, then $\#(\mathcal{T}) \le C\#(\tilde{\mathcal{T}})$, where c, C are fixed constants independent of $\mathbf{u}$ and ε.

A simpler alternative to building tree approximants is to perform thresholding on the residual sequence $\tilde{u}_\lambda$. Indeed, one readily verifies that $\mu \preceq \lambda$ implies $\tilde{u}_\lambda \ge \tilde{u}_\mu$, i.e., for any $\eta > 0$ the set

$$(2.26) \qquad \mathcal{T}_\eta = \mathcal{T}_\eta(\mathbf{u}) := \{\lambda \ : \ |\tilde{u}_\lambda| > \eta\}$$

has tree structure. Thus, thresholding with respect to the modified sequences $\tilde{\mathbf{u}}$ creates trees. This motivates us to define

$$(2.27) \qquad {}_t\ell_\tau^w(\mathcal{J}) := \{\mathbf{u} \in \ell_2(\mathcal{J}) : \tilde{\mathbf{u}} \in \ell_\tau^w(\mathcal{J})\}, \quad \|\mathbf{u}\|_{{}_t\ell_\tau^w(\mathcal{J})} := \|\tilde{\mathbf{u}}\|_{\ell_\tau^w(\mathcal{J})}.$$

Clearly, we have $\|\mathbf{u}\|_{\ell_\tau^w(\mathcal{J})} \le \|\mathbf{u}\|_{{}_t\ell_\tau^w(\mathcal{J})}$ and

$$(2.28) \qquad \#\mathcal{T}_\eta(\mathbf{u}) \le \eta^{-\tau} \|\mathbf{u}\|_{{}_t\ell_\tau^w(\mathcal{J})}^\tau.$$

Therefore, the spaces ${}_t\ell_\tau^w(\mathcal{J})$ can also be used to quantify the sparseness of sequences subject to the tree structure constraint. In fact, one has the following counterpart to Proposition 2.1.

PROPOSITION 2.2. *Let* $\mathbf{u}_\eta := \mathbf{u}|_{\mathcal{T}_\eta}$. *Then* $\mathbf{u} \in {}_t\ell_\tau^w(\mathcal{J})$ *implies the error estimate*

$$(2.29) \qquad \|\mathbf{u} - \mathbf{u}_\eta\| \ \lesssim \ \eta^{1-\tau/2} \|\mathbf{u}\|_{{}_t\ell_\tau^w(\mathcal{J})}^{\tau/2} \ \lesssim \ [\#(\mathcal{T}_\eta)]^{-s} \|\mathbf{u}\|_{{}_t\ell_\tau^w(\mathcal{J})},$$

with $s = 1/\tau - 1/2$. *Therefore* ${}_t\ell_\tau^w(\mathcal{J})$ *is contained in* $\mathcal{A}^s$.

Proof. Let $\mathcal{L}_\eta := \mathcal{L}(\mathcal{T}_\eta)$ denote the set of outer leaves of the tree $\mathcal{T}_\eta$. By (2.19), (2.24) and using (2.28), one has

$$\|\mathbf{u} - \mathbf{u}_\eta\|^2 \ = \ \sum_{\lambda \notin \mathcal{T}_\eta} |u_\lambda|^2 = \sum_{\lambda \in \mathcal{L}_\eta} \tilde{u}_\lambda^2 \le \#\mathcal{L}_\eta \eta^2$$
$$(2.30) \qquad\qquad\qquad \lesssim \ \#\mathcal{T}_\eta \eta^2 \le \|\mathbf{u}\|_{{}_t\ell_\tau^w(\mathcal{J})}^\tau \eta^{2-\tau},$$

where we have used (2.17). This confirms the first estimate in (2.29). Since again by definitions (2.27) and (2.28), $\eta \leq \|\mathbf{u}\|_{t\ell_\tau^w(\mathcal{J})}(\#\mathcal{T}_\eta)^{-1/\tau}$, the second estimate follows from (2.30).   □

Note, however, that in the above proposition, we do not have a converse result which would state that the decay property $\|\mathbf{u} - \mathbf{u}_\eta\| \lesssim [\#(\mathcal{T}_\eta)]^{-s}$ implies that $\mathbf{u}$ is in $_t\ell_\tau^w(\mathcal{J})$. In particular $_t\ell_\tau^w(\mathcal{J})$ is strictly contained in $\mathcal{A}^s$.

Of course, the question arises which property of u implies that the array of wavelet coefficients $\mathbf{u}$ belongs to $_t\ell_\tau^w(\mathcal{J})$ and in turn to $\mathcal{A}^s$.

Remark 2.3. Let $H = H^t$. Then $u \in B_q^{t+sd}(L_{\tau'})$ implies $\mathbf{u} \in {}_t\ell_\tau^w(\mathcal{J})$ whenever $\frac{1}{\tau'} < \frac{1}{\tau} = s + \frac{1}{2}$ and $0 < q \leq \infty$.

Sketch of proof. It is enough to prove this for $B_\infty^{t+sd}(L_{\tau'})$ and $\tau < \tau' \leq 2$, because the remaining cases follow by embeddings. The condition $u \in B_\infty^{t+sd}(L_{\tau'})$ says that the H^t-normalized wavelet coefficients u_λ of u satisfy

$$\left(\sum_{|\lambda|=j} |u_\lambda|^{\tau'} \right)^{1/\tau'} \lesssim 2^{-jd\delta},$$

where $\delta := s + \frac{1}{2} - \frac{1}{\tau'} > 0$ is the discrepancy measuring the "distance" of $B_\infty^{t+sd}(L_{\tau'})$ from the critical embedding line. From this one derives also that $(\sum_{|\lambda|=j} |\tilde{u}_\lambda|^{\tau'})^{1/\tau'} \lesssim 2^{-jd\delta}$, $j \in \mathbb{N}$. This, in turn, implies that the function $\tilde{u}$ with wavelet coefficients $\tilde{\mathbf{u}}$ belongs to $B_\infty^{t+sd}(L_{\tau'})$. By Corollary 4.2 in [8], the best N-term approximation of $\tilde{u}$ in H^t has order N^{-s}. Therefore, by Proposition 2.1, $\tilde{\mathbf{u}} \in \ell_\tau^w(\mathcal{J})$, which, by (2.27) means that $\mathbf{u} \in {}_t\ell_\tau^w(\mathcal{J})$ as claimed.   □

We therefore have at our disposal two distinct notions of tree approximation rates expressed by the spaces $\mathcal{A}^s$ and $_t\ell_\tau^w(\mathcal{J})$. We can now restate the above questions in the following way:

- Does $\mathbf{F}$ map a sequence $\mathbf{u} \in \mathcal{X}$ into a sequence $\mathbf{w} = \mathbf{F}(\mathbf{u}) \in \mathcal{X}$ for $\mathcal{X} = \mathcal{A}^s$ or $_t\ell_\tau^w(\mathcal{J})$?
- Can we compute asymptotically optimal sparse approximations of $\mathbf{w} = \mathbf{F}(\mathbf{u})$ from asymptotically optimal sparse approximations of $\mathbf{u}$ in one of the two senses above?

Note that a positive answer to the second question gives a positive answer to the first question in a constructive way. Our next section will give precise answers to these questions for both $\mathcal{A}^s$ and $_t\ell_\tau^w(\mathcal{J})$.

3. Sparsity preserving discrete operators.

3.1. General assumptions. We shall use two general assumptions on the function $\mathbf{F}$. The first assumption expresses the fact that $\mathcal{F}$ is a *stable transformation* from H to H'.

Assumption 1. $\mathbf{F}$ is a Lipschitz map from ℓ_2 into itself. More precisely, we assume that we have

(3.1) $\|\mathbf{F}(\mathbf{u}) - \mathbf{F}(\mathbf{v})\| \leq C\|\mathbf{u} - \mathbf{v}\|, \quad \text{with} \ \ C = C(\sup\{\|\mathbf{u}\|, \|\mathbf{v}\|\}),$

where $x \mapsto C(x)$ is a positive nondecreasing function.

The fact that the constant C might grow with the norm of $\mathbf{u}$ and $\mathbf{v}$ accounts for the nonlinearity of the transformation. In the context of solving operator equations of the type (2.1), the norms of the arguments of $\mathbf{F}$ will remain bounded (by the $\|\cdot\|_H$-norm of the solution up to the achieved precision) so we can think of C as a constant.

We shall actually use a local version of this stability assumption which will be a direct consequence of (3.1) whenever the nonlinear function $\mathcal{F}$ is local in the physical space: if D is a subdomain of Ω, we have

$$(3.2) \qquad \|(\mathbf{F}(\mathbf{u}) - \mathbf{F}(\mathbf{v}))|_{\{\lambda : S_\lambda \subset D\}}\| \leq C\|(\mathbf{u} - \mathbf{v})|_{\{\lambda : S_\lambda \cap D \neq \emptyset\}}\|,$$

with C depending on $\|\mathbf{u}\|$ and $\|\mathbf{v}\|$ as for the global estimate.

The second assumption describes the *local action of* $\mathbf{F}$ in the space-scale domain of wavelet coefficients.

Assumption 2. If $\mathbf{w} = \mathbf{F}(\mathbf{u})$ for a finitely supported $\mathbf{u}$, we have the estimate

$$(3.3) \qquad |w_\lambda| \leq C \sup_{\mu \,:\, S_\mu \cap S_\lambda \neq \emptyset} |u_\mu| 2^{-\gamma(|\lambda| - |\mu|)} \text{ with } C = C(\|\mathbf{u}\|)$$

for all $\lambda \in \mathcal{J}_\psi$, where $\gamma > d/2$ and $x \mapsto C(x)$ is a positive nondecreasing function.

A typical value of γ is

$$(3.4) \qquad \gamma := r + t + d/2,$$

where r reflects the smoothness and order of vanishing moments of the wavelets, i.e., $\psi_\lambda \in C^r$ and $\int_\Omega x^m \psi_\lambda(x) dx = 0$ for $|m| = m_1 + \cdots + m_d < r$. We shall see in the next section that all these assumptions are fulfilled for a fairly general class of local composition operators.

3.2. Tree expansions. Given a tree $\mathcal{T}$, we shall make use of the following expansion process. Given any $\lambda \in \mathcal{J}$, we define $\Phi_0(\lambda) = \{\lambda\}$. If $\Phi_{k-1}(\lambda)$ has already been defined, then we define $\Phi_k(\lambda)$ as the set of all μ, $|\mu| = |\lambda| - k$ such that $S_\mu \cap S_{\mu'} \neq \emptyset$ for some $\mu' \in \Phi_{k-1}(\lambda)$. We define $\Phi(\lambda) := \cup_{k=0}^{|\lambda|} \Phi_k(\lambda)$. We then define the expansion $\tilde{\mathcal{T}}$ as

$$(3.5) \qquad \tilde{\mathcal{T}} := \cup_{\lambda \in \mathcal{T}} \Phi(\lambda).$$

Let us note that by construction $\tilde{\mathcal{T}}$ has the following property.

EXPANSION PROPERTY. *If $\mu \in \tilde{\mathcal{T}}$ and $\mu' \in \mathcal{J}$, then*

$$(3.6) \qquad \left. \begin{matrix} |\mu'| < |\mu| \\ S_{\mu'} \cap S_\mu \neq \emptyset \end{matrix} \right\} \implies \mu' \in \tilde{\mathcal{T}}.$$

The following lemma (see, e.g., [13], [15]) shows that $\tilde{\mathcal{T}}$ has size comparable to $\mathcal{T}$ and that the supports S_λ associated to the outer leaves $\mathcal{L}(\tilde{\mathcal{T}})$ do not overlap too much, a property that we shall use when dealing with the space $_t\ell^w_\tau(\mathcal{J})$.

LEMMA 3.1. *There exist constants C_1 and C_2 such that for any finite tree $\mathcal{T}$, we have the following.*

(i) $\#(\tilde{\mathcal{T}}) \leq C_1 \#(\mathcal{T})$.

(ii) *For all $\lambda \in \mathcal{L}(\tilde{\mathcal{T}})$ there exist at most C_2 indices $\mu \in \mathcal{L}(\tilde{\mathcal{T}})$ such that $S_\mu \cap S_\lambda \neq \emptyset$.*

Proof. We show first the existence of the constant C_1. To this end, it suffices to show that for each $\mu \in \tilde{\mathcal{T}}$ there exists a *reference element* $\lambda \in \mathcal{T}$ such that $|\lambda| = |\mu|$ and dist$(S_\lambda, S_\mu) \leq C_0 2^{-|\mu|}$, with C_0 the constant of (2.3). Now any $\mu \in \tilde{\mathcal{T}}$ is in $\Phi_k(\lambda')$ for some $\lambda' \in \mathcal{T}$. We prove by induction on k that there is such a reference

element. For $k = 0$, $\mu = \lambda'$ so we can take $\lambda = \lambda'$. Suppose that we have proven the existence of such a reference element for all $\mu' \in \Phi_{k-1}(\lambda')$ and let μ be an index that has been added in the construction of $\Phi_k(\lambda')$. By the definition of $\Phi_k(\lambda')$ there is a $\mu' \in \Phi_{k-1}(\lambda')$ such that $S_\mu \cap S_{\mu'} \neq \emptyset$. By our induction assumption, there is a reference element $\bar{\lambda} \in \mathcal{T}$, with $|\bar{\lambda}| = |\mu'|$, such that $\mathrm{dist}(S_{\mu'}, S_{\bar{\lambda}}) \leq C_0 2^{-|\mu'|}$. It follows that

$$\mathrm{dist}(S_\mu, S_{\bar{\lambda}}) \leq C_0 2^{-|\mu'|} + \mathrm{diam}(S_{\mu'}) \leq C_0 2^{-|\mu'|} + C_0 2^{-|\mu'|} = C_0 2^{-|\mu|}.$$

Hence, we can take the parent $\lambda \in \mathcal{T}$ of $\bar{\lambda}$ as our reference element for μ.

To confirm the existence of C_2, note that when $\nu, \mu \in \mathcal{L}(\tilde{\mathcal{T}})$ and $S_\nu \cap S_\mu \neq \emptyset$, then $||\nu| - |\mu|| \leq 1$. In fact, suppose that $|\nu| < |\mu| - 1$. Then, for the parent μ' of μ we have $S_\nu \cap S_{\mu'} \neq \emptyset$, since $S_\mu \subset S_{\mu'}$ according to (2.4). Since $\mu' \in \tilde{\mathcal{T}}$ and $|\mu'| > |\nu|$ we conclude $\nu \in \tilde{\mathcal{T}}$, which is a contradiction. This completes the proof. $\qquad\square$

3.3. The main result in the $\mathcal{A}^s$ case. We first consider the questions raised at the end of the previous section in the $\mathcal{A}^s$ case. Given a tolerance ε we wish to construct a near best tree for approximating $\mathbf{F}(\mathbf{u})$ with accuracy ε, based on the knowledge of $\mathbf{u}$. To this end, suppose that $\mathcal{T}(\varepsilon, \mathbf{u})$ is the near best tree obtained by one of the algorithms in [2] and which satisfies (2.25). For $j = 0, 1, \dots$, we define

$$(3.7) \qquad \mathcal{T}_j := \mathcal{T}\left(\frac{2^j \varepsilon}{1 + j}, \mathbf{u} \right),$$

and the corresponding expanded trees $\tilde{\mathcal{T}}_j$ according to the above procedure. By construction these trees are nested in the sense that $\tilde{\mathcal{T}}_j \subset \tilde{\mathcal{T}}_{j-1}$. We define the difference sets

$$(3.8) \qquad \Delta_j := \tilde{\mathcal{T}}_j \setminus \tilde{\mathcal{T}}_{j+1}.$$

In order to build a tree which will be adapted to $\mathbf{w} = \mathbf{F}(\mathbf{u})$, we introduce

$$(3.9) \qquad \alpha := \frac{2}{2\gamma - d} > 0,$$

where γ is the constant in (3.3), and for each $\mu \in \Delta_j$, we define the *influence set*

$$(3.10) \qquad \Lambda_{\varepsilon, \mu} := \{\lambda \; : \; S_\lambda \cap S_\mu \neq \emptyset \text{ and } |\lambda| \leq |\mu| + \alpha j\}.$$

We then define $\mathcal{T}$ by

$$(3.11) \qquad \mathcal{T} := \mathcal{J}_\phi \cup \left(\cup_{\mu \in \tilde{\mathcal{T}}_0} \Lambda_{\varepsilon, \mu} \right).$$

We notice that by construction $\mathcal{T}$ has the structure of an expanded tree. Note that an equivalent way of defining $\mathcal{T}$ would be

$$(3.12) \qquad \mathcal{T} := \mathcal{J}_\phi \cup \tilde{\mathcal{T}}_0 \cup \left(\cup_{\mu \in \tilde{\mathcal{T}}_0} \tilde{\Lambda}_{\varepsilon, \mu} \right),$$

with

$$(3.13) \qquad \tilde{\Lambda}_{\varepsilon, \mu} := \{\lambda \; : \; S_\lambda \cap S_\mu \neq \emptyset \text{ and } |\mu| \leq |\lambda| \leq |\mu| + \alpha j\}.$$

THEOREM 3.2. *Given any* $\mathbf{u}$ *and* $\mathcal{T}$ *defined by* (3.11)*, we have the error estimate*

$$(3.14) \qquad \|\mathbf{F}(\mathbf{u}) - \mathbf{F}(\mathbf{u})|_{\mathcal{T}}\| \lesssim \varepsilon.$$

Moreover, if $\mathbf{u} \in \mathcal{A}^s$ *for* $0 < s < \frac{2\gamma - d}{2d}$*, we have the estimate*

$$(3.15) \qquad \#(\mathcal{T}) \lesssim \|\mathbf{u}\|_{\mathcal{A}^s}^{1/s} \varepsilon^{-1/s} + \#(\mathcal{J}_\phi).$$

We therefore have $\mathbf{F}(\mathbf{u}) \in \mathcal{A}^s$ *and*

$$(3.16) \qquad \|\mathbf{F}(\mathbf{u})\|_{\mathcal{A}^s} \lesssim 1 + \|\mathbf{u}\|_{\mathcal{A}^s}.$$

The constants in these above inequalities depend only on $\|\mathbf{u}\|$*, the space dimension* d*, and the parameter* s*.*

Proof. In order to prove (3.14), we first notice that according to Assumption 1, we have

$$(3.17) \qquad \|\mathbf{F}(\mathbf{u}) - \mathbf{F}(\mathbf{u}|_{\tilde{\mathcal{T}}_0})\| \lesssim \varepsilon,$$

with a constant depending on $\|\mathbf{u}\|$. Since one therefore has the trivial estimate $\|(\mathbf{F}(\mathbf{u}) - \mathbf{F}(\mathbf{u}|_{\tilde{\mathcal{T}}_0})|_{\mathcal{T}}\| \lesssim \varepsilon$, it suffices to show that $\|\mathbf{w}_\varepsilon - \mathbf{w}_\varepsilon|_{\mathcal{T}}\| \lesssim \varepsilon$, where $\mathbf{w}_\varepsilon := \mathbf{F}(\mathbf{u}|_{\tilde{\mathcal{T}}_0}) = (w_{\varepsilon,\lambda})$. We then remark that Assumption 2 implies the cruder estimate

$$(3.18) \qquad |w_{\varepsilon,\lambda}|^2 \lesssim \sum_{\mu \in \tilde{\mathcal{T}}_0, S_\mu \cap S_\lambda \neq \emptyset} |u_\mu|^2 2^{-2\gamma(|\lambda| - |\mu|)},$$

and for $\lambda \notin \mathcal{T}$ and $\mu \in \Delta_j$ such that $S_\mu \cap S_\lambda \neq \emptyset$, we always have $|\lambda| - |\mu| \geq \alpha j$. Finally we notice that by definition

$$(3.19) \qquad \sum_{\mu \in \Delta_j} |u_\mu|^2 = \|\mathbf{u}|_{\tilde{\mathcal{T}}_j} - \mathbf{u}|_{\tilde{\mathcal{T}}_{j+1}}\|^2 \lesssim \frac{2^{2j}\varepsilon^2}{(1+j)^2}.$$

Combining these facts, we obtain

$$
\begin{aligned}
\|\mathbf{w}_\varepsilon - \mathbf{w}_\varepsilon|_{\mathcal{T}}\|^2 &= \sum_{\lambda \notin \mathcal{T}} |w_{\varepsilon,\lambda}|^2 \lesssim \sum_{\lambda \notin \mathcal{T}} \sum_{\mu \in \tilde{\mathcal{T}}_0, S_\mu \cap S_\lambda \neq \emptyset} |u_\mu|^2 2^{-2\gamma(|\lambda| - |\mu|)} \\
&= \sum_{\mu \in \tilde{\mathcal{T}}_0} |u_\mu|^2 \sum_{\lambda \notin \mathcal{T}, S_\mu \cap S_\lambda \neq \emptyset} 2^{-2\gamma(|\lambda| - |\mu|)} \\
&= \sum_{j \geq 0} \sum_{\mu \in \Delta_j} |u_\mu|^2 \sum_{\lambda \notin \mathcal{T}, S_\mu \cap S_\lambda \neq \emptyset} 2^{-2\gamma(|\lambda| - |\mu|)} \\
&\lesssim \sum_{j \geq 0} \sum_{\mu \in \Delta_j} |u_\mu|^2 \sum_{k \geq 0} 2^{(d-2\gamma)(\alpha j + k)} \lesssim \sum_{j \geq 0} 2^{(d-2\gamma)\alpha j} \sum_{\mu \in \Delta_j} |u_\mu|^2 \\
&\lesssim \sum_{j \geq 0} 2^{(d-2\gamma)\alpha j} \frac{2^{2j}\varepsilon^2}{(1+j)^2} = \sum_{j \geq 0} \frac{\varepsilon^2}{(1+j)^2} \lesssim \varepsilon^2.
\end{aligned}
$$

In order to prove (3.15), we notice that according to (3.12), for each $\mu \in \Delta_j$, we add at most $C2^{\alpha j d}$ indices to $\mathcal{J}_\phi \cup \tilde{\mathcal{T}}_0$ in the construction of $\mathcal{T}$. Therefore, we have

$$\#(\mathcal{T}) \leq \#(\mathcal{J}_\phi) + \#(\tilde{\mathcal{T}}_0) + C \sum_{j \geq 0} 2^{\alpha j d} \#(\Delta j)$$

$$\lesssim \ \#(\mathcal{J}_\phi) + \|\mathbf{u}\|_{\mathcal{A}^s}^{1/s} \varepsilon^{-1/s} \left(1 + \sum_{j \geq 0} 2^{(\alpha d - 1/s)j} (1+j)^{1/s} \right)$$

$$\lesssim \ \#(\mathcal{J}_\phi) + \|\mathbf{u}\|_{\mathcal{A}^s}^{1/s} \varepsilon^{-1/s},$$

where we have used the fact that $\alpha d - 1/s < 0$, which is equivalent to the restriction $s < \frac{2\gamma - d}{2d}$. The estimate (3.16) is then a direct consequence of (3.15). $\qquad\square$

Note that since we have used Assumptions 1 and 2 in the above proof, the constants in both estimates (3.15) and (3.16) are of the form $C(\|\mathbf{u}\|)$, where $x \mapsto C(x)$ is a positive nondecreasing function.

3.4. The main result in the $_t\ell_\tau^w(\mathcal{J})$ case. In order to deal with the $_t\ell_\tau^w(\mathcal{J})$ case, we shall build the tree for $\mathbf{F}(\mathbf{u})$ in a slightly different way. To this end, we fix $\eta > 0$ and we define the tree $\mathcal{T}_\eta$ obtained by thresholding the local residuals $\tilde{u}_\lambda$ at level η according to (2.26) and its expanded version $\tilde{\mathcal{T}}_\eta$.

Then for all $\mu \in \tilde{\mathcal{T}}_\eta$, we define the number $n(\mu)$ satisfying

$$(3.20) \qquad \eta 2^{\gamma n(\mu)} \leq |u_\mu| < \eta 2^{\gamma(n(\mu)+1)}.$$

We then define the *influence set*

$$(3.21) \qquad \Lambda_{\eta,\mu} := \{\lambda \ : \ S_\lambda \cap S_\mu \neq \emptyset \text{ and } |\lambda| \leq |\mu| + [n(\mu)]_+\}$$

and a tree for the approximation of $\mathbf{F}(\mathbf{u})$ by

$$(3.22) \qquad \mathcal{T} := \mathcal{J}_\phi \cup \left(\cup_{\mu \in \tilde{\mathcal{T}}_\eta} \Lambda_{\eta,\mu} \right).$$

We notice that by construction $\mathcal{T}$ has the structure of an expanded tree. Note that an equivalent way of defining $\mathcal{T}$ would be

$$(3.23) \qquad \mathcal{T} := \mathcal{J}_\phi \cup \tilde{\mathcal{T}}_\eta \cup \left(\cup_{\mu \in \tilde{\mathcal{T}}_\eta} \tilde{\Lambda}_{\eta,\mu} \right),$$

with

$$(3.24) \qquad \tilde{\Lambda}_{\eta,\mu} := \{\lambda \ : \ S_\lambda \cap S_\mu \neq \emptyset \text{ and } |\mu| \leq |\lambda| \leq |\mu| + [n(\mu)]_+\}.$$

THEOREM 3.3. *Given any* $\mathbf{u} \in \ell_2(\mathcal{J})$ *and* $\mathcal{T}$ *defined by (3.23), one has the coefficient size estimate*

$$(3.25) \qquad |\tilde{w}_\lambda| \ \lesssim \ \eta \ \text{ if } \lambda \notin \mathcal{T},$$

where the $\tilde{w}_\lambda$ *are defined for* $\mathbf{w} = \mathbf{F}(\mathbf{u})$ *according to (2.23). If in addition* $\mathbf{u} \in {}_t\ell_\tau^w(\mathcal{J})$ *for some* $d/\gamma < \tau < 2$, *then we have the cardinality estimate*

$$(3.26) \qquad \#(\mathcal{T}) \ \lesssim \ \|\mathbf{u}\|_{_t\ell_\tau^w(\mathcal{J})}^\tau \eta^{-\tau} + \#(\mathcal{J}_\phi).$$

Moreover, we have $\mathbf{F}(\mathbf{u}) \in {}_t\ell_\tau^w(\mathcal{J})$ *and*

$$(3.27) \qquad \|\mathbf{F}(\mathbf{u})\|_{_t\ell_\tau^w(\mathcal{J})} \ \lesssim \ 1 + \|\mathbf{u}\|_{_t\ell_\tau^w(\mathcal{J})}.$$

The constants in these above inequalities depend only on $\|\mathbf{u}\|$, *the space dimension* d, *and the parameter* τ *in the case of (3.26).*

Proof. In order to prove (3.25), we first consider the restricted vector $\mathbf{u}_\eta = \mathbf{u}|_{\tilde{\mathcal{T}}_\eta}$ and its image $\mathbf{w}_\eta := \mathbf{F}(\mathbf{u}_\eta) = (w_{\lambda,\eta})$. For $\lambda \notin \mathcal{T}$ and for all $\mu \in \tilde{\mathcal{T}}_\eta$ such that $S_\mu \cap S_\lambda \neq \emptyset$, we have by (3.21) the inequality $|\lambda| - |\mu| \geq [n(\mu)]_+$. Therefore, remarking that $\lambda \in \mathcal{J}_\psi$, the local action assumption (3.3) implies

$$(3.28) \qquad\qquad |w_{\lambda,\eta}| \lesssim \eta.$$

Moreover, if ν is such that $S_\nu \cap S_\lambda \neq \emptyset$ and $|\nu| = |\lambda| + l$, we also have $|\nu| - |\mu| \geq [n(\mu)]_+ + l$ and therefore, for each $\mu \in \tilde{\mathcal{T}}_\eta$, the better estimate

$$(3.29) \qquad\qquad |w_{\nu,\eta}| \lesssim 2^{-\gamma l}\eta.$$

It follows that

$$(3.30) \qquad\qquad |\tilde{w}_{\lambda,\eta}|^2 \lesssim \eta^2 \left(\sum_{l \geq 0} 2^{(d-2\gamma)l} \right) \lesssim \eta^2,$$

since by assumption $\gamma > d/2$.

Next, we remark that for $\lambda \in \mathcal{L}(\mathcal{T})$, we have

$$
\begin{aligned}
\left| |\tilde{w}_{\lambda,\eta}|^2 - |\tilde{w}_\lambda|^2 \right| &\leq (|\tilde{w}_{\lambda,\eta}| + |\tilde{w}_\lambda|)|\tilde{w}_{\lambda,\eta} - \tilde{w}_\lambda| \\
&\leq (2|\tilde{w}_{\lambda,\eta}| + |\tilde{w}_{\lambda,\eta} - \tilde{w}_\lambda|)|\tilde{w}_{\lambda,\eta} - \tilde{w}_\lambda| \\
&\lesssim (\eta + \|(\mathbf{w} - \mathbf{w}_\eta)|_{\Gamma_\lambda}\|)\|(\mathbf{w} - \mathbf{w}_\eta)|_{\Gamma_\lambda}\|.
\end{aligned}
$$
$$(3.31)$$

Now observe that, according to (3.2),

$$
\begin{aligned}
\|(\mathbf{w} - \mathbf{w}_\eta)|_{\Gamma_\lambda}\| &= \|(\mathbf{w} - \mathbf{w}_\eta)|_{\{\mu : S_\mu \subseteq S_\lambda\}}\| \\
&\lesssim \|(\mathbf{u} - \mathbf{u}_\eta)|_{\{\mu : S_\mu \cap S_\lambda \neq \emptyset\}}\| \\
&= \left(\sum_{\mu \notin \tilde{\mathcal{T}}_\eta, S_\mu \cap S_\lambda \neq \emptyset} |u_\mu|^2 \right)^{1/2} \\
&\leq \left(\sum_{\mu \in \mathcal{L}(\tilde{\mathcal{T}}_\eta), S_\mu \cap S_\lambda \neq \emptyset} |\tilde{u}_\mu|^2 \right)^{1/2} \lesssim \eta,
\end{aligned}
$$

where the last inequality involves the constant C_2 from Lemma 3.1. Combining this with (3.30) and (3.31), we obtain the size estimate (3.25).

To prove (3.26) we define the trees

$$(3.32) \qquad\qquad \tilde{\mathcal{T}}_j := \tilde{\mathcal{T}}_{\eta 2^{\gamma j}}.$$

From (2.28) and Lemma 3.1, we infer that $\mathbf{u} \in {}_t\ell_\tau^w(\mathcal{J})$ implies

$$(3.33) \qquad\qquad \#(\tilde{\mathcal{T}}_j) \lesssim \eta^{-\tau} 2^{-\gamma\tau j} \|\mathbf{u}\|_{{}_t\ell_\tau^w(\mathcal{J})}^\tau.$$

Writing

$$(3.34) \qquad\qquad \tilde{\Lambda}_\eta = \bigcup_{j \geq 0} \bigcup_{\mu \in \tilde{\mathcal{T}}_j \setminus \tilde{\mathcal{T}}_{j+1}} \Lambda_{\eta,\mu},$$

so that

$$\bigcup_{\mu\in\tilde{\mathcal{T}}_j\setminus\tilde{\mathcal{T}}_{j+1}} \Lambda_{\eta,\mu} \subseteq \tilde{\mathcal{T}}_j \cup \bigcup_{\mu\in\tilde{\mathcal{T}}_j\setminus\tilde{\mathcal{T}}_{j+1}} \{\lambda : S_\lambda \cap S_\mu \neq \emptyset, \; |\mu| < |\lambda| \leq |\mu| + [n(\mu)]_+\},$$

and remarking that, by (3.20) and (3.21), $n(\mu) = j$ for $\mu \in \tilde{\mathcal{T}}_j \setminus \tilde{\mathcal{T}}_{j+1}$, we obtain in view of (3.33)

$$\#\left(\bigcup_{\mu\in\tilde{\mathcal{T}}_j\setminus\tilde{\mathcal{T}}_{j+1}} \Lambda_{\eta,\mu}\right) \lesssim 2^{dj}\#(\tilde{\mathcal{T}}_j \setminus \tilde{\mathcal{T}}_{j+1}) + \#(\tilde{\mathcal{T}}_j)$$

$$(3.35) \qquad\qquad\qquad\qquad \lesssim \|\mathbf{u}\|_{t\ell_\tau^w(\mathcal{J})}^\tau \eta^{-\tau} 2^{(d-\gamma\tau)j}.$$

Since $d - \gamma\tau < 0$, by summing over $j \geq 0$, we obtain

$$(3.36) \qquad\qquad \#(\tilde{\Lambda}_\eta) \lesssim \|\mathbf{u}\|_{t\ell_\tau^w(\mathcal{J})}^\tau \eta^{-\tau},$$

and adding the cardinality of $\mathcal{J}_\phi$, we thus obtain (3.26).

In order to obtain the estimate (3.27), we first notice that (3.36) already indicates that we have the estimate

$$(3.37) \qquad\qquad \|(\tilde{w}_\lambda)_{\lambda\in\mathcal{J}_\psi}\|_{\ell_\tau^w(\mathcal{J})} \lesssim \|\mathbf{u}\|_{t\ell_\tau^w(\mathcal{J})}.$$

For the remaining indices $\lambda \in \mathcal{J}_\phi$, we can write

$$\|(\tilde{w}_\lambda)_{\lambda\in\mathcal{J}_\phi}\|_{\ell_\tau^w(\mathcal{J})} \leq \|(\tilde{w}_\lambda)_{\lambda\in\mathcal{J}_\phi}\|_{\ell_\tau} \leq [\#(\mathcal{J}_\phi)]^{1/\tau-1/2}\|(\tilde{w}_\lambda)_{\lambda\in\mathcal{J}_\phi}\| \lesssim \|\mathbf{w}\|,$$

so that we have $\|\mathbf{w}\|_{t\ell_\tau^w(\mathcal{J})} \lesssim \|\mathbf{u}\|_{t\ell_\tau^w(\mathcal{J})} + \|\mathbf{w}\|$. Since by Assumption 1,

$$(3.38) \qquad\qquad \|\mathbf{w}\| \lesssim \|\mathbf{F}(0)\| + \|\mathbf{u}\| \lesssim 1 + \|\mathbf{u}\| \lesssim 1 + \|\mathbf{u}\|_{t\ell_\tau^w(\mathcal{J})},$$

the estimate (3.27) follows. ☐

Note that since we have used Assumptions 1 and 2 in the above proof, the constants in both estimates (3.25) and (3.27) are of the form $C(\|\mathbf{u}\|)$, where $x \mapsto C(x)$ is a positive nondecreasing function.

Remark 3.1. The limitation $d/\gamma < \tau < 2$ in Theorem 3.3 is exactly equivalent to the limitation $0 < s < \frac{2\gamma-d}{2d}$ in Theorem 3.2 with $s = 1/2 - 1/\tau$.

Remark 3.2. In both Theorems 3.2 and 3.3, the presence of the constant term in the right side of (3.16) and (3.27), and of the $\#(\mathcal{J}_\phi)$ term on the right side of (3.15) and (3.26), can be avoided if the local action estimate (3.3) remains valid also for $\lambda \in \mathcal{J}_\phi$. The tree $\mathcal{T}$ is then constructed without systematic inclusion of $\mathcal{J}_\phi$, and the proofs of the new estimates are analogous. This is the case for certain classes of linear mappings; see Proposition 7.4 in [11].

3.5. Adaptive evaluation schemes. Adaptive wavelet schemes for variational problems of the type (2.2) rest on two conceptual steps. First (2.2) is formulated in wavelet coordinates as an equivalent problem over $\ell_2(\mathcal{J})$ as follows:

$$(3.39) \qquad\qquad \mathbf{Au} + \mathbf{F}(\mathbf{u}) = \mathbf{f},$$

where $\mathbf{A} = (\langle\nabla\psi_\lambda, \nabla\psi_\nu\rangle)_{\lambda,\nu\in\mathcal{J}}$ is the wavelet representation of Δ and $\mathbf{f} = (\langle f, \psi_\lambda\rangle)_{\lambda\in\mathcal{J}}$. The second step is to devise an iterative scheme for numerically solving (3.39). This

iteration requires the approximate evaluation of $\mathbf{A}\mathbf{u}^n$ and $\mathbf{F}(\mathbf{u}^n)$ with some dynamically updated tolerance, where $\mathbf{u}^n$ is the current finitely supported iterate. How to deal with the linear part $\mathbf{A}\mathbf{u}^n$ has been explained in [9]. The remaining task may therefore be formulated as follows: given a target accuracy $\varepsilon > 0$, and some finitely supported $\mathbf{v} \in \ell_2(\mathcal{J})$, compute $\mathbf{F}(\mathbf{v})$ with accuracy ε at a possibly moderate computational expense.

The way of tackling this task involves two steps : (i) identify an optimal tree such that the restriction of $\mathbf{F}(\mathbf{v})$ to such a tree can be predicted to approximate $\mathbf{F}(\mathbf{v})$ at the desired accuracy, and (ii) numerically compute the coordinates of $\mathbf{F}(\mathbf{v})$ restricted to this tree. We shall not engage (ii) except to mention the paper [15], which treats this topic once (i) has been solved. On the other hand, the results in sections 3.3 and 3.4 allow us to solve (i). We shall again distinguish between near best tree approximation and thresholding the local residual.

When working with near best tree approximation as in section 3.3, Theorem 3.2 provides us with a construction of the required tree, according to the following algorithm.

ALGORITHM EV1.   *Given the inputs $\varepsilon > 0$ and $\mathbf{v}$ with finite support do the following:*

Step 1. *Invoke the algorithm in* [2] *to compute the trees*

$$(3.40) \qquad \mathcal{T}_j := \mathcal{T}\left(\frac{2^j \varepsilon}{C_0(j+1)}, \mathbf{v}\right),$$

where $C_0 = C_0(\|\mathbf{v}\|)$ is the constant involved in (3.14), *for $j = 0, \dots, J$, and stop for the smallest J such that $\mathcal{T}_J$ is empty (we always have $J \lesssim \log_2(\|\mathbf{v}\|/\varepsilon)$).*

Step 2.   *Derive the expanded trees $\tilde{\mathcal{T}}_j$, the layers Δ_j, and the outcome tree $\mathcal{T}$ according to* (3.11).

The following theorem summarizes the properties of Algorithm EV1.

THEOREM 3.4.   *Given the inputs $\epsilon > 0$, a nonlinear function F such that $\mathbf{F}$ satisfies Assumptions* 1 *and* 2, *and a finitely supported vector $\mathbf{v}$, the output tree $\mathcal{T}$ has the following properties:*

P1: $\|\mathbf{F}(\mathbf{v}) - \mathbf{F}(\mathbf{v})|_{\mathcal{T}}\| \leq \epsilon.$

P2: *For any $0 < s < \frac{2\gamma - d}{2d}$ (see Theorem* 3.2*),*

$$(3.41) \qquad \#(\mathcal{T}) \leq C\|\mathbf{v}\|_{\mathcal{A}^s}^{1/s}\epsilon^{-1/s} + \#(\mathcal{J}_\phi) =: N_\varepsilon,$$

with C a constant depending only on the constants appearing in Theorem 3.2.

P3: *Moreover, the number of computations needed to find $\mathcal{T}$ is bounded by $C(N_\varepsilon + \#\mathcal{T}(\mathbf{v}))$, where N_ε is the right-hand side of* (3.41) *and $\mathcal{T}(\mathbf{v})$ is the smallest tree containing* supp $\mathbf{v}$.

Proof. Properties P1 and P2 directly follow from Theorem 3.2. Property P3 is a consequence of the optimality property of the algorithm in [2]. First, the application of this algorithm requires the array $\tilde{\mathbf{v}}$ whose entries serve as local error terms. $\tilde{\mathbf{v}}$ can be computed by summing the squares of the v_λ starting from the leaves of $\mathcal{T}(\mathbf{v})$ towards the roots. The number of operations remains proportional to $\#(\mathcal{T}(\mathbf{v}))$. Moreover, the additional cost of computing each tree $\mathcal{T}_j$, making use of the previously computed tree, is bounded by $C\#(\mathcal{T}_j)$ and therefore by $C\|\mathbf{v}\|_{\mathcal{A}^s}^{1/s}\epsilon^{-1/s}2^{-j/s}(j+1)^{1/s}$. Summing over j we obtain that the total cost remains bounded by the right side of (3.41) and $\#(\mathcal{T}(\mathbf{v}))$.   □

When working with trees obtained by thresholding the local residuals as in section 3.4, Theorem 3.2 does not provide us with a direct construction of the required tree.

In order to build this tree, we first note that if $\mathcal{T}$ is the tree obtained by (3.23) for some given threshold η, and if $\mathcal{L}$ is the set of its outer leaves, we have the estimate

$$(3.42) \qquad\qquad \|\mathbf{v} - \mathbf{F}(\mathbf{v})\|^2 \le C_0^2 \#(\mathcal{L})\eta^2,$$

where C_0 is the constant of the inequality (3.25). Therefore we can invoke the following algorithm based on geometrically updated thresholds $\eta_j = 2^{-j}$.

ALGORITHM EV2. *Given the inputs $\epsilon > 0$ and $\mathbf{v}$ with finite support, initialize with $j = 0$ and do the following:*

Step 1. Given j compute the predicted tree $\mathcal{T}$ and its outer leaves $\mathcal{L}$ for $\eta = \eta_j = 2^{-j}$. Compute the corresponding error estimator $\varepsilon_0 = C_0^2 \#(\mathcal{L})\eta_j^2$ and proceed to Step 2.

Step 2. If $\epsilon_j \le \epsilon$, terminate the algorithm and take $\mathcal{T}$ as output. If $\epsilon_j > \epsilon$, replace j by $j + 1$ and return to Step 1.

The following theorem summarizes the properties of Algorithm EV2.

THEOREM 3.5. *Given the inputs $\epsilon > 0$, a nonlinear function F such that $\mathbf{F}$ satisfies Assumptions 1 and 2, and a finitely supported vector $\mathbf{v}$, the output tree $\mathcal{T}$ has the following properties:*

P1: $\|\mathbf{F}(\mathbf{v}) - \mathbf{F}(\mathbf{v})|_{\mathcal{T}}\| \le \epsilon$.

P2: *For any $d/\gamma < \tau < 2$ (see Theorem 3.3),*

$$(3.43) \qquad\qquad \#(\mathcal{T}) \le C\|\mathbf{v}\|_{t\ell_\tau^w(\mathcal{J})}^{1/s}\epsilon^{-1/s} + \#(\mathcal{J}_\phi),$$

with C a constant depending only on the constants appearing in Theorem 3.3.

P3: *Moreover, the number of computations needed to find $\mathcal{T}$ is also bounded by the right side of (3.43) plus $\#(\mathcal{T}(\mathbf{v}))$, where $\mathcal{T}(\mathbf{v})$ denotes again the smallest tree containing the support of $\mathbf{v}$.*

Proof. Since the vector $\mathbf{v}$ is finite, it belongs to all $_t\ell_\tau^w(\mathcal{J})$. From (3.26) of Theorem 3.3, we have at step j

$$(3.44) \qquad\qquad \#(\mathcal{T}) \lesssim 2^{j\tau}\|\mathbf{v}\|_{t\ell_\tau^w(\mathcal{J})}^\tau + \#(\mathcal{J}_\phi),$$

from which it follows that ϵ_j tends to 0 as $j \to \infty$. Therefore, the algorithm must terminate at some finite value j^*. It follows from (3.42) that for the tree $\mathcal{T}$ obtained at step j^*, we have

$$(3.45) \qquad\qquad \|\mathbf{F}(\mathbf{v}) - \mathbf{F}(\mathbf{v})|_{\mathcal{T}}\|^2 \le \epsilon^2,$$

which proves P1.

In order to prove P2, we start with (3.26), which gives for $\mathcal{T}$ obtained at step j^*

$$(3.46) \quad \#(\mathcal{T}) \lesssim \|\mathbf{v}\|_{t\ell_\tau^w(\mathcal{J})}^\tau \eta_{j^*}^{-\tau} + \#(\mathcal{J}_\phi) = C_0^\tau \|\mathbf{v}\|_{t\ell_\tau^w(\mathcal{J})}^\tau \epsilon_{j^*}^{-\tau}(\#(\mathcal{L}))^{\tau/2} + \#(\mathcal{J}_\phi).$$

We continue under the assumption that the first term on the far right side of (3.46) is bigger than the second, since otherwise we are done. We recall now that $\#(\mathcal{T}) \sim \#(\mathcal{L})$ and that $\epsilon \le \epsilon_{j^*-1} \le 4\varepsilon_{j^*}$ because the sets $\mathcal{L}$ increase when j increases. Using this information back in (3.46) gives

$$(3.47) \qquad\qquad \#(\mathcal{T})^{1-\tau/2} \lesssim \|\mathbf{v}\|_{t\ell_\tau^w(\mathcal{J})}^\tau \epsilon^{-\tau}.$$

Therefore P2 follows by raising both sides of (3.47) to the power $1/(1 - \tau/2)$ because $s = 1/\tau - 1/2$, and hence $1 - \tau/2 = s\tau$.

Finally, to prove P3 we note that the thresholding requires the knowledge of $\tilde{\mathbf{v}}$. The same arguments as in the proof of P3 in Theorem 3.4 yield the contribution of $\#(\mathcal{T}(\mathbf{v}))$ to the operations count. Moreover, once $\tilde{\mathbf{v}}$ is known, the number of additional computations used in the algorithm is bounded by $C(\|\mathbf{v}\|^\tau_{t\ell^w_\tau(\mathcal{J})}\eta_0^{-\tau}+\#(\mathcal{J}_\phi))$ at iteration 0. Given the tree at stage j, which allows one to avoid corresponding comparisons, the additional work needed to determine the tree at stage $j+1$ is bounded by $C\|\mathbf{v}\|^\tau_{t\ell^w_\tau(\mathcal{J})}\eta_j^{-\tau}$. Summing up over j, we derive P3.   □

4. Verification of the basic assumptions. We shall show in this section that Assumptions 1 and 2 hold for nonlinear mappings of the form $\mathcal{F}(u)(x) = \mathcal{F}(u(x))$, where $\mathcal{F}$ is a univariate function which satisfies growth conditions at infinity of the type

$$(4.1) \qquad |\mathcal{F}^{(n)}(x)| \leq C(1+|x|)^{[p-n]_+}, \quad x \in \mathbb{R}, \ n = 0, 1, \ldots, n^*,$$

for some $p \geq 0$ and n^* a positive integer. Clearly $\mathcal{F}(u) = u^p$ is of this type for all n^* if p is an integer, and with n^* the integer part of p otherwise.

4.1. Verification of Assumption 1. The verification of Assumption 1 is a classical result in the case where $H = H^t(\Omega)$, $t \geq 0$, or when H is a closed subspace of $H^t(\Omega)$ determined, e.g., by homogeneous boundary conditions, such as $H_0^t(\Omega)$ (the closure in the $\|\cdot\|_{H^t}$-norm of smooth functions with compact support in the open bounded domain Ω).

PROPOSITION 4.1. *Assume that $\mathcal{F}$ satisfies* (4.1) *for some $p \geq 0$ and $n^* \geq 0$. Then $\mathcal{F}$ maps H to H' under the restriction*

$$(4.2) \qquad 0 \leq p \leq p^* := \frac{d+2t}{d-2t}$$

when $t < d/2$, and with no restriction otherwise. If in addition $n^* \geq 1$, then we also have under the same restriction

$$(4.3) \qquad \|\mathcal{F}(u) - \mathcal{F}(v)\|_{H'} \leq C\|u - v\|_H,$$

where $C = C(\max\{\|u\|_H, \|v\|_H\})$ and $x \rightarrow C(x)$ is nondecreasing, and therefore Assumption 1 holds.

Proof. For $u \in H$ and $\varphi \in H$, we write

$$(4.4) \qquad |\langle \mathcal{F}(u), \varphi \rangle| \leq C\left[\int_\Omega |\varphi| + \int_\Omega |\varphi||u|^p\right].$$

The first term is bounded according to

$$(4.5) \qquad \int_\Omega |\varphi| \leq |\Omega|^{1/2}\|\varphi\|_{L_2} \leq |\Omega|^{1/2}\|\varphi\|_H.$$

For the second term, we use Hölder's inequality to obtain

$$(4.6) \qquad \int_\Omega |\varphi||u|^p \leq \|\varphi\|_{L_q}\|u\|^p_{L_{pq'}},$$

where $\frac{1}{q} + \frac{1}{q'} = 1$. Taking q such that $q = pq' = pq/(q-1)$, i.e., $q = p+1$, this gives

$$(4.7) \qquad \int_\Omega |\varphi||u|^p \leq \|\varphi\|_{L_{p+1}}\|u\|^p_{L_{p+1}}.$$

We then note that when $t < d/2$, $H = H^t$ is continuously embedded in L_{p+1} if and only if $p \leq p^*$, and this embedding holds for all $p \geq 0$ when $t \geq d/2$. We therefore conclude that

$$
(4.8) \qquad \qquad \|\mathcal{F}(u)\|_{H'} \leq C(1 + \|u\|_H^p).
$$

Therefore, $\mathcal{F}$ maps H to H' provided that $p \leq p^*$ when $t < d/2$, and for all $p \geq 0$ otherwise.

For the stability property, we use the inequality

$$
(4.9) \qquad \qquad |\mathcal{F}(u) - \mathcal{F}(v)| \leq C|u - v|(1 + |u| + |v|)^{[p-1]_+},
$$

which is a consequence of (4.1) with $n = 1$. Therefore, one has for all $\varphi \in H$

$$
(4.10) \quad |\langle \mathcal{F}(u) - \mathcal{F}(v), \varphi \rangle| \leq C \left[\int_\Omega |\varphi||u - v| + \int |\varphi||u - v|(|u| + |v|)^{[p-1]_+} \right].
$$

The first term is simply bounded by

$$
(4.11) \qquad \qquad \int_\Omega |\varphi||u - v| \leq \|\varphi\|_{L_2}\|u - v\|_{L_2} \leq \|\varphi\|_H \|u - v\|_H.
$$

If $p \leq 1$, the second term is bounded analogously. If $p > 1$, we apply Hölder's inequality twice, again with $q = p + 1$, to obtain

$$
(4.12) \int_\Omega |\varphi||u - v|(|u| + |v|)^{p-1} \leq \|\varphi\|_{L_{p+1}}\|u - v\|_{L_{p+1}} \left(\|u\|_{L_{p+1}} + \|v\|_{L_{p+1}} \right)^{p-1}.
$$

Using again the Sobolev embedding, these factors are controlled by $\|\varphi\|_H$, $\|u - v\|_H$, and $(\|u\|_H + \|v\|_H)^{p-1}$, so that we obtain

$$
(4.13) \qquad \qquad \|\mathcal{F}(u) - \mathcal{F}(v)\|_{H'} \leq C\|u - v\|_H,
$$

which is exactly (3.1).   □

Next we want to prove the local version (3.2) of Assumption 1. For a given subdomain D, we define a vector $\bar{\mathbf{v}} = (\bar{v}_\lambda)$ such that $\bar{v}_\lambda = v_\lambda$ if $S_\lambda \cap D \neq \emptyset$, and $\bar{v}_\lambda = u_\lambda$ otherwise. It follows that

$$
(4.14) \qquad \qquad \|(\mathbf{u} - \mathbf{v})|_{\{\lambda : S_\lambda \cap D \neq \emptyset\}}\| = \|\mathbf{u} - \bar{\mathbf{v}}\|.
$$

Denoting by $v, \bar{v}$ the corresponding functions $v = \sum_{\lambda \in \mathcal{J}} v_\lambda \psi_\lambda$, $\bar{v} = \sum_{\lambda \in \mathcal{J}} \bar{v}_\lambda \psi_\lambda$, we clearly have $v = \bar{v}$ on D so that

$$
(4.15) \qquad \qquad \mathbf{F}(\bar{\mathbf{v}})_\lambda = \langle \mathcal{F}(\bar{v}), \psi_\lambda \rangle = \langle \mathcal{F}(v), \psi_\lambda \rangle = \mathbf{F}(\mathbf{v})_\lambda
$$

whenever $S_\lambda \subset D$. It follows that

$$
(4.16) \qquad \qquad \|(\mathbf{F}(\mathbf{u}) - \mathbf{F}(\mathbf{v}))|_{\{\lambda : S_\lambda \subset D\}}\| \leq \|\mathbf{F}(\mathbf{u}) - \mathbf{F}(\bar{\mathbf{v}})\|.
$$

Therefore, the local stability estimate (3.2) follows by combining (4.14) and (4.16) together with the global stability estimate (3.1).   □

4.2. Verification of Assumption 2. For the verification of Assumption 2, we shall assume that either (4.1) is valid or that

$$(4.17) \qquad |\mathcal{F}^{(n)}(x)| \le C(1+|x|)^{p-n}, \;\; n \le p, \qquad \text{and} \;\; \mathcal{F}^{(n)}(x) = 0, \;\; n > p,$$

with p an integer. We shall show in the next theorem that whenever $\mathcal{F}$ satisfies either (4.1) or (4.17), then it satisfies Assumption 2 with $\gamma = r + t + d/2$. Our reason for separating the two cases (4.1) and (4.17) is that in the latter case we can take a larger value for r. We recall the critical index p^* defined by (4.2).

THEOREM 4.1. *Assume that the wavelets ψ_λ belong to C^m and have (for those $\lambda \in \mathcal{J}_\psi$) vanishing moments of order m (i.e., are orthogonal to $\mathbb{P}_{m-1}$ the space of polynomials of total degree at most $m-1$) for some positive integer m. Then Assumption 2 holds for $\gamma = r + t + d/2$ with the following value of r:*
   (i) *If $t \ge d/2$ and $\mathcal{F}$ satisfies (4.1) for some $p \ge 0$, then $r = \min\{m, n^*\}$.*
   (ii) *If $t < d/2$ and $\mathcal{F}$ satisfies (4.1) with $0 \le p < p^*$, then $r = \lceil \min\{m, p, n^*\} \rceil$.*
   (iii) *If $t \ge d/2$ and $\mathcal{F}$ satisfies (4.17) for some $p > 0$, then $r = m$.*
   (iv) *If $t < d/2$ and $\mathcal{F}$ satisfies (4.17) for some $0 \le p < p^*$, then $r = m$.*

Proof. Suppose that u has a finite wavelet expansion. We assume that $r \ge 1$ and leave the simpler case $r = 0$ to the reader (this case only occurs in (ii) when $p < 1$). Since the wavelets $\psi_\lambda \in \mathcal{J}_\psi$ have at least r vanishing moments, we have

$$|w_\lambda| \;=\; |\langle w, \psi_\lambda \rangle| = \inf_{P \in \Pi_{r-1}} |\langle w - P, \psi_\lambda \rangle|$$

$$(4.18) \qquad\quad \lesssim \; |w|_{W^r(L_\infty(S_\lambda))} 2^{-r|\lambda|} 2^{-(t+d/2)|\lambda|} = |w|_{W^r(L_\infty(S_\lambda))} 2^{-\gamma|\lambda|},$$

where $w(x) = \mathcal{F}(u(x))$. Using the chain rule, any rth order derivative of w can be written as a finite sum of functions of the form

$$(4.19) \qquad\qquad \mathcal{F}^{(k)}(u) D^{\beta_1} u \cdots D^{\beta_k} u, \quad k = 1, \dots, r,$$

where $|\beta_1| + \cdots + |\beta_k| = r$ with the usual notation $|\beta_i| := \beta_{i,1} + \cdots + \beta_{i,d}$. Therefore, one has

$$(4.20) \qquad |w|_{W^r(L_\infty(S_\lambda))} \;\lesssim\; \max_{k=1,\dots,r} \max_{|\beta_1|+\cdots+|\beta_k|=r} \|\mathcal{F}^{(k)}(u)\|_{L_\infty(S_\lambda)} \prod_{i=1}^{k} \|D^{\beta_i} u\|_{L_\infty(S_\lambda)}.$$

To bound the right side of (4.20), we recall that

$$(4.21) \qquad\qquad \|D^{\beta_i} u\|_{L_\infty(S_\lambda)} \;\lesssim\; \|u\|_{L_\infty(S_\lambda)}^{1-|\beta_i|/r} |u|_{W^r(L_\infty(S_\lambda))}^{|\beta_i|/r}.$$

This gives for $k = 1, \dots, r$

$$\|\mathcal{F}^{(k)}(u)\|_{L_\infty(S_\lambda)} \prod_{i=1}^{k} \|D^{\beta_i} u\|_{L_\infty(S_\lambda)} \;\lesssim\; \|\mathcal{F}^{(k)}(u)\|_{L_\infty(S_\lambda)} \|u\|_{L_\infty(S_\lambda)}^{k-1}$$

$$(4.22) \hspace{9cm} \times |u|_{W^r(L_\infty(S_\Lambda))}.$$

We shall finish the proof by separating it into two cases depending on the size of $\|u\|_{L_\infty(S_\lambda)}$.

Case $\|u\|_{L_\infty(S_\lambda)} \ge 1$. In this case, (4.20), (4.22), and the bounds (4.1) and (4.17) give

$$(4.23) \qquad\qquad |w|_{W^r(L_\infty(S_\lambda))} \;\lesssim\; \|u\|_{L_\infty(S_\lambda)}^{M} |u|_{W^r(L_\infty(S_\lambda))},$$

where $M := \max\{p, r\} - 1$ in cases (i) and (ii), and $M := p - 1$ in cases (iii) and (iv). We can bound the norms on the right side of (4.23) by using the Besov spaces $B_\infty^s(L_\infty)$, $s \geq 0$. They satisfy the norm equivalences

$$(4.24) \quad \|v\|_{B_\infty^s(L_\infty(S_\lambda))} \sim \sup_{S_\mu \cap S_\lambda \neq \emptyset} \left(2^{s|\mu|} \|v_\mu \psi_\mu\|_{L_\infty} \right) = \sup_{S_\mu \cap S_\lambda \neq \emptyset} \left(2^{(s+\delta)|\mu|} |v_\mu| \right)$$

with $\delta := \frac{d}{2} - t$. Here we used the fact that the H-normalization of the wavelets implies $\|\psi_\mu\|_{L_\infty} \sim 2^{\delta|\mu|}$. We also recall the embedding estimates

$$(4.25) \quad \|u\|_{W^s(L_\infty(S_\lambda))} \lesssim \|u\|_{B_\infty^{s+\varepsilon}(L_\infty(S_\lambda))}$$

for any fixed $\varepsilon \in]0, 1[$ and all $s \geq 0$. Using all of this in (4.23), we obtain

$$|w|_{W^r(L_\infty(S_\lambda))} \lesssim \left(\sup_{S_\mu \cap S_\lambda \neq \emptyset} 2^{(\delta+\varepsilon)|\mu|} |u_\mu| \right)^M \left(\sup_{S_\mu \cap S_\lambda \neq \emptyset} 2^{(r+\delta+\varepsilon)|\mu|} |u_\mu| \right)$$

$$(4.26) \qquad = \left(2^{(\delta+\varepsilon)|\mu_0|} |u_{\mu_0}| \right)^M \left(2^{(r+\delta+\varepsilon)|\mu_1|} |u_{\mu_1}| \right),$$

where μ_0 and μ_1 are the maximizing indices. If $\delta < 0$ (i.e., $t > d/2$), we can take $\epsilon < |\delta|$ and obtain the bound

$$(4.27) \quad |w|_{W^r(L_\infty(S_\lambda))} \lesssim \|\mathbf{u}\|^M (2^{(r+\delta+\varepsilon)|\mu_1|} |u_{\mu_1}|) \leq \|\mathbf{u}\|^M \sup_{S_\mu \cap S_\lambda \neq \emptyset} 2^{r|\mu|} |u_\mu|,$$

which verifies Assumption 2 in this case. If $\delta > 0$, then $|\mu_1| \geq |\mu_0|$ and $p < p_*$ and $M = p - 1$, and so we obtain

$$(4.28) \quad |w|_{W^r(L_\infty(S_\lambda))} \lesssim \|\mathbf{u}\|^M (2^{(r+p\delta+p\varepsilon)|\mu_1|} |u_{\mu_1}|) \leq \|\mathbf{u}\|^M \sup_{S_\mu \cap S_\lambda \neq \emptyset} 2^{(r+t+d/2)\mu|} |u_\mu|,$$

provided $p\epsilon < (p^* - p)\delta$. So we have completed the proof in this case.

$\quad$ *Case* $\|u\|_{L_\infty(S_\lambda)} < 1$. In this case, starting from (4.22) and using either (4.1) or (4.17), we obtain

$$|w|_{W^r(L_\infty(S_\lambda))} \lesssim |u|_{W^r(L_\infty(S_\lambda))} \lesssim \|u\|_{B_\infty^{r+\varepsilon}(L_\infty(S_\lambda))}$$

$$(4.29) \qquad \lesssim \sup_{S_\mu \cap S_\lambda \neq \emptyset} 2^{(r+\delta+\epsilon)|\mu|} |u_\mu| \lesssim \sup_{S_\mu \cap S_\lambda \neq \emptyset} 2^{(r+d/2+t)\mu|} |u_\mu|,$$

provided $\varepsilon < 2t$. Therefore, we have verified Assumption 2 in this case as well. $\quad \square$

$\quad$ **5. Multiple arguments and derivatives.** In this final section, we shall extend the previous results to more general nonlinear operators of the form

$$(5.1) \qquad (u_1, \ldots, u_n) \mapsto w = \mathcal{F}(D^{\alpha_1} u_1, \ldots, D^{\alpha_n} u_n),$$

acting from $H \times \cdots \times H$ to its dual H' (note that $\alpha_i = (\alpha_{i,1}, \ldots, \alpha_{i,d})$ are multi-indices). These include multilinear operators as particular cases. Here we shall indicate the appropriate generalizations of the results in the two previous sections with brief sketches of proofs since they are quite similar but notationally heavier.

5.1. Sparsity preserving discrete operators. Denoting by $\mathbf{u}_i = (u_{i,\lambda})$ the arrays of the wavelet coefficients of the function u_i, $\mathbf{u} = (\mathbf{u}_1, \ldots, \mathbf{u}_n)$, and $\mathbf{F}$ the corresponding discrete mapping

$$(5.2) \qquad \mathbf{F}(\mathbf{u}) := (\langle \mathcal{F}(D^{\alpha_1} u_1, \ldots, D^{\alpha_n} u_n), \psi_\lambda \rangle_{\lambda \in \mathcal{J}},$$

we introduce the following generalization of the basic assumptions.

Assumption 1. $\mathbf{F}$ is a Lipschitz map from $(\ell_2)^n$ into ℓ_2:

$$(5.3) \qquad \|\mathbf{F}(\mathbf{u}) - \mathbf{F}(\mathbf{v})\| \leq C \sum_{i=1}^{n} \|\mathbf{u}_i - \mathbf{v}_i\|,$$

with $C = C(\max_i \{\|\mathbf{u}_i\|, \|\mathbf{v}_i\|\})$, where $x \mapsto C(x)$ is a positive nondecreasing function.

The local version of this stability assumption now reads

$$(5.4) \qquad \|(\mathbf{F}(\mathbf{u}) - \mathbf{F}(\mathbf{v}))|_{\{\lambda : S_\lambda \subset D\}}\| \leq C \sum_{i=1}^{n} \|(\mathbf{u}_i - \mathbf{v}_i)|_{\{\lambda : S_\lambda \cap D \neq \emptyset\}}\|$$

for any domain D.

Assumption 2. For any finitely supported $\mathbf{u}$ (i.e., with all $\mathbf{u}_i$ finitely supported) and $\mathbf{w} = \mathbf{F}(\mathbf{u})$, we have the estimate

$$(5.5) \qquad |w_\lambda| \leq C \sup_{\mu \,:\, S_\lambda \cap S_\mu \neq \emptyset} \left(\sum_{i=1}^{n} |u_{i,\mu}| \right) 2^{-\gamma(|\lambda| - |\mu|)}$$

for all $\lambda \in \mathcal{J}_\psi$, where $\gamma > d/2$, $C = C(\max_i \|\mathbf{u}_i\|)$, and $x \mapsto C(x)$ is a positive nondecreasing function.

In order to generalize the construction of the near best approximation tree from section 3.3, we construct for a prescribed accuracy ε the trees $\tilde{\mathcal{T}}_{j,i}$ for each component $\mathbf{u}_i$ in the same way as we constructed $\tilde{\mathcal{T}}_j$ in section 3.3. We then define

$$(5.6) \qquad \tilde{\mathcal{T}}_j := \cup_{i=0}^{n} \tilde{\mathcal{T}}_{j,i} \text{ and } \Delta_j := \tilde{\mathcal{T}}_j \setminus \tilde{\mathcal{T}}_{j+1}.$$

The tree $\mathcal{T}$ is then constructed as before, according to (3.11).

THEOREM 5.1. *With this definition of $\mathcal{T}$ and under the above generalized Assumptions 1 and 2, if $\mathbf{u} \in (\mathcal{A}^s)^n$, we obtain that the same conclusions as in Theorem 3.2 also hold.*

Sketch of proof. As in the proof of Theorem 3.2, we start by invoking the stability property, which leads us to estimate $\|\mathbf{w}_\varepsilon - \mathbf{w}_\varepsilon|_{\mathcal{T}}\|$, where $\mathbf{w}_\varepsilon := \mathbf{F}(\mathbf{u}|_{\tilde{\mathcal{T}}_0})$. We use the estimate

$$(5.7) \qquad |w_{\varepsilon,\lambda}|^2 \lesssim \sum_{\mu \in \tilde{\mathcal{T}}_0, S_\mu \cap S_\lambda \neq \emptyset} \left[\sum_{i=1}^{n} |u_{i,\mu}|^2 \right] 2^{-2\gamma(|\lambda| - |\mu|)},$$

together with

$$(5.8) \qquad \sum_{\mu \in \Delta_j} \sum_{i=1}^{n} |u_{i,\mu}|^2 = \|\mathbf{u}|_{\tilde{\mathcal{T}}_j} - \mathbf{u}|_{\tilde{\mathcal{T}}_{j+1}}\|^2 \lesssim \frac{2^{2j}\varepsilon^2}{(1+j)^2},$$

in order to derive $\|\mathbf{w}_\varepsilon - \mathbf{w}_\varepsilon|_{\mathcal{T}}\| \lesssim \varepsilon$ in a similar way as in the proof of Theorem 3.2. The estimate on $\#(\mathcal{T})$ also remains the same using that $\#(\Delta_j) \leq \#(\tilde{\mathcal{T}}_j) \leq \sum_{i=1}^{n} \#(\mathcal{T}_{j,i})$. $\square$

In order to generalize the construction of the residual thresholding tree from section 3.4, we fix a threshold $\eta > 0$ and define for all $\mu \in \mathcal{J}$ the number $n(\mu)$ satisfying

$$(5.9) \qquad \eta 2^{\gamma n(\mu)} \leq \max_i |u_{i,\mu}| < \eta 2^{\gamma(n(\mu)+1)},$$

where $u_{i,\mu}$ are the coefficients of $\mathbf{u}_i$. We then define the influence sets $\Lambda_{\eta,\mu}$ and the tree $\mathcal{T}$ in a similar way as in (3.21) and (3.23), with $\tilde{\mathcal{T}}_\eta = \cup_{i=1}^n \tilde{\mathcal{T}}_\eta(\mathbf{u}_i)$ and $\tilde{\mathcal{T}}_\eta(\mathbf{u}_i)$ the expansion of the tree $\mathcal{T}_\eta(u_i)$.

THEOREM 5.2. *With this definition of $\mathcal{T}$ and under the above generalized Assumptions 1 and 2, if $\mathbf{u} \in (\ell_\tau^w(\mathcal{J}))^n$, we obtain that the same conclusions as in Theorem 3.3 hold.*

Sketch of proof. As in the proof of Theorem 3.3, in order to prove (3.25) we first consider the restricted vector $\mathbf{u}_\eta = \mathbf{u}|_{\tilde{\mathcal{T}}_\eta}$ and its image $\mathbf{w}_\eta := \mathbf{F}(\mathbf{u}_\eta) = (w_{\lambda,\eta})$. Using (5.5), we obtain that for any $\lambda \notin \mathcal{T}$, we have $|\tilde{w}_{\lambda,\eta}| \lesssim \eta$. We then use (5.4) in a similar way in order to derive (3.25).

In order to prove (3.26), we use the trees $\tilde{\mathcal{T}}_j := \tilde{\mathcal{T}}_{\eta 2^{\gamma j}}$ to decompose $\mathcal{T}$ into layers indexed by j as in the proof of Theorem 3.3. We then proceed in a similar way to derive (3.26), remarking that

$$(5.10) \qquad \#(\tilde{\mathcal{T}}_j) \lesssim \eta^{-\tau} 2^{-\gamma \tau j} \sup_i \|\mathbf{u}_i\|_{t\ell_\tau^w(\mathcal{J})}^\tau,$$

and that, according to (5.9) and (3.21), $n(\mu) = j$ for $\mu \in \tilde{\mathcal{T}}_j \setminus \tilde{\mathcal{T}}_{j+1}$.

Finally, we prove

$$(5.11) \qquad \|\mathbf{w}\|_{t\ell_\tau^w(\mathcal{J})} \lesssim 1 + \sup_i \|\mathbf{u}_i\|_{t\ell_\tau^w(\mathcal{J})}$$

by the same arguments as in the proof for Theorem 3.3. $\square$

The generalization of Algorithms EV1 and EV2 is straightforward, as is the following proposition.

PROPOSITION 5.1. *If $\mathbf{u} \in \mathcal{X}$ with $\mathcal{X} = (\mathcal{A}^s)^n$ (resp., $(\ell_\tau^w(\mathcal{J}))^n$), the tree $\mathcal{T}$ produced by Algorithm EV1 (resp., EV2) for target accuracy ε satisfies*

$$(5.12) \qquad \#(\mathcal{T}) \lesssim \sup_i \|\mathbf{u}_i\|_{\mathcal{X}}^{1/s} \varepsilon^{-1/s} + \#(\mathcal{J}_\phi),$$

with $s = 1/\tau - 1/2$, under the restriction $d/\gamma < \tau < 2$.

5.2. Verification of the basic assumptions. Recalling that the nonlinear map has the form $\mathcal{F}(D^{\alpha_1} u_1, \ldots, D^{\alpha_n} u_n)$, we shall therefore replace (4.1) by growth assumptions of the type

$$(5.13) \qquad |D^\beta \mathcal{F}(x_1, \ldots, x_n)| \leq C \prod_{i=1}^n (1 + |x_i|)^{[p_i - \beta_i]_+}, \quad |\beta| = 0, 1, \ldots, n^*,$$

for some $p_i \geq 0$ and n^* a positive integer. For notational simplicity, we shall write

$$(5.14) \qquad \mathcal{F}(u) = \mathcal{F}(D^{\alpha_1} u_1, \ldots, D^{\alpha_n} u_n), \quad \text{with } u = (u_1 \ldots, u_n).$$

We then obtain the following generalization of Proposition 4.1.

PROPOSITION 5.2. *Assume that the growth assumptions* (5.13) *hold at least with* $n^* = 0$. *Then* $\mathcal{F}$ *maps* $H \times \cdots \times H$ *to* H' *whenever* $H = H^t$ *and* $t \geq 0$ *satisfies*

$$(5.15) \qquad \left[\frac{1}{2} - \frac{t}{d}\right]_+ + \sum_{i=1}^{n} p_i \left[\frac{1}{2} - \frac{t}{d} + \frac{|\alpha_i|}{d}\right]_+ < 1.$$

If in addition $n^* = 1$, *then we also have under the same restriction*

$$(5.16) \qquad \|\mathcal{F}(u) - \mathcal{F}(v)\|_{H'} \leq C \sum_{i=1}^{n} \|u_i - v_i\|_H,$$

where $C = C(\max_i\{\|u_i\|_H, \|v_i\|_H\})$ *and* $x \to C(x)$ *is nondecreasing, and therefore Assumption* 1 *holds.*

Sketch of proof. For $u_i \in H$ and $\varphi \in H$, we write

$$(5.17) \qquad |\langle \mathcal{F}(u), \varphi \rangle| \leq C \int_\Omega |\varphi| \prod_{i=1}^{n} (1 + |D^{\alpha_i} u_i|)^{p_i}.$$

In view of (5.15), we can choose positive numbers r and r_i, $i = 1, \ldots, n$, such that $\frac{1}{r} + \sum_{i=1}^{n} \frac{p_i}{r_i} = 1$ and

$$(5.18) \qquad \frac{1}{r} > \frac{1}{2} - \frac{t}{d} \quad \text{and} \quad \frac{1}{r_i} > \frac{1}{2} - \frac{t}{d} + \frac{|\alpha_i|}{d}.$$

It follows that H^t is continuously embedded in L_r and $W^{|\alpha_i|}(L_{r_i})$. We can apply Hölder's inequality to obtain

$$(5.19) \qquad |\langle \mathcal{F}(u), \varphi \rangle| \leq C \|\varphi\|_{L_r} \prod_{i=1}^{n} (1 + \|D^{\alpha_i} u_i\|_{L_{r_i}}^{p_i}),$$

where we have used the fact that Ω is a bounded domain in order to control $\int_\Omega 1$ by a constant. In this way, we obtain

$$(5.20) \qquad \|\mathcal{F}(u)\|_{H'} \leq C \prod_{i=1}^{n} (1 + \|D^{\alpha_i} u_i\|_{H}^{p_i}).$$

For the stability property, we use the inequality

$$(5.21) \quad |\mathcal{F}(u) - \mathcal{F}(v)| \leq C \sum_{i=1}^{n} |D^{\alpha_i} u_i - D^{\alpha_i} v_i| \prod_{k=1}^{n} (1 + |D^{\alpha_k} u_k| + |D^{\alpha_k} v_k|)^{[p_k - \delta_{i,k}]_+},$$

with δ the Kronecker delta. Therefore, when estimating $|\langle \mathcal{F}(u) - \mathcal{F}(v), \varphi \rangle|$ for $\varphi \in H$, we are led to expressions of the form

$$(5.22) \qquad E_i = \int_\Omega |\varphi| |D^{\alpha_i} u_i - D^{\alpha_i} v_i| \prod_{k=1}^{n} (1 + |D^{\alpha_k} u_k| + |D^{\alpha_k} v_k|)^{[p_k - \delta_{i,k}]_+}$$

for each i. Using Hölder's inequality, we obtain

$$E_i \leq C \|\varphi\|_{L_r} \|D^{\alpha_i} u_i - D^{\alpha_i} v_i\|_{L_q} \prod_{k=1}^{n} \left(1 + \|D^{\alpha_k} u_k\|_{L_{r_k}}^{[p_k - \delta_{i,k}]_+} + \|D^{\alpha_k} v_k\|_{L_{r_k}}^{[p_k - \delta_{i,k}]_+}\right)$$

whenever

$$(5.23) \qquad \frac{1}{r} + \frac{1}{q} + \sum_{k=1}^{n} \frac{[p_k - \delta_{i,k}]_+}{r_k} = 1.$$

In view of (5.15), we can choose positive numbers r, q, and r_i satisfying condition (5.23) such that

$$(5.24) \qquad \frac{1}{r} > \frac{1}{2} - \frac{t}{d}, \quad \frac{1}{q} > \frac{1}{2} - \frac{t}{d} + \frac{|\alpha_i|}{d}, \quad \text{and} \quad \frac{1}{r_k} > \frac{1}{2} - \frac{t}{d} + \frac{|\alpha_k|}{d}.$$

Therefore, the Sobolev embedding gives

$$(5.25) \qquad E_i \leq C\|\varphi\|_H \|u_i - v_i\|_H,$$

and therefore Assumption 1 holds.   ☐

The local version (3.2) of Assumption 1 is derived in the same way as in section 4.2. Note that the condition (5.15) does not yield the optimal condition (4.2) in the simple case $n = 1$ and $\alpha_1 = 0$ due to the strict inequality but that we anyway need this strict inequality in order to obtain the validity of Assumption 2 according to Theorem 4.1.

For the proof of Assumption 2, we again treat separately the polynomial case for which we have the growth condition

$$(5.26) \qquad |D^\beta \mathcal{F}(x_1, \ldots, x_n)| \leq C \prod_{i=1}^{n} (1 + |x_i|)^{p_i - \beta_i}, \quad \beta_i \leq p_i,$$

and $D^\beta \mathcal{F} = 0$ if $\beta_i > p_i$ for some i, where p_i are positive integers.

THEOREM 5.3. *Assume that the wavelets belong to C^m and have vanishing moments of order m (i.e., are orthogonal to $\mathbb{P}_{m-1}$ the space of polynomials of total degree at most $m - 1$) for some positive integer m. Then Assumption 2 holds for $\gamma = r + t + d/2$ with the following values of r:*
  (i) *If $\mathcal{F}$ satisfies (5.13) with p such that $\sum_{i=1}^{n} p_i [d/2 - t + |\alpha_i|]_+ < d/2 + t$, then $r = \lceil \min\{m, n^*, p^*\} \rceil$, where $p^* = \min\{p_i : i \text{ s.t. } d/2 - t + |\alpha_i| > 0\}$.*
  (ii) *If $\mathcal{F}$ satisfies (5.26) with p such that $\sum_{i=1}^{n} p_i [d/2 - t + |\alpha_i|]_+ < d/2 + t$, then $r = m$.*

Sketch of proof. We shall prove (i); the other case is similar. We shall also assume that $r \geq 1$ and leave the simpler case $r = 0$ to the reader. As in the proof of Theorem 4.1 we start from the estimate

$$(5.27) \qquad |w_\lambda| \lesssim |w|_{W^r(L_\infty(S_\lambda))} 2^{-\gamma|\lambda|},$$

where $w(x) = \mathcal{F}(u(x))$. Using the chain rule, any rth order derivative of w can be written as a finite sum of functions of the form

$$(5.28) \qquad D^\nu \mathcal{F}(D^{\alpha_1} u_1, \ldots, D^{\alpha_n} u_n) G_\nu, \quad |\nu| = 1, \ldots, r,$$

where

$$(5.29) \qquad G_\nu = \prod_{i=1}^{n} \prod_{j=1}^{\nu_i} D^{\beta_{i,j} + \alpha_i} u_i,$$

and $\sum_{i=1}^n \sum_{j=1}^{\nu_i} |\beta_{i,j}| = r$. Therefore, one has

$$
(5.30) \qquad |w|_{W^r(L_\infty(S_\lambda))} \lesssim \max_{|\nu| \leq r} A_\nu,
$$

where

$$
(5.31) \qquad A_\nu := \|D^\nu \mathcal{F}(D^{\alpha_1} u, \ldots, D^{\alpha_n} u_n)\|_{L_\infty(S_\lambda)} \|G_\nu\|_{L_\infty(S_\lambda)}.
$$

To bound $\|G_\nu\|_{L_\infty(S_\lambda)}$, we use the estimate for intermediate derivatives (4.21) and find with $r_i := \sum_{j=1}^{\nu_i} |\beta_{i,j}|$ that

$$
(5.32) \qquad \|G_\nu\|_{L_\infty(S_\lambda)} \lesssim \prod_{i=1}^n |u_i|_{W^{|\alpha_i|}(L_\infty(S_\lambda))}^{\nu_i-1} |u_i|_{W^{r_i+|\alpha_i|}(L_\infty(S_\lambda))}.
$$

We now invoke (5.13) and obtain that

$$
A_\nu \leq \prod_{i=1}^n (1 + |u_i|_{W^{|\alpha_i|}(L_\infty(S_\lambda))})^{(p_i-\nu_i)_+} |u_i|_{W^{|\alpha_i|}(L_\infty(S_\lambda))}^{\nu_i-1} |u|_{W^{r_i+|\alpha_i|}(L_\infty(S_\lambda))}
$$

$$
(5.33) \qquad \leq \prod_{i=1}^n |u_i|_{W^{|\alpha_i|}(L_\infty(S_\lambda))}^{M_i} |u|_{W^{r_i+|\alpha_i|}(L_\infty(S_\lambda))},
$$

where $M_i = \max(p_i, r_i) - 1$ if $|u_i|_{W^{|\alpha_i|}(L_\infty(S_\lambda))} \geq 1$, and $M_i = 0$ otherwise.

Each term appearing in the last product in (5.33) can be bounded by Besov norms. The arguments used in deriving (4.28) and (4.29) give

$$
(5.34) \qquad |u_i|_{W^{|\alpha_i|}(L_\infty(S_\lambda))}^{M_i} |u|_{W^{r_i+|\alpha_i|}(L_\infty(S_\lambda))} \lesssim \|\mathbf{u}\|^{M_i} 2^{(r_i+(M_i+1)(|\alpha_i|+\delta+\epsilon))|\mu_i|} |u_{i,\mu_i}|,
$$

where μ_i is a maximizing index. Let $\mu^* := \max_i \mu_i$. We place (5.34) into (5.33). Each term $|u_{i,\mu_i}|$, $\mu_i \neq \mu^*$, we pull out of the product by the majorant $\|\mathbf{u}\|$. This then gives

$$
(5.35) \quad A_\nu \lesssim \prod_{i=1}^n \|\mathbf{u}\|^{M_i} 2^{(r_i+(M_i+1)(|\alpha_i|+\delta+\epsilon))|\mu_i|} |u_{i,\mu_i}| \lesssim \|\mathbf{u}\|^M \left[\sum_{i=1}^n |u_{i,\mu^*}| \right] 2^{\tilde\gamma|\mu^*|},
$$

with $M = \sum_{i=1}^n M_i$ and

$$
(5.36) \qquad \tilde\gamma = r + \sum_{i=1}^n (1 + M_i)(\varepsilon + [d/2 - t + |\alpha_i|]_+).
$$

Now, consider any term in the sum which is not zero. If $M_i \neq 0$, then $M_i + 1 = \max(p_i, r_i) \leq \max(p_i, r) = p_i$ because $r \leq p_i$. If $M_i = 0$, then $M_i + 1 = 1 \leq p_i$ because by definition $r \leq p^* \leq p_i$, and we have assumed $r \geq 1$. Using this information in (5.36) shows that

$$
(5.37) \qquad \tilde\gamma \leq r + \sum_{i=1}^n p_i(\varepsilon + [d/2 - t + |\alpha_i|]_+) \leq \gamma,
$$

provided ϵ is sufficiently small. $\square$

REFERENCES

[1] S. BERTOLUZZA, *Adaptive wavelet collocation for the solution of steady state equation*, in SPIE Proc. Wavelet Appl. II, 2491 (1995), pp. 947–956.

[2] P. BINEV AND R. DEVORE, *Fast computations in adaptive tree approximation*, Numer. Math., to appear.

[3] G. BOURDAUD, *Fonctions qui opèrent sur les espaces de Besov et de Triebel*, Ann. Inst. H. Poincaré Anal. Non Linéaire, 10 (1993), pp. 413–422.

[4] G. BOURDAUD AND D. KATEB, *Fonctions qui opèrent sur les espacs de Besov*, Math. Ann., 303 (1995), pp. 653–675.

[5] C. CANUTO AND I. CRAVERO, *Wavelet-based adaptive methods for advection-diffusion problems*, Math. Models Methods Appl. Sci., 7 (1997), pp. 265–289.

[6] C. CANUTO, A. TABACCO, AND K. URBAN, *The wavelet element method, part* I: *Construction and analysis*, Appl. Comput. Harmon. Anal., 6(1999), pp. 1–52.

[7] A. COHEN AND R. MASSON, *Wavelet adaptive methods for second order elliptic problems, boundary conditions and domain decomposition*, Numer. Math., 86 (2000), pp. 193–238.

[8] A. COHEN, W. DAHMEN, I. DAUBECHIES, AND R. DEVORE, *Tree approximation and optimal encoding*, Appl. Comp. Harm. Anal., 11 (2001), pp. 192–226.

[9] A. COHEN, W. DAHMEN, AND R. DEVORE, *Adaptive wavelet methods for elliptic operator equations—Convergence rates*, Math. Comp., 70 (2001), pp. 27–75.

[10] A. COHEN, W. DAHMEN, AND R. DEVORE, *Adaptive wavelet methods* II: *Beyond the elliptic case*, Found. Comput. Math., 2 (2002), pp. 203–245.

[11] A. COHEN, W. DAHMEN, AND R. DEVORE, *Adaptive Wavelet Schemes for Nonlinear Variational Problems*, IGPM Report 221, RWTH Aachen, Aachen, Germany, 2002.

[12] A. COHEN, S.M. KABER, S. MÜLLER, AND M. POSTEL, *Fully adaptive multiresolution finite volume schemes for conservation laws*, Math. Comp., 72 (2003), pp. 183–225.

[13] W. DAHMEN, *Adaptive approximation by smooth multivariate splines*, J. Approx. Theory, 36 (1982), pp. 119–140.

[14] W. DAHMEN AND R. SCHNEIDER, *Composite Wavelet Bases for Operator Equations*, Math. Comp., 68 (1999), pp. 1533–1567.

[15] W. DAHMEN, R. SCHNEIDER, AND Y. XU, *Nonlinear functions of wavelet expansions: Adaptive reconstruction and fast evaluation*, Numer. Math., 86 (2000), pp. 49–101.

[16] R. DEVORE, *Nonlinear Approximation*, in Acta Numer. 7, Cambridge University Press, Cambridge, UK, 1998, pp. 51–150.

[17] M. FARGE, M. GRIEBEL, F. KOSTER, AND K. SCHNEIDER, *Adaptive wavelet methods for the Navier-Stokes equations*, in Numerical Flow Simulation II, E. H. Hirschel, ed., Notes Numer. Fluid Mech. 75, Springer-Verlag, Berlin, 2001, pp. 303–318.

[18] M. FARGE AND K. SCHNEIDER, *Numerical simulation of a mixing layer in adaptive wavelet basis*, C. R. Acad. Sci. Paris Sér. II Fasc. b Méc., 328 (2000), pp. 263–269.

[19] A. HARTEN, *Adaptive multiresolution schemes for shock computations*, J. Comput. Phys., 115 (1994), pp. 319–338.

[20] Y. MADAY, V. PERRIER, AND J.C. RAVEL, *Adaptativité dynamique sur bases d'ondelettes pour l'approximation d'équations aux dérivées partielles*, C. R. Acad. Sci. Paris Sér. I, 1 (1991), pp. 405–410.

[21] T. RUNST AND W. SICKEL, *Sobolev Spaces of Fractional Order, Nemitskij Operators, and Nonlinear Partial Differential Equations*, de Gruyter Ser. Nonlinear Anal. Appl. 3, de Gruyter, New York, 1996.

[22] W. SICKEL, *Necessary conditions on composition operators acting between Besov spaces. The case $1 < s < n/p$*, Forum Math., 10 (1998), pp. 303–327.