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. The main result 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 $u_1, \cdots, u_n$ and their derivatives, i.e. $\mathcal{F}(D^{\alpha_1} u_1, \cdots, D^{\alpha_n} u_n)$.

Key Words: Nonlinear mappings, thresholding, tree structures, adaptive evaluation of nonlinear operators

AMS Subject Classification: 35A15, 35A22, 41A60, 46B15, 46A45, 47H30

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 partial differential equations. Such schemes typically rely on the sparsity of the wavelet representation of the solution allowing for data compression, as well as in 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 [11]. 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], [4], [8], [9] and [23].

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

*This work has been supported in part by the Office of Naval Research Contract N0014-91-J1343, the Army Research Office contract Nr. DAAD 19-02-1-0028, the TMR network "Wavelets in Numerical Simulation" and the SFB 401, funded by the German Research Foundation

†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, U.S.A., devore@math.sc.edu, http://www.math.sc.edu/~devore/

function with respect to certain scales of Besov spaces, see e.g. [15]. Hence the above issue is closely connected with the question how the regularity of a given function u is 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 has been treated by several authors and the reader is referred e.g. to [2], [3], [22], and to the book by Runst and Sickel [21] for a detailed treatment. 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 neither tell us *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, while this 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 provide in Section 3 a proof 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 a specific algorithm that constructs 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 [10]. 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 $\mathcal{F}(D^{\alpha_1}u_1, \cdots, D^{\alpha_n}u_n)$ between a multivariate map $\mathcal{F}$ and the derivatives of several functions $u_1, \cdots, u_n$ 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

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

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: find $u \in H_0^1(\Omega)$ such that

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

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 the perhaps 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. [9] 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 only consider compactly supported wavelets, i.e., the supports of the wavelets scale as follows

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

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 domain.

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

$$c\|(v_\lambda)_{\lambda\in\mathcal{J}}\| \leq \|\sum_{\lambda\in\mathcal{J}} v_\lambda \psi_\lambda\|_H \leq C\|(v_\lambda)_{\lambda\in\mathcal{J}}\|, \tag{2.4}$$

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 [5, 6, 13].

Note that, by duality, (2.4) is equivalent to

$$C^{-1}\|(\langle w, \psi_\lambda\rangle)_{\lambda\in\mathcal{J}}\| \leq \|w\|_{H'} \leq c^{-1}\|(\langle w, \psi_\lambda\rangle)_{\lambda\in\mathcal{J}}\|, \quad w \in H', \tag{2.5}$$

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

$$w_\lambda = \langle w, \psi_\lambda\rangle, \quad \lambda \in \mathcal{J}. \tag{2.6}$$

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

$$\mathbf{u} \mapsto \mathbf{w} = \mathbf{F}(\mathbf{u}) = (\langle \mathcal{F}(u), \psi_\lambda\rangle)_{\lambda\in\mathcal{J}}. \tag{2.7}$$

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* (2.4) and (2.5), this can be restated as follows: Suppose 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 $\mathbf{u} \in \ell^w_\tau(\mathcal{J})$. Here $\ell^w_\tau(\mathcal{J})$ is the collection of all $\mathbf{u} \in \ell_2(\mathcal{J})$ which satisfy

$$\#\{\lambda \in \mathcal{J} : |u_\lambda| > \eta\} \leq C\eta^{-\tau}, \quad \eta > 0. \tag{2.8}$$

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

$$\|\mathbf{u}\|_{\ell_\tau^w(\mathcal{J})} := \sup_{\eta>0} \eta[\#\{\lambda \in \mathcal{J} : |u_\lambda| > \eta\}]^{1/\tau}. \tag{2.9}$$

An equivalent norm is given by the quantity

$$\sup_{n>0} n^{1/\tau} u_n^*, \tag{2.10}$$

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

$$\|\mathbf{u}\| \lesssim \|\mathbf{u}\|_{\ell_\tau^w(\mathcal{J})}. \tag{2.11}$$

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

$$\sigma_N(\mathbf{u}) := \inf_{\#\mathrm{supp}\,\mathbf{v}\le N} \|\mathbf{u}-\mathbf{v}\| = \left(\sum_{n>N} |u_n^*|^2\right)^{1/2}, \tag{2.12}$$

one has the following characterization [8].

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*

$$\frac{1}{\tau} = s + \frac{1}{2}. \tag{2.13}$$

Moreover,

$$\sigma_N(\mathbf{u}) \lesssim N^{-s}\|\mathbf{u}\|_{\ell_\tau^w(\mathcal{J})}. \tag{2.14}$$

Thus the smaller τ 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 [8] 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 $S_\lambda \subset S_\mu$. Given a μ there are at most K indices $\lambda \in \mathcal{J}_{|\mu|+1}$ such that $S_\lambda \subset S_\mu$; these λ are called the *children* of μ and μ is a parent of λ. Similarly, every λ has at most K parents. Note that, by definition, whenever a λ belongs to a tree all of its parents belong to it as well.

If the tree $\mathcal{T} \subset \mathcal{J}$ is finite, we define the set $\mathcal{L}^+ = \mathcal{L}^+(\mathcal{T})$ of *inner leaves* as

$$\mathcal{L}^+ := \{\lambda \in \mathcal{T} : \mu \text{ a child of } \lambda \implies \mu \notin \mathcal{T}\}. \tag{2.15}$$

Additionally, the set $\mathcal{L}^- = \mathcal{L}^-(\mathcal{T})$ of *outer leaves* is the set of those indices outside the tree such that all of their parents belong to the tree

$$\mathcal{L}^- := \{\lambda \in \mathcal{J} : \lambda \notin \mathcal{T},\ S_\lambda \subset S_\mu \implies \mu \in \mathcal{T}\}. \tag{2.16}$$

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

We shall make use of the following easily verifiable inequalities

$$\#\mathcal{T} \sim \#\mathcal{L}^+(\mathcal{T}) \sim \#\mathcal{L}^-(\mathcal{T}), \tag{2.17}$$

where the constants depend only on K. Defining

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

one easily verifies that

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

Note that for any tree $\mathcal{T}$ both collections $\mathcal{L}^+(\mathcal{T})$ and $\mathcal{L}^-(\mathcal{T})$ share the property that none of their elements is contained in any other of their elements. This fact will be seen to have an important consequence that will be utilized several times. A collection $\mathcal{C} = \{C_I : I \in \mathcal{I}\}$ is said to have the *finite incidence property* (FIP) if the following holds: there exists a fixed integer M, such that for any subset $\mathcal{G} \subset \mathcal{C}$ with the property that no element of $\mathcal{G}$ is contained in any other element of $\mathcal{G}$, at most M elements of $\mathcal{G}$ have a nonempty intersection.

A prototype instance of (FIP) can be described as follows. Suppose that $C \subset \mathbb{R}^d$ satisfies $C \subseteq [0, L]^d$ for some fixed positive integer L and define $\mathcal{C}_L := \{C_{j,k} := 2^{-j}(k + C) : k \in \mathbb{Z}^d,\ j \in \mathbb{N}_0\}$. It is shown in [16] that $\mathcal{C}_L$ has the (FIP) for every $L \in \mathbb{N}$, where the number M depends on (the smallest) L (for which $C \subseteq [0, L]^d$) and the spatial dimension d. Thus, when Ψ is a wavelet basis on all of $\mathbb{R}^d$ generated by a classical multiresolution sequence of shift-invariant spaces, the collection of supports $\mathcal{C}(\Psi) := \mathbb{Z}^k \cup \{S_\lambda : \lambda = (j, k, e) \in \mathbb{N}_0 \times \mathbb{Z}^d \times \{0, 1\}^d \setminus \{0\}\}$ has the (FIP). It is not hard to see that $\mathcal{C}(\Psi^\pi)$ has the (FIP) when Ψ^π is a wavelet basis on the torus obtained by periodizing a basis on $\mathbb{R}^d$. For the simplest example of a bounded domain, namely the unit cube $\Box$ in $\mathbb{R}^d$, say, boundary adapted wavelet bases $\Psi^\Box$ can be constructed by taking tensor products of boundary adapted wavelet bases on the unit interval. The bases on $\Box$, although no longer being shift invariant, still have the property that the supports of the wavelets on the level j are contained in some cube $2^{-j}(k + [0, L]^d)$, for a fixed positive integer L and some $k \in \mathbb{Z}^d$. The arguments in [16] can still be used to show that $\mathcal{C}(\Psi^\Box)$ has the (FIP) also in this case. A common strategy for constructing wavelet bases on bounded domains or compact manifolds Ω is to partition the domain into a smooth parametric images of the unit cube and to build a basis Ψ^Ω on the whole domain from parametric liftings of a basis $\Psi^\Box$ on the unit cube $\Box$. Thus, intersections of supports of wavelets in the physical domain correspond to intersections in the parameter domain. Thus one can infer the (FIP) for $\mathcal{C}(\Psi^\Omega)$ from that of $\mathcal{C}(\Psi^\Box)$. Throughout the remainder of the paper we shall assume that the wavelet bases referred to possess the (FIP).

Now we associate to any sequence $\mathbf{u} = (u_\lambda)$ in $\ell_2(\mathcal{J})$ the corresponding sequence $\tilde{\mathbf{u}} = (\tilde{u}_\lambda)$ defined by

$$\tilde{u}_\lambda := \left(\sum_{\mu \in \Gamma_\lambda} |u_\mu|^2 \right)^{1/2}. \tag{2.20}$$

The coefficients $\tilde{u}_\lambda$ can be viewed as local error bounds. In fact, one has for any tree $\mathcal{T}$ that

$$\sum_{\lambda \in \mathcal{L}^-(\mathcal{T})} \tilde{u}_\lambda^2 \lesssim \|\mathbf{u} - \mathbf{u}|_{\mathcal{T}}\|^2. \tag{2.21}$$

To see this, recall that the collection of supports S_λ, $\lambda \in \mathcal{J}$ possesses the (FIP). Hence, any $\mu \notin \mathcal{T}$ is contained at most a uniformly bounded finite number of elements in $\mathcal{L}^-(\mathcal{T})$, which confirms (2.21).

One readily verifies that $S_\mu \subseteq S_\lambda$ implies $\tilde u_\lambda \ge \tilde u_\mu$, i.e. for any $\eta > 0$ the set

$$\mathcal{T}_\eta = \mathcal{T}_\eta(\mathbf{u}) := \{\lambda \, : \, |\tilde u_\lambda| > \eta\} \tag{2.22}$$

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

$$_t\ell^w_\tau(\mathcal{J}) := \{\mathbf{u} \in \ell_2(\mathcal{J}) : \tilde{\mathbf{u}} \in \ell^w_\tau(\mathcal{J})\}, \quad \|\mathbf{u}\|_{_t\ell^w_\tau(\mathcal{J})} := \|\tilde{\mathbf{u}}\|_{\ell^w_\tau(\mathcal{J})}. \tag{2.23}$$

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

$$\#\mathcal{T}_\eta(\mathbf{u}) \le \eta^{-\tau}\|\mathbf{u}\|^\tau_{_t\ell^w_\tau(\mathcal{J})}. \tag{2.24}$$

Therefore the spaces $_t\ell^w_\tau(\mathcal{J})$ 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 $\mathbf{u} \in {_t\ell^w_\tau}(\mathcal{J})$ implies the error estimate*

$$\|\mathbf{u} - \mathbf{u}_\eta\| \ \lesssim \ \eta^{1-\tau/2}\|\mathbf{u}\|^{\tau/2}_{_t\ell^w_\tau(\mathcal{J})} \ \lesssim \ [\#(\mathcal{T}_\eta)]^{-s}\|\mathbf{u}\|_{_t\ell^w_\tau(\mathcal{J})}, \tag{2.25}$$

with $s = 1/\tau - 1/2$. Conversely, $\|\mathbf{u} - \mathbf{u}_\eta\| \le C[\#(\mathcal{T}_\eta)]^{-s}$ for all $\eta > 0$ implies $\mathbf{u} \in {_t\ell^w_\tau}(\mathcal{J})$ and $\|\mathbf{u}\|_{_t\ell^w_\tau(\mathcal{J})} \le D$ with D proportional to C.

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) , (2.21) and using (2.24), one has

$$\begin{aligned} \|\mathbf{u} - \mathbf{u}_\eta\|^2 \ &= \ \sum_{\lambda \notin \mathcal{T}_\eta} |u_\lambda|^2 \le \sum_{\lambda \in \mathcal{L}^-_\eta} \tilde u^2_\lambda \le \#\mathcal{L}_\eta \eta^2 \\ &\lesssim \ \#\mathcal{T}_\eta \eta^2 \le \|\mathbf{u}\|^\tau_{_t\ell^w_\tau(\mathcal{J})}\eta^{2-\tau}, \end{aligned} \tag{2.26}$$

where we have used (2.17). This confirms the first estimate in (2.25). Since again by definition (2.23) and (2.24), $\eta \le \|\mathbf{u}\|_{_t\ell^w_\tau(\mathcal{J})}(\#\mathcal{T}_\eta)^{-1/\tau}$, the second estimate follows from (2.26).

As for the converse, let $N_\eta := \#\mathcal{T}_\eta$ and $M_\eta := \#\mathcal{L}^-_\eta$ so that, by (2.17), $N_\eta \sim M_\eta$ uniformly in η. Then, denoting by $(\tilde u^*_n)$ a non-increasing rearrangement of $\tilde{\mathbf{u}}$, we have by assumption

$$M_\eta(\tilde u^*_{N_\eta+M_\eta})^2 \le \sum_{\lambda \in \mathcal{L}^-_\eta} \tilde u^2_\lambda \ \lesssim \ \|\mathbf{u} - \mathbf{u}_\eta\|^2 \ \lesssim \ \#\mathcal{T}_\eta^{-2s},$$

where we have again used (2.21) . Thus $\tilde u^*_{N_\eta+M_\eta} \ \lesssim \ (\#\mathcal{T}_\eta)^{-(s+1/2)} = (\#\mathcal{T}_\eta)^{-1/\tau}$ for s and τ related as above. Since $\|\mathbf{u}\|^2 \ \gtrsim \ \eta^2\#\mathcal{L}^+_\eta$, where $\mathcal{L}^+_\eta := \mathcal{L}^+(\mathcal{T}_\eta)$, we conclude that $\#\mathcal{T}_\eta$ grows at most like η^{-2}. Hence $\#\mathcal{T}_{1/\sqrt{n}}$ grows at most like n which, in view of the above observation for $N = N_\eta$ confirms that $\mathbf{u} \in {_t\ell^w_\tau}(\mathcal{J})$. This completes the proof. □

Of course, the question arises which property of u implies that the array of wavelet coefficients $\mathbf{u}$ belongs to $_t\ell^w_\tau(\mathcal{J})$.

REMARK 2.3. *Let $H = H^t$ then $u \in B^{t+sd}_q(L_{\tau'})$ implies $\mathbf{u} \in {_t\ell^w_\tau}(\mathcal{J})$, whenever $\frac{1}{\tau'} < \frac{1}{\tau} = s + \frac{1}{2}$ and $0 < q \le \infty$.*

Sketch of proof: It is enough to prove this for $B^{t+sd}_\infty(L_{\tau'})$ and $\tau < \tau' \le 2$, because the remaining cases follow by embeddings. The condition $u \in B^{t+sd}_\infty(L_{\tau'})$ 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 that also $\left(\sum_{|\lambda|=j} |\tilde{u}_\lambda|^{\tau'}\right)^{1/\tau'}$ $\lesssim 2^{-jd\delta}$, $j \in \mathbb{N}$. 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 [7], 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 (2.23) means that $\mathbf{u} \in {}_t\ell_\tau^w(\mathcal{J})$ as claimed. □

We can now restate the above questions in the following way:

- Does $\mathbf{F}$ map a sequence $\mathbf{u} \in {}_t\ell_\tau^w(\mathcal{J})$ into a sequence $\mathbf{w} = \mathbf{F}(\mathbf{u}) \in {}_t\ell_\tau^w(\mathcal{J})$?
- Can we compute asymptotically optimal sparse approximations of $\mathbf{w} = \mathbf{F}(\mathbf{u})$ from asymptotically optimal sparse approximations of $\mathbf{u}$?

Note that a positive answer to the second question gives a positive answer to the first question in a constructive way.

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*

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

where $x \mapsto C(x)$ is a positive non-decreasing function.

The fact that the constant C might grow with the norm of $\mathbf{u}$ and $\mathbf{v}$ accounts for the nonlinearity of the tranformation. 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

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

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*

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

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

A typical value of γ is

$$\gamma := r + t + d/2, \tag{3.4}$$

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 fullfilled for a fairly general class of local composition operators.

3.2. Tree expansions. Given a tree $\mathcal{T}$, we have defined already the set $\mathcal{L}^-(\mathcal{T})$ of outer leaves for $\mathcal{T}$. It will be important in the trees that we construct that there is not too much overlap in the elements of $\mathcal{L}^-(\mathcal{T})$. We can always accomplish this by expanding the tree slightly as follows. 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 $\Phi(\lambda) := \cup_{k=0}^{|\lambda|}\Phi_k(\lambda)$.

Now, given a tree $\mathcal{T}$, we define the expansion $\tilde{\mathcal{T}}$ as

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

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*

$$\left.\begin{array}{c} |\mu'| < |\mu| \\ S_{\mu'} \cap S_\mu \neq \emptyset \end{array}\right\} \implies \mu' \in \tilde{\mathcal{T}}. \tag{3.6}$$

The following lemma, see e.g. [12, 14] shows that $\tilde{\mathcal{T}}$ has comparable size to $\mathcal{T}$.

LEMMA 3.1. *There exist constants C_1 and C_2 such that for any finite tree $\mathcal{T}$, we have:*
(i) $\#(\tilde{\mathcal{T}}) \le C_1\#(\mathcal{T})$,
(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 $\operatorname{dist}(S_\lambda, S_\mu) \le C_0 2^{-|\mu|}$ 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 $\operatorname{dist}(S_{\mu'}, S_{\bar{\lambda}}) \le C_0 2^{-|\mu'|}$. It follows that

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

Hence, we can take any 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 $||\nu| - |\mu|| \le 1$. In fact, suppose that $|\nu| < |\mu| - 1$. Then, for any parent μ' of μ we have $S_\nu \cap S_{\mu'} \neq \emptyset$. Since $\mu' \in \tilde{\mathcal{T}}$ and $|\mu'| > |\nu|$ we conclude $\nu \in \tilde{\mathcal{T}}$ which is a contradiction. This completes the proof. □

3.3. The main result. We wish to *predict* next the significant coefficients of $\mathbf{w} = \mathbf{F}(\mathbf{u})$. To this end, we fix $\eta > 0$ and for the constant γ of (3.3), we define for all $\mu \in \mathcal{J}$ the number $n(\mu)$ satisfying

$$\eta 2^{\gamma n(\mu)} \le |\tilde{u}_\mu| < \eta 2^{\gamma(n(\mu)+1)}, \tag{3.7}$$

and the *influence set*

$$\Lambda_{\eta,\mu} := \{\lambda \ : \ S_\lambda \cap S_\mu \neq \emptyset \text{ and } |\lambda| \le |\mu| + [n(\mu)]_+\}. \tag{3.8}$$

We then define a set of coefficients for the approximation of $\mathbf{w}$ by

$$\Lambda_\eta := \tilde{\Lambda}_\eta \cup \mathcal{J}_\phi, \tag{3.9}$$

with

$$\tilde\Lambda_\eta := \cup_{\mu\in\tilde{\mathcal T}_\eta}\Lambda_{\eta,\mu} \tag{3.10}$$

where $\mathcal{T}_\eta$ is the tree for $\mathbf{u}$ defined by (2.20) and $\tilde{\mathcal{T}}_\eta$ is its expansion defined in the previous section. We notice that Λ_η has a tree structure.

THEOREM 3.2. *Given any $\mathbf{u}\in\ell_2(\mathcal{J})$, one has the coefficient size estimate*

$$|\tilde w_\lambda| \ \lesssim\ \eta \ \text{ if } \ \lambda\notin\Lambda_\eta, \tag{3.11}$$

where the $\tilde w_\lambda$ are defined for $\mathbf{w}=\mathbf{F}(\mathbf{u})$ according to (2.20). If in addition $\mathbf{u}\in {}_t\ell^w_\tau(\mathcal{J})$ for some $d/\gamma<\tau<2$, then we have the cardinality estimate

$$\#(\Lambda_\eta) \ \lesssim\ \|\mathbf{u}\|^\tau_{{}_t\ell^w_\tau(\mathcal{J})}\eta^{-\tau}+\#(\mathcal{J}_\phi). \tag{3.12}$$

Moreover we have $\mathbf{w}\in {}_t\ell^w_\tau(\mathcal{J})$ and

$$\|\mathbf{w}\|_{{}_t\ell^w_\tau(\mathcal{J})} \ \lesssim\ 1+\|\mathbf{u}\|_{{}_t\ell^w_\tau(\mathcal{J})}. \tag{3.13}$$

The constants in these above inequalities depend only on the constants in Assumptions 1 and 2, the space dimension d, and the parameter τ in the case of (3.12).

Proof: In order to prove (3.11) , 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 $\lambda\notin\Lambda_\eta$ and for all $\mu\in\tilde{\mathcal{T}}_\eta$ such that $S_\mu\cap S_\lambda\neq\emptyset$, we have by (3.8) and (3.10) the inequality $|\lambda|-|\mu|\geq[n(\mu)]_+$. Therefore, remarking that $\lambda\in\mathcal{J}_\psi$, the local action assumption (3.3) implies

$$|w_{\lambda,\eta}| \ \lesssim\ \eta. \tag{3.14}$$

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

$$|w_{\nu,\eta}| \ \lesssim\ 2^{-\gamma l}\eta. \tag{3.15}$$

It follows that

$$|\tilde w_{\lambda,\eta}|^2 \ \lesssim\ \eta^2\Big(\sum_{l\geq 0}2^{(d-2\gamma)l}\Big) \ \lesssim\ \eta^2, \tag{3.16}$$

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

Next, we remark that for $\lambda\in\mathcal{L}^-(\Lambda_\eta)$, we have

$$\begin{aligned}\big||\tilde w_{\lambda,\eta}|^2-|\tilde w_\lambda|^2\big| &\leq (|\tilde w_{\lambda,\eta}|+|\tilde w_\lambda|)|\tilde w_{\lambda,\eta}-\tilde w_\lambda|\\ &\leq (2|\tilde w_{\lambda,\eta}|+|\tilde w_{\lambda,\eta}-\tilde w_\lambda|)|\tilde w_{\lambda,\eta}-\tilde w_\lambda|\\ &\lesssim (\eta+\|(\mathbf{w}-\mathbf{w}_\eta)|_{\Gamma_\lambda}\|)\|(\mathbf{w}-\mathbf{w}_\eta)|_{\Gamma_\lambda}\|.\end{aligned} \tag{3.17}$$

Now observe that, according to (3.2) ,

$$\begin{aligned}\|(\mathbf{w}-\mathbf{w}_\eta)|_{\Gamma_\lambda}\| &= \|(\mathbf{w}-\mathbf{w}_\eta)|_{\{\mu:S_\mu\subset S_\lambda\}}\|\\ &\leq C(\|\mathbf{u}\|)\|(\mathbf{u}-\mathbf{u}_\eta)|_{\{\mu:S_\mu\cap S_\lambda\neq\emptyset\}}\|\\ &= C(\|\mathbf{u}\|)\left(\sum_{\mu\notin\tilde{\mathcal{T}}_\eta,S_\mu\cap S_\lambda\neq\emptyset}|u_\mu|^2\right)^{1/2}\\ &\leq C(\|\mathbf{u}\|)\left(\sum_{\mu\in\mathcal{L}^-(\tilde{\mathcal{T}}_\eta),S_\mu\cap S_\lambda\neq\emptyset}|\tilde u_\mu|^2\right)^{1/2} \ \lesssim\ C_2C(\|\mathbf{u}\|)\eta,\end{aligned}$$

where C_2 is the constant from Lemma 3.1. Combining this with (3.16) and (3.17), we obtain the size estimate (3.11) .

To prove (3.12) we define the trees

$$\tilde{\mathcal{T}}_j := \tilde{\mathcal{T}}_{\eta 2^{\gamma j}}. \tag{3.18}$$

From (2.24) and Lemma 3.1, we infer that $\mathbf{u} \in {}_t\ell^w_\tau(\mathcal{J})$ implies

$$\#(\tilde{\mathcal{T}}_j) \lesssim \eta^{-\tau} 2^{-\gamma\tau j} \|\mathbf{u}\|^\tau_{{}_t\ell^w_\tau(\mathcal{J})}. \tag{3.19}$$

Writing

$$\tilde{\Lambda}_\eta = \bigcup_{j\geq 0} \bigcup_{\mu\in\tilde{\mathcal{T}}_j\setminus\tilde{\mathcal{T}}_{j+1}} \Lambda_{\eta,\mu}, \tag{3.20}$$

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 (3.7) and (3.8), $n(\mu) = j$ for $\mu \in \tilde{\mathcal{T}}_j \setminus \tilde{\mathcal{T}}_{j+1}$ we obtain in view of (3.19)

$$\begin{aligned} \#(\bigcup_{\mu\in\tilde{\mathcal{T}}_j\setminus\tilde{\mathcal{T}}_{j+1}} \Lambda_{\eta,\mu}) &\lesssim 2^{dj}\#(\tilde{\mathcal{T}}_j \setminus \tilde{\mathcal{T}}_{j+1}) + \#(\tilde{\mathcal{T}}_j) \\ &\lesssim \|\mathbf{u}\|^\tau_{{}_t\ell^w_\tau(\mathcal{J})} \eta^{-\tau} 2^{(d-\gamma\tau)j}. \end{aligned} \tag{3.21}$$

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

$$\#(\tilde{\Lambda}_\eta) \lesssim \|\mathbf{u}\|^\tau_{{}_t\ell^w_\tau(\mathcal{J})} \eta^{-\tau}, \tag{3.22}$$

and adding the cardinality of $\mathcal{J}_\phi$, we thus obtain (3.12).

In order to obtain the estimate (3.13), we first notice that (3.22) already indicates that we have the estimate

$$\|(\tilde{w}_\lambda)_{\lambda\in\mathcal{J}_\psi}\|_{\ell^w_\tau(\mathcal{J})} \lesssim \|\mathbf{u}\|_{{}_t\ell^w_\tau(\mathcal{J})}. \tag{3.23}$$

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

$$\begin{aligned} \|(\tilde{w}_\lambda)_{\lambda\in\mathcal{J}_\phi}\|_{\ell^w_\tau(\mathcal{J})} &\leq \|(\tilde{w}_\lambda)_{\lambda\in\mathcal{J}_\phi}\|_{\ell_\tau} \\ &\leq [\#(\mathcal{J}_\phi)]^{1/\tau-1/2} \|(\tilde{w}_\lambda)_{\lambda\in\mathcal{J}_\phi}\| \\ &\leq C\|\mathbf{w}\|, \end{aligned}$$

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

$$\|\mathbf{w}\| \lesssim \|\mathbf{F}(0)\| + \|\mathbf{u}\| \lesssim 1 + \|\mathbf{u}\| \lesssim 1 + \|\mathbf{u}\|_{{}_t\ell^w_\tau(\mathcal{J})}, \tag{3.24}$$

the estimate (3.13) follows □

Note that, since we have used Assumption 1 and its local version in the above proof, the constants in both estimate (3.11) and (3.13) are of the form $C(\|\mathbf{u}\|)$ where $x \mapsto C(x)$ is a positive non-decreasing function.

3.4. An Adaptive Evaluation Scheme. 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

$$\mathbf{A}\mathbf{u} + \mathbf{F}(\mathbf{u}) = \mathbf{f}, \tag{3.25}$$

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 (3.25). 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 [8]. 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. We shall discuss and analyze a numerical method for such evaluations which is suggested by Theorem 3.2.

Let us fix $\mathbf{v}$ with finite support and consider the numerical approximation of $\mathbf{w} = \mathbf{F}(\mathbf{v})$. We introduce the notation $\eta_j := 2^{-j}$ and $\bar\Lambda_j := \Lambda_{\eta_j}(\mathbf{v})$ where these sets are defined as in (3.9) for $\mathbf{v}$. Recall also the set $\mathcal{L}_j^- := \mathcal{L}_{\eta_j}^-$ of outer leaves of $\bar\Lambda_j$. We introduce the computable quantities

$$\epsilon_j^2 := C_0^2 \#(\mathcal{L}_j^-)\eta_j^2 \tag{3.26}$$

where C_0 is the constant of the inequality (3.11). The following algorithm **EV** which takes as input a tolerance $\epsilon > 0$, the nonlinear map $\mathbf{F}$, and the finitely supported vector and gives as output $\mathbf{EV}(\epsilon, \mathbf{F}, \mathbf{v}) = (\bar\Lambda_\epsilon, \mathbf{w}_\epsilon)$ where $\bar\Lambda_\epsilon$ is a finite set and $\mathbf{w}_\epsilon = \mathbf{w}|_{\bar\Lambda_\epsilon}$.

Algorithm EV *Given the inputs* $\epsilon > 0$, $\mathbf{F}$, $\mathbf{v}$, *do the following:*
Step 1: *Calculate* $\bar\Lambda_0$, $\mathcal{L}_0^-$, *and* ϵ_0. *If* $\epsilon_0 \le \epsilon$, *terminate the algorithm and take as output* $\bar\Lambda_\epsilon := \bar\Lambda_0$ *and* $\mathbf{w}_\epsilon := \mathbf{w}|_{\bar\Lambda_\epsilon}$. *If* $\epsilon_0 > \epsilon$ *set* $j = 1$ *and proceed to* **Step 2**.
Step 2: *Given* j *compute* $\bar\Lambda_j$, $\mathcal{L}_j^-$ *and* ϵ_j *and proceed to* **Step 3**.
Step 3: *If* $\epsilon_j \le \epsilon$, *terminate the algorithm and take as output* $\bar\Lambda_\epsilon := \bar\Lambda_j$ $\mathbf{w}_\epsilon := \mathbf{w}|_{\bar\Lambda_\epsilon}$. *If* $\epsilon_j > \epsilon$ *replace* j *by* $j+1$ *and return to* **Step 2**.

REMARK 3.1. *The above algorithm assumes that given a finite set* Λ, *we can compute* $\mathbf{w}_\Lambda$. *This is a numerical issue that we shall not engage except to mention the paper [14] which treat this topic. In going further, we assume that given any finite set* Λ *with tree structure,* $\mathbf{w}_\Lambda$ *can always be computed at a cost proportional to* $\#(\Lambda)$.

The following theorem summarizes the properties of Algorithm **EV**.

THEOREM 3.3. *Given the inputs* $\epsilon > 0$, *a nonlinear function* F *such that* $\mathbf{F}$ *satisfies Assumptions 1 and 2, and a finitely supported vector* $\mathbf{v}$, *then the output* $\mathbf{w}_\epsilon$ *has the following properties:*
P1: $\|\mathbf{w} - \mathbf{w}_\epsilon\| \le \epsilon$.
P2: *For any* $d/\gamma < \tau < 2$ *(see Theorem 3.2),*

$$\#(\mathrm{supp}(\mathbf{w}_\epsilon)) \le C\|\mathbf{v}\|_{{}_t\ell_\tau^w(\mathcal{J})}^{1/s}\epsilon^{-1/s} + \#(\mathcal{J}_\phi) \tag{3.27}$$

with C *a constant depending only on the constants appearing in Theorem 3.2. Moreover, the number of computations needed to find* $\mathbf{w}_\epsilon$ *is also bounded by the right side of (3.27).*

Proof: Since the vector $\mathbf{v}$ is finite, it belongs to all ${}_t\ell_\tau^w(\mathcal{J})$. From (3.12) of Theorem 3.2, we have $\#(\Lambda_j) \lesssim 2^{j\tau} + \#(\mathcal{J}_\phi)$ from which it follows that ϵ_j tends to 0 as $j \to \infty$. Therefore, the algorithm must terminate at some finite value j^*. Now every μ which is not in $\bar\Lambda_\epsilon$ is in Γ_ν for some $\nu \in \mathcal{L}_{j^*}^-$. It follows from (3.11) that

$$\|\mathbf{w} - \mathbf{w}_\epsilon\|^2 \le \sum_{\nu\in\mathcal{L}_{j^*}^-} |\tilde{\mathbf{w}}_\mu|^2 \le C_0^2\#(\mathcal{L}_{j^*}^-)\eta_{j^*}^2 = \epsilon_{j^*}^2 \le \epsilon^2, \tag{3.28}$$

which proves bf P1.

To prove **P2**, we start with (3.12) which gives

$$(3.29)\ \#(\bar\Lambda_{j^*}) \ \lesssim\ \|\mathbf{v}\|^{\tau}_{t\ell^w_\tau(\mathcal{J})}\eta_{j^*}^{-\tau} + \#(\mathcal{J}_\phi) = C_0^{\tau}\|\mathbf{v}\|^{\tau}_{t\ell^w_\tau(\mathcal{J})}\epsilon_{j^*}^{-\tau}(\#(\mathcal{L}_{j^*}^{-}))^{\tau/2} + \#(\mathcal{J}_\phi).$$

We continue on under the assumption that the first term on the far right side of (3.29) is bigger than the second since otherwise we are done. We recall now that $\#(\bar\Lambda_{j^*}) \sim \#(\mathcal{L}_{j^*}^{-})$ and $\epsilon_{j^*} \ \gtrsim\ \epsilon_{j^*-1} \geq \epsilon$ because the sets $\bar\Lambda_j$ increase when j increases. Using this information back in (3.29) gives

$$\#(\bar\Lambda_{j^*})^{1-\tau/2} \ \lesssim\ \|\mathbf{v}\|^{\tau}_{t\ell^w_\tau(\mathcal{J})}\epsilon^{-\tau} \tag{3.30}$$

Therefore **P2** follows by raising both sides of (3.30) 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 number of computations N_0 used at iteration $j = 0$ of the algorithm is bounded by $\#(\bar\Lambda_0)$ and

$$\#(\bar\Lambda_0) \ \lesssim\ \|\mathbf{v}\|^{\tau}_{t\ell^w_\tau(\mathcal{J})}\eta_0^{-\tau} + \#(\mathcal{J}_\phi). \tag{3.31}$$

At each iteration $j \geq 1$ the number of computations N_j needed to execute this iteration from the computations already in hand at iteration $j-1$ is bounded by

$$N_j \ \lesssim\ \|\mathbf{v}\|^{\tau}_{t\ell^w_\tau(\mathcal{J})}\eta_j^{\tau}. \tag{3.32}$$

So the total number N of computations is bounded by

$$N \ \lesssim\ \|\mathbf{v}\|^{\tau}_{t\ell^w_\tau(\mathcal{J})}\eta_{j^*}^{-\tau} + \#(\mathcal{J}_\phi). \tag{3.33}$$

This is the same estimate as we had for $\bar\Lambda_{j^*}$ in (3.29) and therefore we derive **P3** in the same way we derived **P2** from (3.29). □

4. Verification of the basic assumptions. We shall show in this section that the 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

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

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*

$$0 \leq p \leq p^* := \frac{d+2t}{d-2t}, \tag{4.2}$$

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

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

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

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

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

The first term is bounded according to

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

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

$$\int_\Omega |\varphi||u|^p \leq \|\varphi\|_{L_q} \|u\|_{L_{pq'}}^p, \tag{4.6}$$

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

$$\int_\Omega |\varphi||u|^p \leq \|\varphi\|_{L_{p+1}} \|u\|_{L_{p+1}}^p. \tag{4.7}$$

We then remark 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

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

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

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

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

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

The first term is simply bounded by

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$\begin{aligned} |w_\lambda| &= |\langle w, \psi_\lambda\rangle| = \inf_{P\in\Pi_{r-1}} |\langle w - P, \psi_\lambda\rangle| \\ &\lesssim |w|_{W^r(L_\infty(S_\lambda))} 2^{-r|\lambda|} 2^{-(t+d/2)|\lambda|} = |w|_{W^r(L_\infty(S_\lambda))} 2^{-\gamma|\lambda|}, \end{aligned} \tag{4.18}$$

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

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

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

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

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

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

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

$$\begin{aligned} \|\mathcal{F}^{(k)}(u)\|_{L_\infty(S_\lambda)} \prod_{i=1}^k \|D^{\beta_i}u\|_{L_\infty(S_\lambda)} &\lesssim \|\mathcal{F}^{(k)}(u)\|_{L_\infty(S_\lambda)} \|u\|_{L_\infty(S_\lambda)}^{k-1} \\ &\quad \times |u|_{W^r(L_\infty(S_\Lambda))}. \end{aligned} \tag{4.22}$$

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

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

$$|w|_{W^r(L_\infty(S_\lambda))} \lesssim \|u\|^M_{L_\infty(S_\lambda)} |u|_{W^r(L_\infty(S_\lambda))} \tag{4.23}$$

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 $B^s_\infty(L_\infty)$, $s \geq 0$. They satisfy the norm equivalences

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

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

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

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

$$\begin{aligned} |w|_{W^r(L_\infty(S_\lambda))} &\lesssim \left(\sup_{S_\mu \cap S_\lambda \neq \emptyset} 2^{(\delta+\varepsilon)|\mu|} |u_\mu|\right)^M \left(\sup_{S_\mu \cap S_\lambda \neq \emptyset} 2^{(r+\delta+\varepsilon)|\mu|} |u_\mu|\right) \\ &= \left(2^{(\delta+\varepsilon)|\mu_0|} |u_{\mu_0}|\right)^M \left(2^{(r+\delta+\varepsilon)|\mu_1|} |u_{\mu_1}|\right) \end{aligned} \tag{4.26}$$

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

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

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

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

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

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

$$\begin{aligned} |w|_{W^r(L_\infty(S_\lambda))} &\lesssim |u|_{W^r(L_\infty(S_\lambda))} \lesssim \|u\|_{B^{r+\varepsilon}_\infty(L_\infty(S_\lambda))} \\ &\lesssim \sup_{S_\mu \cap S_\lambda \neq \emptyset} 2^{(r+\delta+\epsilon)|\mu|} |u_\mu| \lesssim \sup_{S_\mu \cap S_\lambda \neq \emptyset} 2^{(r+d/2+t)\mu|} |u_\mu| \end{aligned} \tag{4.29}$$

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

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

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

acting from $H \times \cdots \times H$ to its dual H' (note that $\alpha_i = (\alpha_{i,1}, \cdots, \alpha_{i,d})$ are multi-indices). These include multilinear operators as particular cases. Here we shall indicate the apropriate generalizations of the results in the two previous sections with brief sketches of proofs since they are quite similar but notationally quite 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, $\mathbf{u} = (\mathbf{u}_1, \cdots, \mathbf{u}_n)$ and $\mathbf{F}$ the corresponding discrete mapping

$$\mathbf{F}(\mathbf{u}) := (\langle \mathcal{F}(D^{\alpha_1} u_1, \cdots, D^{\alpha_n} u_n), \psi_\lambda \rangle)_{\lambda \in \mathcal{J}}, \tag{5.2}$$

we introduce the following generalization of the basic assumptions.

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

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

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

The local version of this stability assumption now reads

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

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*

$$|w_\lambda| \le C \sup_{\mu \,:\, S_\lambda \cap S_\mu \neq \emptyset} [\sum_{i=1}^{n} |u_{i\mu}|] 2^{-\gamma(|\lambda| - |\mu|)}], \tag{5.5}$$

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 non-decreasing function.*

In order to predict the significant coefficients of $\mathbf{w} = \mathbf{F}(\mathbf{u})$, we fix $\eta > 0$ and for the constant γ of (5.5), we define for all $\mu \in \mathcal{J}$ the number $n(\mu)$ satisfying

$$\eta 2^{\gamma n(\mu)} \le \max_i |\tilde{u}_{i,\mu}| < \eta 2^{\gamma(n(\mu)+1)}, \tag{5.6}$$

where $\tilde{u}_{i,\mu}$ are the residuals for $\mathbf{u}_i$ (see (2.20)) and we define the influence set $\Lambda_{\eta,\mu}$ in a similar way as in (3.8). The set of coefficients for the approximation of $\mathbf{w}$ is now defined by $\Lambda_\eta := \mathcal{J}_\phi \cup \tilde{\Lambda}_\eta$ with

$$\tilde{\Lambda}_\eta := \cup_{\mu \in \tilde{\mathcal{T}}_\eta} \Lambda_{\eta,\mu}, \tag{5.7}$$

where $\tilde{\mathcal{T}}_\eta = \cup_{i=1}^{n} \tilde{\mathcal{T}}_\eta(\mathbf{u}_i)$ and $\tilde{\mathcal{T}}_\eta(\mathbf{u}_i)$ is the expansion of the tree $\mathcal{T}_\eta(u_i)$ (as before $\mathcal{T}_\eta(u_i)$ consists of all indices μ such that $\tilde{u}_{i,\mu} > \eta$).

THEOREM 5.1. *With this definition of* Λ_η *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 3.2 also hold.*

Sketch of proof: As in the proof of Theorem 3.2, in order to prove (3.11) 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 (3.8), (5.7) and (5.5), we obtain that for any $\lambda \notin \Lambda_\eta$, we have $|\tilde{w}_{\lambda,\eta}| \lesssim \eta$. We then use (5.4) in a similar way in order to derive (3.11) .

In order to prove (3.12) , we use the trees $\tilde{\mathcal{T}}_j := \tilde{\mathcal{T}}_{\eta 2^{\gamma j}}$, in order to decompose Λ_η into layers indexed by j as in the proof of Theorem 3.2. We then proceed in a similar way to derive 3.12, remarking that

$$\#(\tilde{\mathcal{T}}_j) \lesssim \eta^{-\tau} 2^{-\gamma\tau j} \sup_i \|\mathbf{u}_i\|^\tau_{{}_t\ell_\tau^w(\mathcal{J})}, \tag{5.8}$$

and that according to (5.6) and (3.8), $n(\mu) = j$ for $\mu \in \tilde{\mathcal{T}}_j \setminus \tilde{\mathcal{T}}_{j+1}$.

Finally, we prove

$$(5.9) \qquad \|\mathbf{w}\|_{{}_t\ell^w_\tau(\mathcal{J})} \lesssim 1 + \sup_i \|\mathbf{u}_i\|_{{}_t\ell^w_\tau(\mathcal{J})},$$

by the same arguments as in the proof for Theorem 3.2. □

We can also generalize the algorithm of §3.4 which ensures the target accuracy $\varepsilon > 0$ for $\mathbf{w} = \mathbf{F}(\mathbf{u})$:

- For $j = 0, 1, \ldots$, define the threshold $\eta_j = 2^{-j}$ and compute $\Lambda_j := \Lambda_{\eta_j}$ according to 5.7. Define $\mathcal{L}_j^-$ as the set of outer leaves of Λ_j.
- According to Theorem 5.1, we have $\|\mathbf{w} - \mathbf{w}_{|\Lambda_j}\|^2 \lesssim \#(\mathcal{L}_j^-)\eta^2 := \varepsilon_j^2$. Stop for the smallest j such that $\varepsilon_j \le \varepsilon$ and define $\Lambda_\varepsilon = \Lambda_j$.
- Define the corresponding approximation $\mathbf{w}_\varepsilon := \mathbf{w}_{|\Lambda_\varepsilon}$.

This strategy exhibits asymptotically optimal complexity summarized in the following proposition which is a direct consequence of Theorem 5.1.

PROPOSITION 5.1. *If $\mathbf{u} \in ({}_t\ell^w_\tau(\mathcal{J}))^n$ we have the estimate*

$$(5.10) \qquad \#(\Lambda_\varepsilon) \lesssim \sup_i \|\mathbf{u}_i\|^{\tau/2}_{{}_t\ell^w_\tau(\mathcal{J})} \varepsilon^{-1/s} + \#(\mathcal{J}_\phi),$$

with $s = 1/\tau - 1/2$.

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

$$(5.11) \qquad |D^\beta \mathcal{F}(x_1, \cdots, x_n)| \le C \prod_{i=1}^n (1 + |x_i|)^{[p_i - \beta_i]_+}, \quad |\beta| = 0, 1, \cdots, n^*,$$

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

$$(5.12) \qquad \mathcal{F}(u) = \mathcal{F}(D^{\alpha_1}u_1, \cdots, D^{\alpha_n}u_n), \quad \text{with} \quad u = (u_1 \cdots, u_n).$$

We then obtain the following generalization of Proposition 4.1

PROPOSITION 5.2. *Assume that the growth assumptions (5.11) hold at least with $n^* = 0$. Then $\mathcal{F}$ maps $H \times \cdots \times H$ to H' whenever $H = H^t$ and $t \ge 0$ satisfies*

$$(5.13) \qquad [\frac{1}{2} - \frac{t}{d}]_+ + \sum_{i=1}^n p_i[\frac{1}{2} - \frac{t}{d} + \frac{|\alpha_i|}{d}]_+ < 1.$$

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

$$(5.14) \qquad \|\mathcal{F}(u) - \mathcal{F}(v)\|_{H'} \le C \sum_{i=1}^n \|u_i - v_i\|_H,$$

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

$$(5.15) \qquad |\langle \mathcal{F}(u), \varphi \rangle| \le C \int_\Omega |\varphi| \prod_{i=1}^n (1 + |D^{\alpha_i}u_i|)^{p_i}.$$

In view of (5.13), 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

$$(5.16) \qquad \frac{1}{r} > \frac{1}{2} - \frac{t}{d} \quad \text{and} \quad \frac{1}{r_i} > \frac{1}{2} - \frac{t}{d} + \frac{|\alpha_i|}{d}.$$

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

$$|\langle \mathcal{F}(u), \varphi\rangle| \leq C\|\varphi\|_{L_r} \prod_{i=1}^{n}(1+\|D^{\alpha_i}u_i\|_{L_{r_i}}^{p_i}), \tag{5.17}$$

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

$$\|\mathcal{F}(u)\|_{H'} \leq C \prod_{i=1}^{n}(1+\|D^{\alpha_i}u_i\|_{H}^{p_i}). \tag{5.18}$$

For the stability property, we use the inequality

$$|\mathcal{F}(u)-\mathcal{F}(v)| \leq C\sum_{i=1}^{n}|D^{\alpha_i}u_i - D^{\alpha_i}v_i| \prod_{k=1}^{n}(1+|D^{\alpha_k}u_k|+|D^{\alpha_k}v_k|)^{[p_k-\delta_{i,k}]_+}, \tag{5.19}$$

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

$$E_i = \int_\Omega |\varphi||D^{\alpha_i}u_i - D^{\alpha_i}v_i| \prod_{k=1}^{n}(1+|D^{\alpha_k}u_k|+|D^{\alpha_k}v_k|)^{[p_k-\delta_{i,k}]_+}, \tag{5.20}$$

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

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

whenever

$$\frac{1}{r}+\frac{1}{q}+\sum_{k=1}^{n}\frac{[p_k-\delta_{i,k}]_+}{r_k} = 1. \tag{5.21}$$

In view of (5.13), we can choose positive numbers r, q and r_i satisfying condition (5.21) such that

$$\frac{1}{r} > \frac{1}{2}-\frac{t}{d}, \quad \frac{1}{q} > \frac{1}{2}-\frac{t}{d}+\frac{|\alpha_i|}{d} \quad \text{and} \quad \frac{1}{r_k} > \frac{1}{2}-\frac{t}{d}+\frac{|\alpha_k|}{d}. \tag{5.22}$$

Therefore, the Sobolev embedding gives

$$E_i \leq C\|\varphi\|_H\|u_i - v_i\|_H, \tag{5.23}$$

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 (5.13) 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

$$|D^\beta \mathcal{F}(x_1,\cdots,x_n)| \leq C\prod_{i=1}^{n}(1+|x_i|)^{p_i-\beta_i}, \quad \beta_i \leq p_i, \tag{5.24}$$

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

THEOREM 5.2. *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 (5.11) 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 (5.24) 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

$$|w_\lambda| \lesssim |w|_{W^r(L_\infty(S_\lambda))} 2^{-\gamma|\lambda|}, \tag{5.25}$$

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

$$D^\nu \mathcal{F}(D^{\alpha_1} u_1, \cdots, D^{\alpha_n} u_n) G_\nu, \quad |\nu| = 1, \ldots, r, \tag{5.26}$$

where

$$G_\nu = \prod_{i=1}^n \prod_{j=1}^{\nu_i} D^{\beta_{i,j} + \alpha_i} u_i, \tag{5.27}$$

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

$$|w|_{W^r(L_\infty(S_\lambda))} \lesssim \max_{|\nu| \leq r} A_\nu \tag{5.28}$$

where

$$A_\nu := \|D^\nu \mathcal{F}(D^{\alpha_1} u, \cdots, D^{\alpha_n} u_n)\|_{L_\infty(S_\lambda)} \|G_\nu\|_{L_\infty(S_\lambda)}. \tag{5.29}$$

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

$$\|G_\nu\|_{L_\infty(S_\lambda)} \lesssim \prod_{i=1}^n |u_i|^{\nu_i - 1}_{W^{|\alpha_i|}(L_\infty(S_\lambda))} |u_i|_{W^{r_i + |\alpha_i|}(L_\infty(S_\lambda))}. \tag{5.30}$$

We now invoke (5.11) and obtain that

$$\begin{aligned} A_\nu &\leq \prod_{i=1}^n (1 + |u_i|_{W^{|\alpha_i|}(L_\infty(S_\lambda))})^{(p_i - \nu_i)_+} |u_i|^{\nu_i - 1}_{W^{|\alpha_i|}(L_\infty(S_\lambda))} |u|_{W^{r_i + |\alpha_i|}(L_\infty(S_\lambda))} \\ &\leq \prod_{i=1}^n |u_i|^{M_i}_{W^{|\alpha_i|}(L_\infty(S_\lambda))} |u|_{W^{r_i + |\alpha_i|}(L_\infty(S_\lambda))} \end{aligned} \tag{5.31}$$

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 (5.31) can be bounded by Besov norms. The arguments used in deriving (4.28) and (4.29) gives

$$|u_i|^{M_i}_{W^{|\alpha_i|}(L_\infty(S_\lambda))} |u|_{W^{r_i + |\alpha_i|}(L_\infty(S_\lambda))} \lesssim \|\mathbf{u}\|^{M_i} 2^{(r_i + (M_i + 1)(|\alpha_i| + \delta + \epsilon))|\mu_i|} |u_{i,\mu_i}| \tag{5.32}$$

where μ_i is a maximizing index. Let $\mu^* := \max_i \mu_i$. We place (5.32) into (5.31). Each term $|u_{i,\mu_i}|$, $\mu_i \neq \mu^*$, we pull out of the product by the majorant $\|\mathbf{u}\|$. This then gives

$$A_\nu \lesssim \prod_{i=1}^n \|\mathbf{u}\|^{M_i} 2^{(r_i + (M_i + 1)(|\alpha_i| + \delta + \epsilon))|\mu_i|} |u_{i,\mu_i}| \lesssim \|\mathbf{u}\|^M [\sum_{i=1}^n |u_{i,\mu^*}|] 2^{\tilde{\gamma}|\mu^*|}]. \tag{5.33}$$

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

$$\tilde{\gamma} = r + \sum_{i=1}^{n} (1 + M_i)(\varepsilon + [d/2 - t + |\alpha_i|]_+). \tag{5.34}$$

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 (5.34) shows that

$$\tilde{\gamma} \leq r + \sum_{i=1}^{n} p_i(\varepsilon + [d/2 - t + |\alpha_i|]_+) \leq \gamma \tag{5.35}$$

provided ϵ is sufficiently small. □

REFERENCES

[1] S. Bertoluzza, *Adaptive wavelet collocation for the solution of steady state equation*, in: SPIE Proc. Wavelet Appl. II., 2491 (1995).

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

[3] G. Bourdaud and D. Kateb, *Fonctions qui opèrent sur les espacs de Besov*, Math. Ann., 303 (1995), 653–675.

[4] C. Canuto and I. Cravero, *Wavelet-based adaptive methods for advection-diffusion problems*, Math. Mod. Meth. Appl. Sci., 7 (1997), 265-289.

[5] C. Canuto, A. Tabacco, K. Urban, The wavelet element method, part I: Construction and analysis, Appl. Comp. Harm. Anal., 6(1999), 1–52.

[6] A. Cohen, R. Masson, Wavelet adaptive methods for second order elliptic problems, boundary conditions and domain decomposition, Preprint 1997.

[7] A. Cohen, W. Dahmen, I. Daubechies, R. DeVore, Tree approximation and optimal encoding, Appl. Comp. Harm. Anal., 11 (2001), 192–226.

[8] A. Cohen, W. Dahmen, R. DeVore, Adaptive wavelet methods for elliptic operator equations – Convergence rates, Math. Comp. 70 (2001), 27–75.

[9] A. Cohen, W. Dahmen, R. DeVore, Adaptive wavelet methods II - Beyond the elliptic case, IGPM Report, RWTH Aachen, Nov. 2000, to appear in FOCM Journal.

[10] A. Cohen, W. Dahmen, R. DeVore, Adaptive wavelet methods for nonlinear variational problems, in preparation.

[11] A. Cohen, S.M. Kaber, S. Müller, M. Postel, Fully adaptive multiresolution schemes for conservation laws, to appear in Mathematics of Computation, Report No. 00009, May 2000, Laboratoire d'Analyse Numerique, Paris VI

[12] W. Dahmen, Adaptive approximation by smooth multivariate splines, J. Approx. Theory, 36 (1982), 119-140.

[13] W. Dahmen and R. Schneider, Composite Wavelet Bases for Operator Equations, Math. Comp., 68 (1999), 1533–1567.

[14] W. Dahmen and R. Schneider, Y. Xu, Nonlinear functions of wavelet expansions – Adaptive reconstruction and fast evaluation, Numerische Mathematik, 86(2000), 49–101.

[15] R. DeVore, Nonlinear Approximation, *Acta Numerica*, 7 (1998), Cambridge University Press, 51–150.

[16] R. DeVore, P. Petruchev, X.M. Yu, Wavelet approximation in the space C, in: Progress in Approximation Theory, Springer Verlag, New York, 1992.

[17] M. Farge, M. Griebel, F. Koster and K. Schneider, *Adaptive wavelet methods for the Navier-Stokes equations* Lecture Notes in Physics (ED. E. H. Hirschel), Springer (2001), 303–318.

[18] M. Farge and K. Schneider, *Numerical simulation of a mixing layer in adaptive wavelet basis*, CRAS Paris, Série II b, 328 (2000), 263–269.

[19] A. Harten, *Adaptive multiresolution schemes for shock computations*, J. Comp. Phys. 115 (1994), 319-338.

[20] Y. Maday, V. Perrier and J.C. Ravel, *Adaptativité dynamique sur bases d'ondelettes pour l'approximation d'équations aaux dérivées partielles*, CRAS Paris, Série I, 1 (1991), 405-410.

[21] T. Runst, W. Sickel, *Sobolev Spaces of Fractional Order, Nemitskij Operators, and Nonlinear Partial Differential Equations*, de Gruyter Series in Nonlinear Analysis and Applications, 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), 303–327.

[23] M. Verani, *An adaptive Newton scheme based on nonlinear wavelet approximation*, Preprint IAN, Pavia, 2002.