ADAPTIVE WAVELET SCHEMES FOR NONLINEAR VARIATIONAL PROBLEMS*

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

Abstract. We develop and analyze wavelet based adaptive schemes for nonlinear variational problems. We derive estimates for convergence rates and corresponding work counts that turn out to be asymptotically optimal. Our approach is based on a new paradigm that has been put forward recently for a class of linear problems. The original problem is transformed first into an equivalent one which is well posed in the Euclidean metric ℓ_2. Then conceptually one seeks iteration schemes for the infinite dimensional problem that exhibits at least a fixed error reduction per step. This iteration is then realized approximately through an adaptive application of the involved operators with suitable dynamically updated accuracy tolerances. The main conceptual ingredients center around nonlinear tree approximation and the sparse evaluation of nonlinear mappings of wavelet expansions. We prove asymptotically optimal complexity for adaptive realizations of first order iterations and of Newton's method.

Key Words: Variational problems, wavelet representations, semilinear equations, mapping properties, gradient iteration, convergence rates, adaptive application of operators, sparse evaluation of nonlinear mappings of wavelet expansions, tree approximation, Newton's scheme.

AMS Subject Classification: 65J15, 65N12, 65N15, 35A15, 35A35, 35J60, 41A60, 46A45, 47H17

1. Introduction.

1.1. Background and Objectives. Adaptive wavelet schemes for numerically solving a wide class of variational problems have been recently studied in [7, 8] from the perspective of asymptotic estimates for convergence rates and corresponding work counts. The problems covered by this analysis include elliptic boundary integral equations, elliptic boundary value problems, but also indefinite problems of elliptic type such as the Stokes problem. Two requirements were essential in this context: (i) the variational problem induces an operator $\mathcal{L}$ that is an isomorphism from some Hilbert space $\mathcal{H}$ onto its dual; (ii) this Hilbert space permits a wavelet characterization, i.e., the $\mathcal{H}$-norm of an element is equivalent to a weighted ℓ_2-norm of its wavelet coefficients. It could then be shown that certain adaptive schemes exhibit an *asymptotically optimal accuracy/work balance* within a certain range of convergence rates depending on the choice of wavelet bases. The precise meaning of this statement is explained in the Meta-theorem below. To our knowledge for the above range of linear problems such convergence/complexity estimates have been established so far only for wavelet methods. Just recently, a similar result has been proved for adaptive finite element methods for the more restricted class of second order elliptic boundary value problems [4].

*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/

In this paper we wish to exlore the convergence rates and the computational complexity of certain new adaptive wavelet schemes for *nonlinear problems* for which no results of the above type seem to be known so far.

Our primary concern here is *not* to develop a specific algorithm for a concrete application. We are rather interested in developing a numerically realizable *new algorithmic paradigm* in a fairly general context of nonlinear problems and in analyzing its principal complexity features. Therefore the various algorithmic ingredients will at times not be discussed in full detail but only to an extent that clarifies their principal asymptotic complexity.

The new paradigm is based upon the adaptive evaluation of (linear and nonlinear) operators in the course of an ideal iteration for the *infinite dimensional* problem formulated in the wavelet coordinate domain. Such perturbed iterations will lead to an algorithm **SOLVE** that (with a proper initialization) produces for any target accuracy ϵ a finitely supported vector of coeffcients $\bar{\mathbf{u}}(\epsilon)$ that approximates the array of wavelet coefficients of the exact solution (of the underlying variational problem) in ℓ_2 with accuracy ϵ. The choice of wavelet basis will then imply that the corresponding finite expansion approximates the exact solution with accuracy $C\epsilon$ in the energy norm where C depends only on the wavelet basis. In order to identify the essential mechanisms governing such schemes, we will consider nonlinear variational problems on various levels of generality. The results will be purely asymptotic in nature. They reveal asymptotically optimal work/accuracy balances interrelating the achieved target accuracy with the required computational work and associated adaptively generated number of degrees of freedom. More precisely, we shall prove results of the following type.

Meta-Theorem *If the exact solution can be approximated as a linear combination of N wavelets (subject only to certain tree restrictions on the distribution of active coefficients) to accuracy of order N^{-s}, (for a certain range of s), then the support of the output $\bar{\mathbf{u}}(\epsilon)$ of* **SOLVE** *for target accuracy ϵ grows at most like $\epsilon^{-1/s}$, uniformly in ϵ, and the computational complexity stays also proportional to the support size. In this sense the scheme tracks the exact solution at asymptotically minimal cost.*

Note that the above mentioned *tree restriction* on the permitted distribution of active coeffcients is the analogue of locally refined meshes in the finite element context.

We shall outline now how we approach results of the above type.

1.2. The Basic Paradigm. The *classical* approach to numerically solving (linear and nonlinear) variational problems is concerned with the following issues:

(c1) Well-posedness of the given variational problem;
(c2) discretization of the infinite dimensional problem so as to obtain a finite system of algebraic equations;
(c3) well-posedness of the finite system of equations and error analysis;
(c4) numerical solution of the finite system of equations.

It is important to note that (c1) is often hidden in the analysis and that (c3) is, in general, *not* a direct consequence of (c1). Typical examples even in the linear case are *saddle point problems*. It is well known that, for Galerkin discretizations to be stable, the trial spaces for the different solution components have to satisfy a certain compatibility condition (LBB-condition). For nonlinear problems one can often establish only *local* uniqueness of solutions so that some care is required to ensure that the discrete problems approximate the correct solution branch. Thus, the discrete problems do not necessarily inherit the "nice properties" of the original infinite dimensional problem. Depending on the choice of the discretization one might introduce "new difficulties". The typical obstructions encountered in (c4) are the *large size* of the discrete systems and possible *ill-conditioning*. The latter issue interfers with the need to resort to iterative solvers, due to the size and sparseness of the systems. Attempts to reduce computational complexity are often based on adaptive and hence

possibly economic discretizations. A reliable control of adaptive refinements, however, depends usually in a sensitive way on the particular type of the problem and rigorous complexity estimates are generally not available.

A *new paradigm* has been explored in [8] for *linear variational problems*. It aims at closely intertwining the analysis - discretization - solution process. The basic steps there read as follows:

(n1) Well-posedness of the given variational problem;
(n2) transformation of the infinite dimensional problem into an *equivalent* problem in ℓ_2 which is *well posed* in the Euclidean metric;
(n3) the derivation of an iterative scheme for the infinite dimensional ℓ_2-problem that exhibits a fixed error reduction per iteration step;
(n4) numerical realization of the iterative scheme by an *adaptive application* of the involved infinite dimensional operators within some finite dynamically updated accuracy tolerances.

Thus, the starting point (n1) is the same as (c1), although it takes a somewhat more exposed and explicit role in the new setting, as will be explained later. The main difference is that one aims at staying as long as possible with the infinite dimensional problem, hopefully, given in a favorable format. Only at the very end, when it comes to applying the operators in the ideal iteration scheme (n4), one enters the finite dimensional realm. However, the finite number of degrees of freedom is determined at each stage by the adaptive application of the operator, so that at *no* stage is any specific trial space fixed. Roughly speaking, the "nice properties" of the infinite dimensional problem are preserved through adaptive evaluations. In fact, one can show that thereby compatibility conditions like the LBB-condition indeed become void [8, 12].

The main goal of the present paper is to show how to carry over this paradigm, already existing for linear problems to the nonlinear setting. One then encounters three major issues, namely:

(a) the choice of tolerances in (n4) to ensure that the perturbed iteration converges to the correct solution;
(b) the design of economic approximate application schemes for the possibly nonlinear infinite dimensional operators;
(c) estimating the complexity of the scheme.

Here (a) means that any given *target accuracy* ϵ is achieved after finitely many steps. (b) is the most crucial part and will be discussed in detail in the course of the paper. Clearly (b) is closely related to (c). As in [7, 8, 12] we will measure complexity by the *number of adaptively generated degrees of freedom* $N = N(\epsilon)$ required by the adaptive scheme to achieve the target accuracy ϵ and the corresponding number of floating point operations (which, of course, is aimed at staying proportional to $N(\epsilon)$). Estimating the asymptotic *work/accuracy balance* $N(\epsilon) \leftrightarrow \epsilon$ will be a central theme in the subsequent developments. This part differs significantly from the classical error analysis for FEM and resides on concepts from *harmonic analysis* and *nonlinear approximation*.

Finally, a comment on (n3) is in order. Aiming at a fixed error reduction per iteration step means that one is content with a *first order* scheme. So why not go for faster iteration schemes? The answer to this question is not completely clear. Indeed, a higher order method may not automatically win for the following reason. Usually a higher order method is more costly in function evaluations. In the present context this means, according to (n4), it is more costly in the adaptive application of the full infinite dimensional operators within some dynamically updated accuracy tolerance. Preserving the higher order of the ideal iteration also in its perturbed form in connection with the higher demands of function evaluations may very well increase the cost of each iteration step so as to offset the potential gain of a better error reduction. So with regard to the objective of reaching a target accuracy at possibly

low overall computational cost, the role of higher order schemes remains unclear. In fact, it will be seen that *asymptotic optimality* can indeed be achieved already with simple *first order outer iterations*. Nevertheless, we shall show that it is also possible to retain second order convergence of the adaptive version of Newton's scheme so as to arrive at an overall scheme with asymptotically optimal solution complexity, which may offer quantitative advantages over the first order versions.

1.3. Organization of Material. The paper is organized as follows. In Section 2 we describe (n1), (n2) and (n3) for a general setting that will host all subsequent specifications. In Section 3 we distinguish three classes of variational problems to which the subsequent developments will refer frequently, namely (L) *linear* problems, (SL) semilinear elliptic problems and (GNL) more general nonlinear problems where we have to assume the existence of locally unique solutions. In Section 4 we formulate the prototype of an adaptive perturbed first order iteration which is based on two main ingredients, namely approximate *residual evaluations* and a certain *coarsening scheme*. In particular, the residual approximations involve the *adaptive* application of linear or nonlinear (infinite dimensional) operators. Assuming at this stage that these ingredients are indeed available, we address for the most general setting first only issue (a) to clarify for which choice of dynamically updated accuracy tolerances is convergence guaranteed. The remaining sections will be devoted issues (b) and (c) for the problem types (L), (SL) and (GNL).

In Section 5 we review briefly concrete realizations of these ingredients for the linear case (L) and indicate the concepts for their complexity analysis. This serves two purposes. First, these results will be used in the last section in connection with Newton iterations. Second, they motivate our treatment of the nonlinear case. In Section 6 we introduce some new concepts needed to deal with nonlinear problems. They center upon *tree approximation* and related coarsening techniques. This enables us to formulate the notion of τ^*-sparsity as the key criterion for controlling the complexity of the adaptive schemes in the nonlinear case. Drawing on several results from [9] we develop in Section 7 adaptive evaluation schemes that are proven to be τ^*-sparse and thus lead to asymptotically optimal results in the sense of the above **Meta-Theorem**. To our knowledge these are the first convergence and complexity estimates for adaptive solvers for nonlinear problems. Finally, in Section 8 we develop an adaptive Netwon scheme and analyze its complexity. It differs in essential ways from the schemes discussed in the previous sections which are based on first order iterations. In particular, we show that the quadratic convergence of the outer iteration can in some sense be preserved in the adaptive context.

After completion of this paper we became aware of related work concerning the convergence of Newton's method in the wavelet coordinate domain [23]. However, in this latter work no complexity estimates are given and the convergence is asserted under the *assumption* that the nonlinearity is stable on appropriate sequence spaces.

2. The Setting. We describe now the setting for which the above paradigm will be discussed.

2.1. The General Problem Format. The variational problems mentioned in step (n1) above will always have the following format. Let $\mathcal{H}$ be a Hilbert space with norm $\|\cdot\|_{\mathcal{H}}$ and let $\mathcal{H}'$ denote its dual endowed with the norm

$$\|v\|_{\mathcal{H}'} := \sup_{w\in\mathcal{H}} \frac{\langle v, w\rangle}{\|w\|_{\mathcal{H}}},$$

where $\langle\cdot,\cdot\rangle$ is the dual pairing between $\mathcal{H}$ and $\mathcal{H}'$ (with respect to L_2 as the pivot space). Suppose that $f \in \mathcal{H}'$ and

$$F : \mathcal{H} \mapsto \mathcal{H}' \tag{2.1}$$

is a (possibly nonlinear) mapping. We consider the numerical solution of the problem: Find $u \in \mathcal{H}$ such that

$$\langle v, F(u) - f\rangle =: \langle v, R(u)\rangle = 0, \qquad v \in \mathcal{H}. \tag{2.2}$$

The objective in (n1) is the identification of a suitable space $\mathcal{H}$, so that (2.2) is well posed in the following sense. Recall that the Frechét derivative $DR(z) = DF(z)$ is a mapping from $\mathcal{H}$ to $\mathcal{H}'$, defined by the duality

$$\langle v, DR(z)w\rangle = \lim_{h\to 0} \frac{1}{h}\langle v, R(z+hw) - R(z)\rangle. \tag{2.3}$$

The problem (2.2) is called *well–posed* if R has the following properties

A1: R possesses a continuous Frechét derivative, i.e., $R \in C^1(\mathcal{H}, \mathcal{H}')$ as a mapping $v \to R(v)$.

A2: There exists a solution $u \in \mathcal{H}$ to (2.2) and in addition to (2.1) the Frechét derivative DF of F at v in some neighborhood $\mathcal{U}$ of u is an isomorphism from $\mathcal{H}$ onto $\mathcal{H}'$.

Clearly A2 ensures that the solution u is locally unique.

2.2. Wavelet Coordinates and an Equivalent ℓ_2- Problem. The transformations for (n2) will be based on suitable wavelet bases. For a detailed discussion of such bases, we refer the reader to the literature (see e.g. [5, 6, 13, 10]) and collect here only the relevant facts. A *wavelet basis* $\Psi = \{\psi_\lambda : \lambda \in \mathcal{J}\} \subset \mathcal{H}$ has the following properties: The indices $\lambda \in \mathcal{J}$ encode typical information about the wavelet ψ_λ, namely its type its location $k(\lambda)$ and its scale $|\lambda|$.

We shall now explain the meaning of "suitable" in the present context. We will always assume that the wavelets have compact support which scales as $\operatorname{diam}(\operatorname{supp}\psi_\lambda) \sim 2^{-|\lambda|}$. Abbreviating in the following $S_\lambda := \operatorname{supp}\psi_\lambda$, $\lambda \in \mathcal{J}$, we require the collection $\mathcal{S}(\Psi) := \{S_\lambda : \lambda \in \mathcal{J}\}$ of wavelet supports to have the following *finite incidence property* (FIP): There exists an $M \in \mathbb{N}$ such that, whenever $\mathcal{G} \subset \mathcal{S}(\Psi)$ has the property that no element of $\mathcal{G}$ is contained in any other element of $\mathcal{G}$, one has

$$\bigcap_{\lambda\in\mathcal{G}'} S_\lambda \neq \emptyset \ \text{ for some } \ \mathcal{G}' \subseteq \mathcal{G} \ \text{ implies } \ \#\mathcal{G}' \leq M, \tag{2.4}$$

i.e. only a finite number of supports in such a $\mathcal{G}$ overlap at any given point. The (FIP) is known to hold for essentially all known wavelet bases, see [9].

Furthermore, aside from finitely many functions $\psi_\lambda, \lambda \in \mathcal{J}_\phi \subset \mathcal{J}$, $|\lambda| = j_0$, representing the coarsest scale j_0, the wavelets ψ_λ, $\lambda \in \mathcal{J}\backslash\mathcal{J}_\phi$, have *vanishing moments* of some order $m \in \mathbb{N}$, i.e., these wavelets are orthogonal to all polynomials of order m.

Finally, each $v \in \mathcal{H}$ has a unique expansion $\sum_{\lambda\in\mathcal{J}} v_\lambda\psi_\lambda$ such that

$$c_1\|\mathbf{v}\|_{\ell_2(\mathcal{J})} \leq \|\sum_{\lambda\in\mathcal{J}} v_\lambda\psi_\lambda\|_{\mathcal{H}} \leq C_1\|\mathbf{v}\|_{\ell_2(\mathcal{J})}, \tag{2.5}$$

holds for some positive constants c_1, C_1, i.e. Ψ forms a *Riesz basis* for $\mathcal{H}$. Note that (unlike the quoted references) we have normalized the wavelets here in the energy space $\mathcal{H}$ associated with the variational problem (2.1), (2.2), i.e. $\|\psi_\lambda\|_{\mathcal{H}} = 1$, $\lambda \in \mathcal{J}$. Again such bases are known whenever $\mathcal{H}$ is a product of Sobolev spaces (or closed subspaces of Sobolev spaces, determined e.g. by homogeneous boundary conditions or vanishing integral means).

In the following, we will always use boldface notation $\mathbf{v}$ to denote the wavelet coefficients of a given function $v \in \mathcal{H}$ with respect to the basis Ψ. It is often viewed as a column vector with respect to some fixed but unspecified ordering of $\mathcal{J}$ (and analogously for $u, w \in \mathcal{H}$), i.e., $v = \mathbf{v}^T\Psi = \sum_{\lambda\in\mathcal{J}} v_\lambda\psi_\lambda$.

Next note that by duality (2.5) implies

$$C_1^{-1}\|(\langle w,\psi_\lambda\rangle)_{\lambda\in\mathcal{J}}\|_{\ell_2(\mathcal{J})} \le \|w\|_{\mathcal{H}'} \le c_1^{-1}\|(\langle w,\psi_\lambda\rangle)_{\lambda\in\mathcal{J}}\|_{\ell_2(\mathcal{J})}. \tag{2.6}$$

We shall make frequent use of the following fact, see e.g. [8].

REMARK 2.1. *Assume that $T:\mathcal{H}\to\mathcal{H}'$ is a linear isomorphism, i.e., there exists positive constants c_T, C_T such that*

$$c_T\|v\|_{\mathcal{H}} \le \|Tv\|_{\mathcal{H}'} \le C_T\|v\|_{\mathcal{H}}. \tag{2.7}$$

Let

$$\mathbf{T} := (\langle\psi_\lambda, T\psi_\nu\rangle)_{\lambda,\nu\in\mathcal{J}}$$

be its wavelet representation. Then one has

$$c_1^2 c_T\|\mathbf{v}\|_{\ell_2(\mathcal{J})} \le \|\mathbf{T}\mathbf{v}\|_{\ell_2(\mathcal{J})} \le C_1^2 C_T\|\mathbf{v}\|_{\ell_2(\mathcal{J})}. \tag{2.8}$$

We can now transform (2.2) into wavelet coordinates. Defining

$$\mathbf{R}(\mathbf{v}) := (\langle\psi_\lambda, R(v)\rangle : \lambda\in\mathcal{J}) \quad \text{whenever} \quad v=\sum_{\lambda\in\mathcal{J}} v_\lambda\psi_\lambda,$$

the orginal problem (2.2) is obviously equivalent to finding $\mathbf{u}\in\ell_2(\mathcal{J})$ so that

$$\mathbf{R}(\mathbf{u}) = \mathbf{0}. \tag{2.9}$$

2.3. The Basic Iteration. According to (n3) we wish to devise an iterative scheme for the problem (2.9) such that each step reduces the current error at least by a fixed rate $\rho<1$. The schemes we shall consider will have the form

$$\mathbf{u}^{n+1} = \mathbf{u}^n - \mathbf{B}_n\mathbf{R}(\mathbf{u}^n), \tag{2.10}$$

where the (infinite, possibly stage dependent) matrix $\mathbf{B}_n$ is yet to be chosen. For instance, $\mathbf{B}_n=\alpha\mathbf{I}$ corresponds to a fixed point or Richardson iteration, while for $\mathbf{B}_n := D\mathbf{R}(\mathbf{u}^n)^{-1}$ (2.10) becomes Newton's method. We shall make frequent use of the fact that the Jacobian $D\mathbf{R}(\mathbf{u}^n)$ is given by

$$D\mathbf{R}(\mathbf{v}) = (\langle\psi_\lambda, DR(v)\psi_\nu\rangle)_{\lambda,\nu\in\mathcal{J}}\,, \tag{2.11}$$

where $DR=DF$ is the Frechét derivative of the mapping R.

We proceed now in discussing several instances of this setting.

3. The Scope of Reference Problems. We shall address the variational problem (2.2) for the following different levels of generality:

(L) Linear well posed problems;

(SL) Semilinear elliptic boundary value problems;

(GNL) General nonlinear problems.

Beginning with (L) will serve two purposes. First it provides a guideline for the treatment of nonlinear problems. Second, it allows us to formulate some prerequisites for the later discussion of Newton's method.

(SL) is a natural extension of (L). Essential features of nonlinear problems are already encountered in (SL). On the other hand, it is specific enough to permit a complete complexity analysis for a *globally convergent* iteration scheme.

Finally, in (GNL) we relax our assumptions on the structure of R to a great extent. We pay for this by making stronger assumptions on initial guesses and being content with *locally convergent first order* iterations on the infinite dimensional level.

We shall exemplify step (n3) for all three cases (L), (SL) and (GNL) in this order. Except for the last Section this will be based on first order iteration schemes for the underlying infinite dimensional problem. It will be seen along the way that it then suffices to employ stationary "preconditioners" $\mathbf{B}_n = \mathbf{B}$ to obtain asymptotically optimal complexity estimates (although more flexible nonstationary choices may well result in quantitative improvements in practical realizations). The use of truly nonstationary $\mathbf{B}_n$ will be necessary only in connection with Newton's method in Section 8.

3.1. Linear Variational Problems (L). In order to guide the discussion of nonlinear problems we briefly review the case of *linear* variational problems. Let $a(\cdot,\cdot) : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$ be a bilinear form such that the operator $\mathcal{A}$ defined by $a(v,u) = \langle v, \mathcal{A}u\rangle$, for all $v \in \mathcal{H}$, is an isomorphism from $\mathcal{H}$ to $\mathcal{H}'$, i.e. there exist positive constants $c_{\mathcal{A}}, C_{\mathcal{A}}$ such that

$$c_{\mathcal{A}}\|v\|_{\mathcal{H}} \leq \|\mathcal{A}v\|_{\mathcal{H}'} \leq C_{\mathcal{A}}\|v\|_{\mathcal{H}}, \quad v \in \mathcal{H}. \tag{3.1}$$

The simplest example is

$$a(v,u) := \langle \nabla v, \nabla u\rangle + \kappa\langle v,u\rangle, \quad \kappa \geq 0, \quad \langle v,w\rangle = \int_\Omega vw, \tag{3.2}$$

and $\mathcal{H} = H^1_0(\Omega)$ (the space of functions with first order weak derivatives in L_2 whose traces vanish on the boundary $\Gamma = \partial\Omega$) endowed with the norm $\|v\|^2_{\mathcal{H}} := \|\nabla v\|^2_{L_2(\Omega)} + c\|v\|^2_{L_2(\Omega)}$. In this case, $C_{\mathcal{A}} = 1$ while $c_{\mathcal{A}}$ is the constant from the Poincaré inequality. Clearly, for any fixed $f \in \mathcal{H}'$, the problem: find $u \in \mathcal{H}$ such that

$$\langle v, R(u)\rangle := a(v,u) - \langle v, f\rangle = 0, \quad v \in \mathcal{H}, \tag{3.3}$$

is well posed in the sense that $DR = \mathcal{A}$ has the required mapping property relative to $\mathcal{H}$ defined above. Moreover, taking any wavelet basis whose properly scaled versions are Riesz bases in $L_2(\Omega)$ and $H^1_0(\Omega)$, one can find a diagonal scaling such that (2.5) holds *uniformly* in the parameter κ of the zero order term in (3.2), [7]. Denoting by $\mathbf{A} = (a(\psi_\lambda, \psi_\nu))_{\lambda,\nu\in\mathcal{J}}$ the wavelet representation of $\mathcal{A}$ and setting $\mathbf{f} = (\langle \psi_\lambda, f\rangle : \lambda \in \mathcal{J})^T$, the problem (2.9) takes the form

$$\mathbf{A}\mathbf{u} = \mathbf{f}, \tag{3.4}$$

whose solution $\mathbf{u}$ gives the solution $u = \sum_{\lambda\in\mathcal{J}} u_\lambda\psi_\lambda$ in $\mathcal{H}$. In this case $\mathbf{A}$ is symmetric positive definite so that one can find a relaxation parameter α such that

$$\|\mathbf{I} - \alpha\mathbf{A}\|_{\ell_2(\mathcal{J})\to\ell_2(\mathcal{J})} \leq \rho < 1. \tag{3.5}$$

Hence, the iteration

$$\mathbf{u}^{n+1} = \mathbf{u}^n - \alpha(\mathbf{A}\mathbf{u}^n - \mathbf{f}), \quad n = 0,1,2,\ldots, \tag{3.6}$$

has the form (2.10) with $\mathbf{B}_n := \alpha\mathbf{I}$ and provides iterates satisfying

$$\|\mathbf{u} - \mathbf{u}^n\|_{\ell_2(\mathcal{J})} \leq \rho\|\mathbf{u} - \mathbf{u}^{n-1}\|_{\ell_2(\mathcal{J})}. \tag{3.7}$$

This line of thought easily extends to problems where $a(\cdot,\cdot)$ is no longer definite or not even symmetric but where the induced operator $\mathcal{A}$ still satisfies the mapping property (3.1). This is for instance the case for certain *saddle point problems* such

as the Stokes problem, see [8] for more examples. In this case $\mathbf{A}^T\mathbf{A}$ is symmetric positive definite and, since $\mathbf{A}$ satisfies (2.8), is still boundedly invertible on $\ell_2(\mathcal{J})$. Hence one can again find a relaxation parameter $\alpha > 0$ such that

$$(3.8)\quad \mathbf{u}^{n+1} = \mathbf{u}^n - \alpha(\mathbf{A}^T\mathbf{A}\mathbf{u}^n - \mathbf{A}^T\mathbf{f}) = \mathbf{u}^n - \alpha\mathbf{A}^T(\mathbf{A}\mathbf{u}^n - \mathbf{f}), \quad n = 0, 1, 2, \ldots,$$

converges with a fixed error reduction $\rho < 1$ and has the form (2.10) with $\mathbf{B}_n := \alpha\mathbf{A}^T$. For saddle point problems there are actually alternatives that avoid squaring the problem (in wavelet coordinates). One option is to employ an Uzawa iteration for applying the Schur complement operator, which also leads to an iteration of the form (3.6), [12, 16].

As mentioned before, we could use stage dependent relaxation parameters α_n in (3.6) and (3.8) which would give rise to instationary "preconditioners" $\mathbf{B}_n$. However, since this will have no asymptotic effect we shall confine the discussion in connection with the case (L) always to the stationary matrices $\mathbf{B}_n = \mathbf{B} = \alpha\mathbf{I}$ or $\mathbf{B}_n = \mathbf{B} = \alpha\mathbf{A}^T$.

Of course, in either case step (n4) requires eventually approximating the *weighted residual* $\mathbf{B}_n\mathbf{R}(\mathbf{u}^n)$ which in the above linear case amounts to approximating $\mathbf{f}$ and approximately evaluating the infinite matrix $\mathbf{A}$ (respectively $\mathbf{A}^T$). We shall address this issue later in some detail.

3.2. Semilinear Elliptic Problems (SL). To formulate a natural nonlinear generalization of the above problem class (L) suppose again that $a(\cdot,\cdot)$ is a continuous bilinear form on a Hilbert space $\mathcal{H}$ endowed with the norm $\|\cdot\|_{\mathcal{H}}$, which is $\mathcal{H}$-elliptic, i.e., there exist positive constants c, C such that

$$(3.9)\qquad c\|v\|_{\mathcal{H}}^2 \le a(v,v), \quad a(v,w) \le C\|v\|_{\mathcal{H}}\|w\|_{\mathcal{H}}, \quad \forall\, v, w \in \mathcal{H},$$

see (3.2) as an example. In principle, the subsequent analysis will also cover elliptic integral operators with positive order such as the hypersingular operator.

We suppose that $G : \mathbb{R} \to \mathbb{R}$ is a function with the following property:

P1 the mapping $v \mapsto G(v)$ takes $\mathcal{H}$ into its dual $\mathcal{H}'$ and is *stable* in the sense that

$$(3.10)\quad \|G(u) - G(v)\|_{\mathcal{H}'} \le C(\max\{\|u\|_{\mathcal{H}}, \|v\|_{\mathcal{H}}\})\|u - v\|_{\mathcal{H}}, \quad u, v \in \mathcal{H},$$

where $t \to C(t)$ is a nondecreasing function of t.

The problem: Given $f \in \mathcal{H}'$ find $u \in \mathcal{H}$ such that

$$(3.11)\qquad \langle v, F(u)\rangle := a(v,u) + \langle v, G(u)\rangle = \langle v, f\rangle, \quad \forall\ v \in \mathcal{H},$$

is of the form (2.2) with $R(v) = F(v) - f$. We wish to discuss this problem in a little more detail because it will turn out to admit a complete analysis of adaptive solvers and thus may serve as a good point of departure for more difficult problems later.

It is not hard to verify that the weak formulation of the boundary value problem

$$(3.12)\qquad -\Delta u + u^3 = f \quad \text{in } \Omega, \quad u = 0 \ \text{ on } \partial\Omega,$$

is of the form (3.11) where for $\mathcal{H} = H_0^1(\Omega)$ the above assumptions hold for $d \le 3$. We shall describe later more general classes of mappings G satisfying property P1 and covering (3.12) as a special case.

REMARK 3.1. *If we assume in addition that G is monotone, i.e. $(u-v)(G(u) - G(v)) \ge 0$ for $u, v \in \mathbb{R}$, then (3.11) has for every $f \in \mathcal{H}'$ a unique solution $u \in \mathcal{H}$.*

The argument follows standard lines. Under the above assumptions it is easy to show that the operator F, defined by (3.11), is also monotone and coercive. One can then invoke the Browder-Minty Theorem (see e.g [21], Theorem 9.45) to conclude existence while the strict monotonicy guaranteed by the quadratic part also ensures uniqueness.

REMARK 3.2. *Alternatively one can argue that, under the above assumptions, G is of potential type so that (3.11) is the Euler equation of a convex minimization problem with a strictly convex functional, see e.g. [24], Proposition 42.6.*

An equivalent ℓ_2-formulation (n2):. We turn now to step (n2) in the present setting. In order to rewrite (3.11) in wavelet coordinates, we adhere to the notation from Section 3.1 and set in addition $\mathbf{G}(\mathbf{v}) := (\langle \psi_\lambda, G(v)\rangle)_{\lambda\in\mathcal{J}}$. Then $u = \sum_{\lambda\in\mathcal{J}} u_\lambda\psi_\lambda$ is the unique solution of (3.11) if and only if $\mathbf{u}$ solves

$$\mathbf{R}(\mathbf{u}) := \mathbf{A}\mathbf{u} + \mathbf{G}(\mathbf{u}) - \mathbf{f} = \mathbf{0}. \tag{3.13}$$

Note that, in view of (2.6), f belongs to $\mathcal{H}'$ if and only if $\mathbf{f} \in \ell_2(\mathcal{J})$. Clearly $\mathbf{A}$ is symmetric positive definite. Moreover, it follows from (3.9) that the operator $\mathcal{A}$ defined by $a(v,u) = \langle v, \mathcal{A}u\rangle$ for all $v \in \mathcal{H}$, is an isomorphism from $\mathcal{H}$ onto $\mathcal{H}'$. As an immediate consequence of Remark 2.1 we can assert that there exist positive constants c_A, C_A (depending only on c_1, C_1 in (2.5) and on the constants in (3.9)) such that

$$c_A\|\mathbf{v}\|_{\ell_2(\mathcal{J})} \le \|\mathbf{A}\mathbf{v}\|_{\ell_2(\mathcal{J})} \le C_A\|\mathbf{v}\|_{\ell_2(\mathcal{J})}, \quad \mathbf{v} \in \ell_2(\mathcal{J}), \tag{3.14}$$

which, of course, means that $\|\mathbf{A}\|_{\ell_2(\mathcal{J})\to\ell_2(\mathcal{J})} \le C_A$ and $\|\mathbf{A}^{-1}\|_{\ell_2(\mathcal{J})\to\ell_2(\mathcal{J})} < c_A^{-1}$, i.e., $\mathbf{A}$ is a positive definite automorphism of $\ell_2(\mathcal{J})$.

We end this Section with the simple observation that monotonicity of G together with (3.9) implies monotonicity of F which carries over into the discrete setting, namely

$$(\mathbf{u}-\mathbf{v})^T(\mathbf{F}(\mathbf{u}) - \mathbf{F}(\mathbf{v})) \ge 0, \quad \mathbf{u},\mathbf{v} \in \ell_2(\mathcal{J}). \tag{3.15}$$

In fact, denoting by $\tilde{\Psi}$ the dual basis to Ψ, we have by definition of $\mathbf{F}(\mathbf{u})$

$$u - v = \sum_{\lambda\in\mathcal{J}} (u_\lambda - v_\lambda)\psi_\lambda, \quad F(u) - F(v) = \sum_{\lambda\in\mathcal{J}} (\langle F(u) - F(v), \psi_\lambda\rangle\tilde{\psi}_\lambda.$$

Thus (3.15) follows from monotonicity of F and biorthogonality.

Gradient Iterations (n3):. We now address (n3) for the above class of semilinear elliptic problems. The simplest option is to take $\mathbf{B}_n = \alpha\mathbf{I}$ which gives the iteration

$$\mathbf{u}^{n+1} = \mathbf{u}^n - \alpha\mathbf{R}(\mathbf{u}^n), \quad n \in \mathbb{N}_0. \tag{3.16}$$

Keeping Remark 3.2 in mind, one can show that (3.13) are the Euler equations of a strictly convex minimization problem on $\ell_2(\mathcal{J})$, and (3.16) can be viewed as a gradient iteration. However, if we do not insist on determining the optimal step size $\alpha = \alpha_n$ for the steepest descent we can argue directly as follows.

$$\begin{aligned}
\mathbf{u}^{n+1} - \mathbf{u} &= \mathbf{u}^n - \mathbf{u} - \alpha(\mathbf{R}(\mathbf{u}^n) - \mathbf{R}(\mathbf{u})) \\
&= \left(\mathbf{I} - \alpha\int_0^1 (\mathbf{A} + D\mathbf{G}(\mathbf{u} + s(\mathbf{u}^n - \mathbf{u})))ds\right)(\mathbf{u}^n - \mathbf{u}) \\
&=: (\mathbf{I} - \alpha\mathbf{M}(\mathbf{u}^n, \mathbf{u}))\,(\mathbf{u}^n - \mathbf{u}). \qquad (3.17)
\end{aligned}$$

By (3.14) and (3.15), the smallest eigenvalue of the matrix $\mathbf{M}(\mathbf{u}^n, \mathbf{u})$ is bounded from below by c_A. By P1 and (3.14), $\mathbf{F}$ takes bounded sets into bounded sets. Thus the largest eigenvalue of $\mathbf{M}(\mathbf{u}^n, \mathbf{u})$ is bounded from above by some constant $\tilde{C}$ depending on C_A and the norms $\|\mathbf{u}^n\|_{\ell_2(\mathcal{J})}, \|\mathbf{u}\|_{\ell_2(\mathcal{J})}$. In particular, given $\mathbf{u}^0$, $\delta_0 := \|\mathbf{u}-\mathbf{u}^0\|_{\ell_2(\mathcal{J})}$, there exists a positive α such that $\|\mathbf{I} - \alpha\mathbf{M}(\mathbf{v}, \mathbf{u})\|_{\ell_2(\mathcal{J})\to\ell_2(\mathcal{J})} \le \rho < 1$ for all $\mathbf{v} \in B_{\delta_0}(\mathbf{u})$, the ball of radius δ_0 with center $\mathbf{u}$. This confirms the following observation.

REMARK 3.3. *Given $\mathbf{u}^0$ choose $\alpha > 0$ such that*

$$\|\mathbf{I} - \alpha\mathbf{M}(\mathbf{u}^0, \mathbf{u})\|_{\ell_2(\mathcal{J})\to\ell_2(\mathcal{J})} \le \rho < 1, \tag{3.18}$$

then, one has

$$\|\mathbf{u} - \mathbf{u}^n\|_{\ell_2(\mathcal{J})} \le \rho\|\mathbf{u} - \mathbf{u}^{n-1}\|_{\ell_2(\mathcal{J})}, \quad n \in \mathbb{N}. \tag{3.19}$$

3.3. The General Nonlinear Case – Locally Convergent Schemes (GNL). While the assumptions in the previous setting allow us to conclude convergence of the ideal infinite dimensional scheme for *any* initial guess $\mathbf{u}^0$, one often has to be content with weaker assumptions (and correspondingly weaker conclusions). In the literature, variational problems of the type (2.2) are frequently studied under general assumptions on R, such as A1 and A2, that typically guarantee local convergence of an iterative scheme to a locally unique solution provided that a sufficiently good initial guess is known, see e.g. [20, 22].

Our plan here is to exemplify the above paradigm under assumptions A1 and A2, provided that a sufficiently good initial approximation is known. According to (n2) we consider again the equivalent formulation (2.9) in wavelet coordinates and turn to devising a suitable iteration of the form (2.10) that converges for a sufficiently good initial guess. To this end, we assume that

$$\mathbf{u}^0 \in B_\delta(\mathbf{u}) := \{\mathbf{v} : \|\mathbf{v}-\mathbf{u}\|_{\ell_2(\mathcal{J})} < \delta\}, \tag{3.20}$$

where δ will be specified below.

As mentioned before a possible choice for $\mathbf{B}_n$ could involve the Jacobian which leads to Newton's method. But under the above weak assumptions on R we wish to avoid at this point requiring higher order smoothness conditions and consider first the following much simpler option. An analog to the least squares iteration (3.8) would be $\mathbf{B}_n := D\mathbf{R}(\mathbf{u}^n)^T$. An even simpler alternative, which is presumably less computationally demanding, is to take the *stationary* matrix

$$\mathbf{B} = D\mathbf{R}(\mathbf{u}^0)^T, \tag{3.21}$$

provided that δ is sufficiently small. Let us point out next that

$$\mathbf{W}(\mathbf{v}) := \mathbf{v} - \alpha D\mathbf{R}(\mathbf{u}^0)^T\mathbf{R}(\mathbf{v})$$

is a contraction on $B_\delta(\mathbf{u})$. In fact,

$$\begin{aligned}\mathbf{W}(\mathbf{z}) - \mathbf{W}(\mathbf{v}) &= (\mathbf{z}-\mathbf{v}) - \alpha D\mathbf{R}(\mathbf{u}^0)^T(\mathbf{R}(\mathbf{z})-\mathbf{R}(\mathbf{v}))\\ &= \left(\mathbf{I} - \alpha D\mathbf{R}(\mathbf{u}^0)^T D\mathbf{R}(\mathbf{v})\right)(\mathbf{z}-\mathbf{v}) + \mathrm{o}(\|\mathbf{z}-\mathbf{v}\|_{\ell_2(\mathcal{J})})\\ &= \left(\mathbf{I} - \alpha D\mathbf{R}(\mathbf{u}^0)^T D\mathbf{R}(\mathbf{u}^0)\right)(\mathbf{z}-\mathbf{v}) + \mathrm{o}(\|\mathbf{z}-\mathbf{v}\|_{\ell_2(\mathcal{J})})\\ &\quad + \mathrm{O}(\epsilon(\delta)\|\mathbf{z}-\mathbf{v}\|_{\ell_2(\mathcal{J})}),\end{aligned} \tag{3.22}$$

where we have used assumption A1 and where $\epsilon(\delta)$ tends to zero as $\delta \to 0$. By A1 and A2, $DR(u^0)$ is still an isomorphism from $\mathcal{H}$ onto $\mathcal{H}'$ when δ is sufficiently small. Thus, by Remark 2.1, the positive definite matrix $D\mathbf{R}(\mathbf{u}^0)^T D\mathbf{R}(\mathbf{u}^0)$ is an automorphism on $\ell_2(\mathcal{J})$. Therefore, for $\alpha > 0$ satisfying

$$\alpha\|D\mathbf{R}(\mathbf{u}^0)^T D\mathbf{R}(\mathbf{u}^0)\|_{\ell_2(\mathcal{J})\to\ell_2(\mathcal{J})} < 2, \tag{3.23}$$

$\mathbf{W}$ is a contraction on $B_\delta(\mathbf{u})$. Furthermore, the iterates

$$\mathbf{u}^{n+1} = \mathbf{u}^n - \alpha D\mathbf{R}(\mathbf{u}^0)^T\mathbf{R}(\mathbf{u}^n), \quad n = 0, 1, \ldots, \tag{3.24}$$

stay in $B_\delta(\mathbf{u})$. In fact, as above

$$\begin{aligned}\mathbf{u}^{n+1} - \mathbf{u} &= \mathbf{u}^n - \mathbf{u} - \alpha D\mathbf{R}(\mathbf{u}^0)^T(\mathbf{R}(\mathbf{u}^n) - \mathbf{R}(\mathbf{u}))\\ &= \mathbf{u}^n - \mathbf{u} - \alpha D\mathbf{R}(\mathbf{u}^0)^T(\mathbf{R}(\mathbf{u}^n) - \mathbf{R}(\mathbf{u}))\\ &= (\mathbf{I} - \alpha D\mathbf{R}(\mathbf{u}^0)^T D\mathbf{R}(\mathbf{u}^0))(\mathbf{u}^n - \mathbf{u}) + \mathrm{o}(\|\mathbf{u}^n - \mathbf{u}\|_{\ell_2(\mathcal{J})})\\ &\quad + \mathrm{O}(\epsilon(\delta)\|\mathbf{u}^n - \mathbf{u}\|_{\ell_2(\mathcal{J})}).\end{aligned}$$

Hence, for α as above and δ sufficiently small, i.e. $\|\mathbf{I} - \alpha D\mathbf{R}(\mathbf{u}^0)^T D\mathbf{R}(\mathbf{u}^0)\|_{\ell_2(\mathcal{J})\to\ell_2(\mathcal{J})} =: b < 1$ and $\mathrm{o}(1) + \mathrm{O}(\epsilon(\delta)) < 1 - b$, one has $\|\mathbf{u}^{n+1} - \mathbf{u}\|_{\ell_2(\mathcal{J})} < \delta$. We can summarize these observations as follows.

REMARK 3.4. *Under the above assumptions there exists a $\delta_0 > 0$ and a positive α such that for any $\delta \le \delta_0$ and $\mathbf{u}^0 \in B_\delta(\mathbf{u})$ the iteration (3.24) converges to the locally unique solution $\mathbf{u}$ of (2.9). Moreover, there exists some $\rho < 1$ such that*

$$\|\mathbf{u}^n - \mathbf{u}\|_{\ell_2(\mathcal{J})} \le \rho \|\mathbf{u}^{n-1} - \mathbf{u}\|_{\ell_2(\mathcal{J})}, \quad n = 0, 1, \ldots. \tag{3.25}$$

4. A Perturbed First Order Iteration Scheme. We shall now turn to step (n4) under the assumption that (2.10) gives rise to a fixed error reduction ρ per iteration step. Recall that by (3.7), (3.19) and (3.25), that this is indeed the case for (L), (SL) and (GNL) already for the corresponding stationary choices of $\mathbf{B}_n = \mathbf{B}$. In order to minimize technicalities we shall consider only this case in connection with such first order schemes. In order to arrive at computable versions of these schemes we have to *approximate* the weighted residuals $\mathbf{BR}(\mathbf{u}^n)$ in each step. Already in the linear case (L) this requires approximating the application of an infinite matrix to a finitely supported vector and to approximate the given data $\mathbf{f}$. In the nonlinear cases (SL), (GNL) the additional difficulty is to approximately evaluate the *nonlinear* expressions $\mathbf{R}(\mathbf{u}^j)$.

Our strategy can be outlined as follows. In the present section we shall address only issue (a) from Section 1.2, namely how accurate have these approximations to be chosen at a given stage of the iteration, so as to guarantee convergence to the correct solution? We shall do so at this point under the *assumption* that a subroutine for approximating the weighted residuals $\mathbf{BR}(\mathbf{v})$ with desired accuracy is at our disposal. Once (a) has been clarified for the general scope of problems, we shall in subsequent sections then narrow down step by step the specific requirements on the basic subroutine, develop concrete realizations for the various problem types (L), (SL) and (GNL) and analyze their complexity.

Thus for the time being we assume now that a routine with the following property is given:

RES $[\eta, \mathbf{B}, \mathbf{R}, \mathbf{v}] \to \mathbf{w}_\eta$: *determines for any positive tolerance η and any finitely supported input $\mathbf{v}$ a finitely supported $\mathbf{w}_\eta$ satisfying*

$$\|\mathbf{BR}(\mathbf{v}) - \mathbf{w}_\eta\|_{\ell_2(\mathcal{J})} \le \eta. \tag{4.1}$$

The need for the following further ingredient is at this point less obvious. It will be applied after a certain finite number of perturbed iterations based on the application of **RES**. It will be seen later that this is crucial for controling the complexity of the scheme.

CCOARSE $[\eta, \mathbf{v}] \to \mathbf{w}_\eta$: *determines for any positive tolerance η and any finitely supported input vector $\mathbf{v}$ a finitely supported output vector $\mathbf{w}_\eta$ such that*

$$\|\mathbf{v} - \mathbf{w}_\eta\|_{\ell_2(\mathcal{J})} \le \eta, \tag{4.2}$$

while the support of $\mathbf{w}_\eta$ is minimized subject to certain constraints on the distribution of its entries.

The constraints mentioned in **CCOARSE** will depend on the particular application and will be specified later. A perturbed iteration based on these ingredients requires specifying a suitable

Initialization: We distinguish the following three cases for the choice of the initial guess:
(L): In the linear case (3.3) we can set $\mathbf{u}^0 := \mathbf{0}$ so that an initial error bound is

given by $\epsilon_0 := c_A^{-1}\|\mathbf{f}\|_{\ell_2(\mathcal{J})}$ when $\mathbf{A}$ is symmetric positive definite, $\mathbf{B} = \alpha\mathbf{I}$, and $\epsilon_0 := c_A^{-1}C_A\|\mathbf{f}\|_{\ell_2(\mathcal{J})}$ in the general case when using a least squares formulation $\mathbf{B} = \alpha\mathbf{A}^T$.

(SL): In the case of semilinear elliptic problems (3.13), (3.11) we know that the exact solution $\mathbf{u}$ of (3.13) is contained in the ball B_0 around zero. Since $\mathbf{u} = -(D\mathbf{R}(\xi\mathbf{u}))^{-1}\mathbf{R}(\mathbf{0})$ for some $\xi \in [0,1]$, we conclude that

$$\|\mathbf{u}\|_{\ell_2(\mathcal{J})} \leq \max_{s\in[0,1]} \|(D\mathbf{R}(s\mathbf{u}))^{-1}\|_{\ell_2(\mathcal{J})\to\ell_2(\mathcal{J})}\|\mathbf{R}(\mathbf{0})\|_{\ell_2(\mathcal{J})} =: \epsilon_0, \tag{4.3}$$

which is the desired initial error bound for the initial guess $\mathbf{u}^0 = \mathbf{0}$. Now let

$$B := \{\mathbf{v} \in \ell_2(\mathcal{J}) : \|\mathbf{u} - \mathbf{v}\|_{\ell_2(\mathcal{J})} \leq 2\epsilon_0 \tag{4.4}$$

denote the ball around $\mathbf{u}$ with radius $2\epsilon_0$. Then, choose $\alpha > 0$ according to the remarks in the previous section such that

$$\|\mathbf{I} - \alpha D\mathbf{R}(\mathbf{v})\|_{\ell_2(\mathcal{J})\to\ell_2(\mathcal{J})} \leq \rho < 1, \quad \forall \ \mathbf{v} \in B. \tag{4.5}$$

We fix this value of α and set $\mathbf{B} = \alpha\mathbf{I}$ in the case (3.13).

(GNL): For the locally convergent scheme, we adhere to the assumptions made in Section 3.3. For any fixed $\delta < \delta_0$ (the parameter from Remark 3.4) which satisfies $(1+\alpha)\delta < \delta_0$ where α is the constant from (3.23), we choose $\mathbf{u}^0$ according to (3.20). In this case we have $\mathbf{B} = \alpha D\mathbf{R}(\mathbf{u}^0)^T$ and and $\epsilon_0 := \delta$ is a valid initial error bound.

Thus in all cases (L), (SL) and (GNL) one has under the above premises

$$\|\mathbf{u} - \mathbf{u}^0\|_{\ell_2(\mathcal{J})} \leq \epsilon_0. \tag{4.6}$$

The last prerequisite is to determine the number of applications of **RES** before calling **CCOARSE**. To this end, it will be convenient to extract the following fact from the above considerations.

REMARK 4.1. *For each of the above choices of* $\mathbf{B}$ *in* (L), (SL) *and* (GNL) *there exists a neighborhood* B *of the exact solution* $\mathbf{u}$ *and a constant* $\rho < 1$ *such that for* $\mathbf{v}, \mathbf{z} \in B$ *one has*

$$\|(\mathbf{v} - \mathbf{z}) - \mathbf{B}(\mathbf{R}(\mathbf{v}) - \mathbf{R}(\mathbf{z}))\|_{\ell_2(\mathcal{J})} \leq \rho\|\mathbf{v} - \mathbf{z}\|_{\ell_2(\mathcal{J})}. \tag{4.7}$$

Proof: In the linear case (L) this readily follows from (3.5) when $\mathbf{A}$ is symmetric positive definite and $\mathbf{B} = \alpha\mathbf{I}$ for a suitable α. More generally, when (3.1) holds, α is chosen so that $\mathbf{I} - \alpha\mathbf{A}^T\mathbf{A}$ is a contraction which corresponds to $\mathbf{B} = \alpha\mathbf{A}^T$. In either case we can take $B = \ell_2(\mathcal{J})$.

In the case (SL) of the semilinear elliptic problem (3.11), respectively (3.13), this follows for $\mathbf{B} = \alpha\mathbf{I}$ and suitable positive α from the reasoning leading to (3.18), where B is the set from the initialization.

Finally, for (GNL) (see Section 3.3) the claim follows from (3.22) for $\mathbf{B} = \alpha D\mathbf{R}(\mathbf{u}^0)^T$ and α satisfying (3.23) with $B = B_\delta(\mathbf{u})$. □

We can now describe our computable analogue of (2.10). For this we choose any fixed summable sequence $(\omega_j)_{j\in\mathbb{N}_0}$ which, for convenience, we arrange to sum to one $\sum_{j=0}^{\infty}\omega_j = 1$. Then define K as

$$K := \min\{k \in \mathbb{N} : (\rho+1)\rho^{k-1} \leq \frac{1}{2(1+3C^*)}\}, \tag{4.8}$$

where ρ is the constant from (4.7) and where C^* is a fixed constant depending on the realization of the routine **CCOARSE**, see Section 6.2. We have now collected all necessary ingredients to describe the following scheme:

SOLVE $[\epsilon, \mathbf{R}] \to \bar{\mathbf{u}}(\epsilon)$

(i) Set $\bar{\mathbf{u}}^0 = \mathbf{u}^0$ and the corresponding initial bound ϵ_0 according to the above initialization, and define $j = 0$;

(ii) If $\epsilon_j \leq \epsilon$ stop and output $\bar{\mathbf{u}}(\epsilon) := \bar{\mathbf{u}}^j$; else set $\mathbf{v}^0 := \bar{\mathbf{u}}^j$ and for $k = 0, \ldots, K-1$, set $\eta_k := \omega_k \rho^k \epsilon_j$; compute

$$\mathbf{w}^k = \mathbf{RES}\,[\eta_k, \mathbf{B}, \mathbf{R}, \mathbf{v}^k]$$

and update

$$\mathbf{v}^{k+1} = \mathbf{v}^k - \mathbf{w}^k;$$

(iii) $\mathbf{CCOARSE}\,[\frac{3C^*\epsilon_j}{2(1+3C^*)}, \mathbf{v}^K] \to \bar{\mathbf{u}}^{j+1}$, $\epsilon_{j+1} = \epsilon_j/2$, $j+1 \to j$, go to (ii).

Let us confirm first that the choice of accuracy tolerances in **SOLVE** imply convergence.

PROPOSITION 4.2. *The iterates $\bar{\mathbf{u}}^j$ produced by the scheme* **SOLVE** *satisfy*

$$\|\mathbf{u} - \bar{\mathbf{u}}^j\|_{\ell_2(\mathcal{J})} \leq \epsilon_j, \tag{4.9}$$

so that in particular $\|\mathbf{u} - \bar{\mathbf{u}}(\epsilon)\|_{\ell_2(\mathcal{J})} \leq \epsilon$. Moreover, one has

$$\|u - \sum_{\lambda \in \Lambda(\epsilon)} \bar{u}(\epsilon)_\lambda \psi_\lambda\|_{\mathcal{H}} \leq C_1 \epsilon, \tag{4.10}$$

where C_1 is the constant from (2.5) and $\Lambda(\epsilon) := \operatorname{supp} \mathbf{u}(\epsilon)$.

Proof: We assume the above initialization and employ a simple perturbation argument using induction on j. We fix a value of j and let $\mathbf{u}^k := \mathbf{u}^k(\mathbf{v}^0)$ be the exact iterates $\mathbf{u}^{k+1} = \mathbf{u}^k - \mathbf{BR}(\mathbf{u}^k)$ with initial guess $\mathbf{u}^0 = \mathbf{v}^0 = \bar{\mathbf{u}}^j$. Hence

$$\begin{aligned} \mathbf{v}^{k+1} - \mathbf{u}^{k+1} &= \mathbf{v}^k - \mathbf{u}^k - (\mathbf{w}^k - \mathbf{BR}(\mathbf{u}^k)) \\ &= \mathbf{v}^k - \mathbf{u}^k - \mathbf{B}(\mathbf{R}(\mathbf{v}^k) - \mathbf{R}(\mathbf{u}^k)) + (\mathbf{BR}(\mathbf{v}^k) - \mathbf{w}^k). \end{aligned} \tag{4.11}$$

Next we wish to invoke (4.7). To do this we need to make sure that the iterates $\mathbf{v}^k, \mathbf{u}^k$ stay in the neighborhood B mentioned in Remark 4.1. In the linear case (L) there is no constraint, i.e., $B = \ell_2(\mathcal{J})$. Let us look at the semilinear case (SL) next. By induction assumption we know that $\|\mathbf{u} - \bar{\mathbf{u}}^j\|_{\ell_2(\mathcal{J})} \leq \epsilon_j \leq \epsilon_0$. Therefore $\|\mathbf{u} - \mathbf{u}^k\|_{\ell_2(\mathcal{J})} \leq \rho^k \|\mathbf{u} - \mathbf{u}^0\|_{\ell_2(\mathcal{J})} \leq \rho^k \|\mathbf{u} - \bar{\mathbf{u}}^j\|_{\ell_2(\mathcal{J})} \leq \rho^k \epsilon_j$. So $\mathbf{u}^k \in B$ for all $k \leq K$. Also $\mathbf{v}^0 = \bar{\mathbf{u}}^j \in B$. Thus suppose that $\mathbf{v}^k$ is in B. We wish to show that then also $\mathbf{v}^{k+1} \in B$. To this end, we infer from (4.7), (4.11) and the definition of $\mathbf{w}^k$ in step (ii) that

$$\begin{aligned} \|\mathbf{v}^{k+1} - \mathbf{u}^{k+1}\|_{\ell_2(\mathcal{J})} &\leq \rho \|\mathbf{v}^k - \mathbf{u}^k\|_{\ell_2(\mathcal{J})} + \omega_k \rho^k \epsilon_j \\ &\leq \rho^k \epsilon_j \sum_{l=0}^{k} \omega_{k-l} \leq \rho^k \epsilon_j, \end{aligned} \tag{4.12}$$

where we have used that $\mathbf{u}^0 = \mathbf{v}^0$. Since $\rho < 1$ we see, by the previous estimate for the $\mathbf{u}^k$, that also $\mathbf{v}^{k+1} \in B$ and the iteration can be advanced up to K. We now conclude from (4.8) and (4.12) that

$$\|\mathbf{u} - \mathbf{v}^K\|_{\ell_2(\mathcal{J})} \leq (\rho + \alpha)\rho^{K-1}\epsilon_j \leq \frac{\epsilon_j}{2(1+3C^*)}. \tag{4.13}$$

For the locally convergent scheme (GNL) with $\mathbf{B} = \alpha \mathbf{R}(\mathbf{u}^0)^T$ the reasoning is analogous. The choice of the initial guess ensures that $(\rho + \alpha)\epsilon_j \leq (\rho + \alpha)\epsilon \leq (1 + \alpha)\epsilon_0 \leq \delta_0$. Then the above arguments for (SL) yield again (4.13) so that all iterates stay in $B_{\delta_0}(\mathbf{u})$.

Thus in all cases the estimate (4.9) follows now immediately from step (iii) in **SOLVE**, the definition of **CCOARSE** and (4.2). Finally, (4.10) is an immediate consequence of the norm equivalence (2.5) and (4.9). □

Thus for an idealized infinite dimensional scheme of order one in (n3) we know how to choose the tolerances in the routines **RES** and **CCOARSE** so as to guarantee convergence. Let us pause to point out that the scheme **SOLVE** in the above form should be viewed as the simplest proptotype whose practical realization could be varied in several ways depending on the special case at hand. For instance, in step (ii) one could try to avoid performing always K perturbed iterations but terminate earlier based on information provided by the approximate residuals. This might be important if only a poor guess of the reduction factor ρ is available which determines the possibly much too pessimistic value of K. Let us briefly indicate a possible strategy. Note that in all the above examples there exists a constant $\hat{C}$ such that

$$\|\mathbf{u} - \mathbf{v}\|_{\ell_2(\mathcal{J})} \leq \hat{C} \|\mathbf{BR}(\mathbf{v})\|_{\ell_2(\mathcal{J})} \tag{4.14}$$

for $\mathbf{v}$ in a suitable neighborhood of $\mathbf{u}$. In fact, in the semilinear case (SL), the monotonicity of G assures that for $\mathbf{B} = \alpha \mathbf{I}$ we could take $\hat{C} = (\alpha \lambda_{\min}(\mathbf{A}))^{-1}$ where $\lambda_{\min}(\mathbf{A})$ is a lower bound for the spectrum of $\mathbf{A}$. Alternatively, one has in this case

$$\|R(v)\|_{\mathcal{H}'} = \sup_{w \in \mathcal{H}} \frac{a(w, v-u) + \langle w, G(v) - G(u) \rangle}{\|w\|_{\mathcal{H}}} \geq c \|v - u\|_{\mathcal{H}},$$

where c is the ellipticity constant from (3.9). Now the norm equivalences (2.5) and (2.6) yield

$$\|\mathbf{R}(\mathbf{v})\|_{\ell_2(\mathcal{J})} \geq c_1 \|R(v)\|_{\mathcal{H}'} \geq c c_1 \|u - v\|_{\mathcal{H}} \geq c c_1^2 \|\mathbf{u} - \mathbf{v}\|_{\ell_2(\mathcal{J})},$$

so that $\hat{C} = c c_1^2$ would be an admissible choice (without any restriction on $\mathbf{v}$). For the locally convergent scheme with $\mathbf{B} = \alpha D \mathbf{R}(\mathbf{u}^0)^T$ the reasoning leading to (3.23) also leads to an estimate of the type (4.14).

To see how this can be exploited, an inspection of the above proof reveals the following fact. Choose any $\bar{\rho} < 1$ and define $\bar{K}$ by (4.8) with respect to $\bar{\rho}$. Replacing ρ by $\bar{\rho}$ in the definition of the tolerances in step (ii), it would take $M := \max\{K, \bar{K}\}$ steps to ensure that in the $(j+1)$st call of (ii) $\|\mathbf{u} - \mathbf{v}^M\|_{\ell_2(\mathcal{J})} \leq \epsilon_j / (2(1 + 3C^*))$. Now suppose that the ρ is expected to be a too pessimistic estimate of the true reduction rate. Choosing e.g. $\bar{\rho} := 1/2$ and setting $\eta_k := 2^{-k} \epsilon_j$ as tolerances in the $(j+1)$st call of (ii), we infer from (4.14) that

$$\|\mathbf{u} - \mathbf{v}^k\|_{\ell_2(\mathcal{J})} \leq \hat{C} \|\mathbf{BR}(\mathbf{v}^k)\|_{\ell_2(\mathcal{J})} \leq \hat{C} \eta_k + \|\mathbf{w}^k\|_{\ell_2(\mathcal{J})} =: \delta_k.$$

By the previous remarks, we can terminate the iteration in step (ii) of **SOLVE** when either $k = K$ or the computable a-posteriori bound $\delta_k \leq \epsilon_j / (2(1 + 3C^*))$, which might happen much earlier than predicetd by (4.8). Of course, the constant $\hat{C}$ is usually also only estimated. However, a poor estimate enters the termination criterion in a less severe way than a poor estimate for ρ. Nevertheless, in order to keep the exposition as simple as possible we confine the subsequent discussion to the above version of **SOLVE**, bearing in mind that variants of the above sort are automatically covered by the complexity analysis.

We have thus far introduced our paradigm for solving nonlinear problems. This paradigm is built on the availability of numerical algorithms such as **CCOARSE** and

RES. The remainder of this paper is to show how to construct concrete practical realizations of these algorithms in various settings and then to show how, under suitable controls on the computations in these algorithms, we can give complexity estimates for the entire numerical scheme **SOLVE**. More precisely, we wish to determine its *work/accuracy balance* i.e. given any target accuracy ϵ, how many degrees of freedom $N = N(\epsilon) := \#\Lambda(\epsilon)$ where $\Lambda(\epsilon) := \operatorname{supp} \bar{\mathbf{u}}(\epsilon)$, are needed to achieve it, and what is the associated (asymptotic) computational work. Of course, one hopes to keep the latter quantity proportional to $N(\epsilon)$ so that the number of degrees of freedom is a reasonable complexity measure. In the following section we shall address these issues first for the linear case (3.3). We review quickly the the relevant facts from [7, 8] tailored somewhat to the present situation, since this will guide the subsequent developments.

5. Realization and Complexity Analysis in the Linear Case (L). Recall from (3.6) that in the linear case, $\mathbf{BR}(\mathbf{v}) = \alpha(\mathbf{A}\mathbf{v} - \mathbf{f})$ (or $\alpha\mathbf{A}^T(\mathbf{A}\mathbf{v} - \mathbf{f})$). Thus, one part in approximating the residual is to approximate given data, here in the form of the right hand side $\mathbf{f}$ which, in general, is an infinite sequence.

5.1. Coarsening and Best N-term Approximation. We will also assume in the sequel that all coefficients of $\mathbf{f}$ are *known* and thus in principle accessible. In practice this may require a preprocessing step that computes for some overall target accuracy $\bar{\epsilon}$ (depending on the desired solution accuracy) an approximation $\mathbf{f}_{\bar{\epsilon}}$ satisfying $\|\mathbf{f} - \mathbf{f}_{\bar{\epsilon}}\|_{\ell_2(\mathcal{J})}$ and then order the entries by size. Once this has been done any coarser approximations, needed in the course of the iteration process, can be produced by the following simplest version of **CCOARSE**, introduced and analyzed in [7].

COARSE $[\eta, \mathbf{v}] \to \mathbf{v}_\eta$: *associates with any finitely supported input* $\mathbf{v}$ *a vector* $\mathbf{v}_\eta$ *such that*

$$\|\mathbf{v} - \mathbf{v}_\eta\|_{\ell_2(\mathcal{J})} \le \eta, \quad \#\operatorname{supp}\mathbf{w} \ge \#\operatorname{supp}\mathbf{v}_\eta, \text{ whenever } \|\mathbf{v} - \mathbf{w}\|_{\ell_2(\mathcal{J})} \le \eta. \tag{5.1}$$

Thus **COARSE** determines for a given finitely supported vector, a new vector with the smallest possible support deviating no more than a prescribed tolerance from the input. There is no constraint on the distribution of active indices in this case. Ordering the entries of $\mathbf{v}$ sizewise, this can be realized by summing entries in increasing order until the sum of their squares reaches η^2. For a detailed description of this routine see [7]. In fact, a strict ordering is not necessary. The same effect is realized by collecting the entries in binary bins which avoids a log factor at the expense of a fixed factor in the accuracy tolerance, [1].

The routine **COARSE** can be used to approximate the data $\mathbf{f}$ as follows.

$$\mathbf{RHS}\,[\eta, \mathbf{f}] := \mathbf{COARSE}\,[\eta - \bar{\epsilon}, \bar{\mathbf{f}}], \tag{5.2}$$

whenever $\eta > \bar{\epsilon}$.

Note that **COARSE** is a *nonlinear* process that realizes a given accuracy tolerance at the expense of a minimal number of degrees of freedom. It is therefore a version of *best N-term approximation* in $\ell_2(\mathcal{J})$. In fact, defining

$$\sigma_{N,\ell_2(\mathcal{J})}(\mathbf{u}) := \min_{\#\operatorname{supp}\mathbf{v} \le N} \|\mathbf{u} - \mathbf{v}\|_{\ell_2(\mathcal{J})}, \tag{5.3}$$

one has for any $\mathbf{v} \in \ell_2(\mathcal{J})$

$$\sigma_{N,\ell_2(\mathcal{J})}(\mathbf{v}) = \|\mathbf{v} - \mathbf{v}_N\|_{\ell_2(\mathcal{J})} = \left(\sum_{n>N} |v_n^*|^2\right)^{1/2}, \tag{5.4}$$

where $(v_n^*)_{n\in\mathbb{N}}$ is any *nonincreasing rearrangement* of $\mathbf{v}$. Thus $\mathbf{v}_N$ is obtained by retaining the N largest (in modulus) terms of $\mathbf{v}$ and setting all other entries to zero. Depending on the context $\mathbf{v}_N$ will be viewed as a sequence in $\ell_2(\mathcal{J})$ or a vector in $\mathbf{R}^N$.

The best N-term approximation sets a lower bound for the complexity that could ever be achieved by a scheme like **SOLVE**. In fact, it will serve as our bench mark in the case of linear variational problems of the form (3.3).

5.2. Adaptive Application of Compressible Matrices. It remains to approximate the action of $\mathbf{A}$ on a finitely supported vector $\mathbf{v}$. While the treatment of the right hand side data has been already seen to comply with best N-term approximation complexity, the question arises whether $\mathbf{Av}$ can be approximated with a similar efficiency. This has been answered affirmatively in [7] and we briefly recall the relevant facts from there.

Due to the *vanishing moment* property of wavelets the wavelet representation of many operators turns out to be *quasi-sparse*. The following quantification of sparsity is appropriate [7].

A matrix $\mathbf{C}$ is said to be s^*-compressible – $\mathbf{C}\in\mathcal{C}_{s^*}$ – if for any $0<s<s^*$ and every $j\in\mathbb{N}$ there exists a summable sequences $(\alpha_j)_{j=1}^\infty$ ($\sum_j\alpha_j<\infty$) and a matrix $\mathbf{C}_j$ obtained by replacing all but the order of $\alpha_j 2^j$ entries per row and column in $\mathbf{C}$ by zero, while still satisfying

$$\|\mathbf{C}-\mathbf{C}_j\|\le C\alpha_j 2^{-js},\quad j\in\mathbb{N}. \tag{5.5}$$

Specifically, wavelet representations of differential (and also certain singular integral) operators fall into this category. One typically has then estimates of the type

$$|a(\psi_\lambda,\psi_\mu)|\ \lesssim\ 2^{-\sigma||\lambda|-|\mu||}, \tag{5.6}$$

where $\sigma>d/2$ depends on the *regularity* of the wavelets.

In order to describe the essence of an approximate application scheme for compressible matrices, we abbreviate for any finitely supported $\mathbf{v}$ the best 2^j-term approximations by $\mathbf{v}_{[j]}:=\mathbf{v}_{2^j}$ and define

$$\mathbf{w}_j:=\mathbf{A}_j\mathbf{v}_{[0]}+\mathbf{A}_{j-1}(\mathbf{v}_{[1]}-\mathbf{v}_{[0]})+\cdots+\mathbf{A}_0(\mathbf{v}_{[j]}-\mathbf{v}_{[j-1]}), \tag{5.7}$$

as an approximation to $\mathbf{Av}$. Obviously this scheme is *adaptive* in that it exploits directly information on $\mathbf{v}$. In fact, if $\mathbf{A}\in\mathcal{C}_{s^*}$, then the triangle inequality together with the above compression estimates yield for any fixed $s<s^*$

$$\|\mathbf{Av}-\mathbf{w}_j\|_{\ell_2(\mathcal{J})}\le c\underbrace{\|\mathbf{v}-\mathbf{v}_{[j]}\|_{\ell_2(\mathcal{J})}}_{\sigma_{2^j,\ell_2(\mathcal{J})}(\mathbf{v})}+\sum_{l=0}^{j}\alpha_l 2^{-ls}\underbrace{\|\mathbf{v}_{[j-l]}-\mathbf{v}_{[j-l-1]}\|_{\ell_2(\mathcal{J})}}_{\lesssim\ \sigma_{2^{j-l-1},\ell_2(\mathcal{J})}(\mathbf{v})}. \tag{5.8}$$

One can now exploit the *a-posteriori* information offered by the quantities $\sigma_{2^{j-l-1},\ell_2(\mathcal{J})}(\mathbf{v})$ to choose the smallest j for which the right hand side of (5.8) is smaller than a given target accuracy η and set $\mathbf{w}_\eta:=\mathbf{w}_j$. Since the sum is finite for each finitely supported input $\mathbf{v}$ such a j does indeed exist. This leads to a concrete multiplication scheme (see [7, 2] for a detailed description, analysis and implementation) which we summarize as follows:

APPLY $[\eta,\mathbf{A},\mathbf{v}]\ \to\ \mathbf{w}_\eta$: *determines for any finitely supported input* $\mathbf{v}$ *a finitely supported output* $\mathbf{w}_\eta$ *such that*

$$\|\mathbf{Av}-\mathbf{w}_\eta\|_{\ell_2(\mathcal{J})}\le\eta. \tag{5.9}$$

To describe the work/accuracy rate of **APPLY** in precise terms, we recall from [7] the following characterization of those sequences in $\ell_2(\mathcal{J})$ whose error of best N-term approximation decays like N^{-s} for some $s > 0$. For any $\tau < 2$ let

$$|\mathbf{v}|_{\ell_\tau^w(\mathcal{J})} := \sup_n n^{1/\tau} |v_n^*|, \quad \|\mathbf{v}\|_{\ell_\tau^w(\mathcal{J})} := \|\mathbf{v}\|_{\ell_2(\mathcal{J})} + |\mathbf{v}|_{\ell_\tau^w(\mathcal{J})} \tag{5.10}$$

and let $\ell_\tau^w(\mathcal{J})$ denote the subspace of those $\mathbf{v} \in \ell_2(\mathcal{J})$ for which $\|\mathbf{v}\|_{\ell_\tau^w(\mathcal{J})} < \infty$. (Obviously one obtains a strict subspace of $\ell_2(\mathcal{J})$ only when $\tau < 2$.) It is easy to see that

$$\ell_\tau \subset \ell_\tau^w(\mathcal{J})$$

while the difference between both spaces is rather small. $\ell_\tau^w(\mathcal{J})$ is often called *weak* ℓ_τ due to the following alternate characterization, see e.g. [7, 19]:

$$\ell_\tau^w(\mathcal{J}) = \{\mathbf{v} \in \ell_2(\mathcal{J}) : \exists\ C \text{ such that } \ \#\{\lambda : |v_\lambda| > \eta\} \le C\eta^{-\tau},\ \eta > 0\}. \tag{5.11}$$

In fact, the smallest constant $C = C(\mathbf{v})$ above satisfies

$$C(\mathbf{v}) = |\mathbf{v}|^\tau_{\ell_\tau^w(\mathcal{J})}. \tag{5.12}$$

We will make use of the following fact, see [7].

PROPOSITION 5.1. *Let*

$$\frac{1}{\tau} = s + \frac{1}{2}. \tag{5.13}$$

Then $\mathbf{v} \in \ell_\tau^w(\mathcal{J})$ *if and only if* $\sigma_{N,\ell_2(\mathcal{J})}(\mathbf{v}) \lesssim N^{-s}$. *In addition,*

$$\|\mathbf{v} - \mathbf{v}_N\|_{\ell_2(\mathcal{J})} \lesssim N^{-s}\|\mathbf{v}\|_{\ell_\tau^w(\mathcal{J})}, \quad N = 1, 2, \ldots. \tag{5.14}$$

The main result concerning **APPLY** can be formulated as follows [7].

THEOREM 5.1. *Suppose that* $\mathbf{C} \in \mathcal{C}_{s^*}$ *and that for some* $0 < s < s^*$, $\mathbf{v} \in \ell_\tau^w(\mathcal{J})$, *where* $\frac{1}{\tau} = s + \frac{1}{2}$. *Then* $\mathbf{w}_\eta = \mathbf{APPLY}\,[\eta, \mathbf{C}, \mathbf{v}]$ *satisfies:*

(i) $\|\mathbf{w}_\eta\|_{\ell_\tau^w(\mathcal{J})} \lesssim \|\mathbf{v}\|_{\ell_\tau^w(\mathcal{J})}$;

(ii) $\#\text{flops} \sim \#\text{supp}\,\mathbf{w}_\eta \lesssim \|\mathbf{v}\|^{1/s}_{\ell_\tau^w(\mathcal{J})}\eta^{-1/s}$,

where the constants in these estimates depend only on s *when* s *is small. In particular, we see that* $\mathbf{C}$ *is bounded on* $\ell_\tau^w(\mathcal{J})$, $\frac{1}{\tau} = s + \frac{1}{2}$, *as long as* $s < s^*$ *and* τ *are related by (5.13)*

Thus, when dealing with linear problems (3.3), an approximation within the tolerance $\eta > 0$ for to the weighted residual $\mathbf{BR}(\mathbf{v}) = \alpha(\mathbf{A}\mathbf{v} - \mathbf{f})$ for any finitely supported input $\mathbf{v}$ can be computed as follows

$$\mathbf{RES}_{\text{lin}}[\eta, \alpha, \mathbf{A}, \mathbf{f}, \mathbf{v}] := \alpha\left(\mathbf{APPLY}\,[\frac{\eta}{2\alpha}, \mathbf{A}, \mathbf{v}] - \mathbf{RHS}\,[\frac{\eta}{2\alpha}, \mathbf{f}]\right), \tag{5.15}$$

where **RHS** is given by (5.2). The same ideas can be used in the least squares case (3.8), where again **RHS** can be composed of **COARSE** and **APPLY**, see [8] for details.

REMARK 5.2. *Since by Theorem 5.1* $\mathbf{f} \in \ell_\tau^w(\mathcal{J})$, *whenever the solution* $\mathbf{u}$ *belongs to* $\ell_\tau^w(\mathcal{J})$, *the above considerations and analogous facts about* **COARSE** *from [7] show that the output* $\mathbf{f}_\eta$ *of* $\mathbf{RHS}\,[\eta, \mathbf{f}]$ *satisfies* $\|\mathbf{f}_\eta\|_{\ell_\tau^w(\mathcal{J})} \lesssim \|\mathbf{u}\|_{\ell_\tau^w(\mathcal{J})}$ *and* $\#\text{supp}\,\mathbf{f}_\eta \lesssim \eta^{-1/s}\|\mathbf{u}\|^{1/s}_{\ell_\tau^w(\mathcal{J})}$. These observations provide the following result.

PROPOSITION 5.3. *If the sequence of wavelet coefficients* $\mathbf{u}$ *of the exact solution* u *of (3.3) belongs to* $\ell_\tau^w(\mathcal{J})$ *and if* $\mathbf{A}$ *belongs to* $\mathcal{C}_{s^*}$ *where* τ *is related to some* $s <$

s^* by (5.13), then, for any finitely supported input $\mathbf{v}$, the output $\mathbf{w}_\eta$ of the scheme $\mathbf{RES}_{\rm lin}[\eta, \alpha, \mathbf{A}, \mathbf{f}, \mathbf{v}]$ satisfies

$$\begin{aligned} \|\mathbf{w}_\eta\|_{\ell_\tau^w(\mathcal{J})} &\lesssim \left(\|\mathbf{v}\|_{\ell_\tau^w(\mathcal{J})} + \|\mathbf{u}\|_{\ell_\tau^w(\mathcal{J})}\right), \\ \#\,\mathrm{supp}\,\mathbf{w}_\eta &\lesssim \eta^{-1/s}\left(\|\mathbf{u}\|^{1/s}_{\ell_\tau^w(\mathcal{J})} + \|\mathbf{v}\|^{1/s}_{\ell_\tau^w(\mathcal{J})}\right), \quad \eta > 0, \end{aligned} \tag{5.16}$$

where the constants in these estimates depend only on s.

The only missing link now is to explain the role of step (iii) in **SOLVE**. It is based on the following result from [7].

PROPOSITION 5.4. *If $\mathbf{v} \in \ell_\tau^w(\mathcal{J})$ and $\|\mathbf{v} - \mathbf{w}\|_{\ell_2(\mathcal{J})} \le \eta$ with $\#\mathrm{supp}\,\mathbf{w} < \infty$. Then $\bar{\mathbf{w}}_\eta := \mathbf{COARSE}[\mathbf{w}, 4\eta]$ satisfies*

$$\#\mathrm{supp}\,\bar{\mathbf{w}}_\eta \lesssim \|\mathbf{v}\|^{1/s}_{\ell_\tau^w(\mathcal{J})}\eta^{-1/s}, \quad \|\mathbf{v} - \bar{\mathbf{w}}_\eta\|_{\ell_2(\mathcal{J})} \le 5\eta, \tag{5.17}$$

and

$$\|\bar{\mathbf{w}}_\eta\|_{\ell_\tau^w(\mathcal{J})} \lesssim \|\mathbf{v}\|_{\ell_\tau^w(\mathcal{J})}, \tag{5.18}$$

where the constant in these estimates depend only on s when s becomes small.

Thus, in this case $C^* = 4/3$ is a valid choice in the definition of the constant K in (4.8).

By Proposition 5.4, the coarsening step (iii), with the above algorithm **COARSE** used as **CCOARSE**, pulls a current approximation to the unknown $\mathbf{u}$ towards its best N-term approximation and controls the $\ell_\tau^w(\mathcal{J})$ norms of the approximations, while Proposition 5.3 controls the complexity within each iteration block (ii).

Let us now denote by $\mathbf{SOLVE}_{\rm lin}$ the specification of **SOLVE** obtained by using $\mathbf{RES}_{\rm lin}$ and **COARSE** in place of **RES**, respectively **CCOARSE**. We emphasize that adaptivity enters the scheme $\mathbf{SOLVE}_{\rm lin}$ solely through the adaptive application of $\mathbf{A}$.

In order to describe the complexity of $\mathbf{SOLVE}_{\rm lin}$, we define, in analogy to (5.3), the error of best N-term approximation in $\mathcal{H}$ as

$$\sigma_{N,\mathcal{H}}(u) := \min_{\#\Lambda \le N, \mathbf{v} \in \mathbb{R}^{\#\Lambda}} \|u - \sum_{\lambda \in \Lambda} v_\lambda \psi_\lambda\|_{\mathcal{H}}.$$

It follows from the norm equivalence (2.5) that

$$c_1 \sigma_{N,\ell_2(\mathcal{J})}(\mathbf{u}) \le \sigma_{N,\mathcal{H}}(u) \le C_1 \sigma_{N,\ell_2(\mathcal{J})}(\mathbf{u}). \tag{5.19}$$

The main result of this section follows now from Propositions 5.3 and 5.4, see [8] for details.

THEOREM 5.2. *Assume that the wavelet representation $\mathbf{A}$ of the operator $\mathcal{A}$ induced by the bilinear form $a(\cdot,\cdot)$ belongs to $\mathcal{C}_{s^*}$ and that the solution $u = \sum_{\lambda \in \mathcal{J}} u_\lambda \psi_\lambda$ of (3.3) satisfies*

$$\sigma_{N,\mathcal{H}} \lesssim N^{-s}$$

for some $s < s^*$. Then the output $\bar{\mathbf{u}}(\epsilon)$ of $\mathbf{SOLVE}_{\rm lin}[\epsilon, \mathbf{R}]$ and its support $\Lambda(\epsilon)$ have the following properties:

$$\#\Lambda(\epsilon) \lesssim \epsilon^{-1/s}\|\mathbf{u}\|^{1/s}_{\ell_\tau^w(\mathcal{J})}, \quad \|\bar{\mathbf{u}}(\epsilon)\|_{\ell_\tau^w(\mathcal{J})} \lesssim \|\mathbf{u}\|_{\ell_\tau^w(\mathcal{J})}. \tag{5.20}$$

Moreover, the number of operations and storage needed to compute $\bar{\mathbf{u}}(\epsilon)$ remains proportional to $\#\Lambda(\epsilon)$ while

$$\|u - \sum_{\lambda \in \Lambda(\epsilon)} u(\epsilon)_\lambda \psi_\lambda\|_{\mathcal{H}} \le C_1 \epsilon. \tag{5.21}$$

Thus, under the above premises, $\mathbf{SOLVE}_{\mathrm{lin}}$ has in a certain range the same asymptotic work/accuracy rate as the best N-term approximation of the solution and exhibits in this sense optimal complexity.

REMARK 5.5. *We conclude this section by recalling that* $\mathbf{u} \in \ell_\tau^w(\mathcal{J})$ *is for instance implied by a certain Besov regularity of* u. *In fact, when* $\mathcal{H} = H^t$, $u \in B_\tau^{t+ds}(L_\tau)$ *with* $\tau^{-1} = s + 1/2$, *implies* $\mathbf{u} \in \ell_\tau^w(\mathcal{J})$. *This can be used to identify circumstances under which the adaptive scheme performs asymptotically better than a scheme based on uniform refinements. Recall that* $B_\tau^{t+ds}(L_\tau)$ *is the "largest" space of smoothness* $t + sd$ *imbedded in* H^t.

6. The Nonlinear Case. In view of the fact that **SOLVE** has the same structure, regardless of whether the involved operators are linear or nonlinear, our strategy will be to follow closely the above lines also when the variational problem (2.2) is nonlinear. In principle, this will prove successful although some important modifications of the ingredients will be encountered. The main distinction lies in the sparsity measure in that the role of best (unrestricted) N-term approximation will be replaced by *best tree approximation*. This constraint on the distribution of active coefficients arises naturally when analyzing the approximate evaluation of nonlinear expressions $\mathbf{R}(\mathbf{v})$.

6.1. Tree Approximation and Coarsening. To explain this, recall that $S_\lambda := \operatorname{supp}\psi_\lambda$. A set $\mathcal{T} \subset \mathcal{J}$ is called 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 P 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 P 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

$$\mathcal{L}^+ := \{\lambda \in \mathcal{T} : \mu \text{ a child of } \lambda \implies \mu \notin \mathcal{T}\}. \tag{6.1}$$

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

$$\mathcal{L}^- := \{\lambda \in \mathcal{J} : \lambda \notin \mathcal{T},\ S_\lambda \subset S_\mu \implies \mu \in \mathcal{T}\}. \tag{6.2}$$

We shall make use of the following facts.

REMARK 6.1. *a) One has*

$$\#\mathcal{T} \sim \#\mathcal{L}^+(\mathcal{T}) \sim \#\mathcal{L}^-(\mathcal{T}), \tag{6.3}$$

where the constants depend only on the number P *of parents.*
b) There exists an $M \in \mathbb{N}$ *such that at most* M *elements of* $\mathcal{L}^\pm(\mathcal{T})$ *have a nonempty intersection.*

Proof: a) is easily verified. As for b), 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. Recall that the collection $\mathcal{S}(\Psi)$ of wavelet supports is assumed to have the (FIP). The claim b) follows then from (2.4). □

Of course, regular thresholding does not generally preserve tree structures. This leads us to the following notions. We associate to any sequence $\mathbf{v} = (v_\lambda)$ in $\ell_2(\mathcal{J})$ another sequence $\tilde{\mathbf{v}} = (\tilde{v}_\lambda)$ defined by

$$\tilde{v}_\lambda := \left(\sum_{S_\mu \subseteq S_\lambda, |\mu| \geq |\lambda|} |v_\mu|^2 \right)^{1/2}. \tag{6.4}$$

One readily confirms that $S_\mu \subset S_\lambda$ implies $\tilde{v}_\lambda \geq \tilde{v}_\mu$, i.e. for any $\eta > 0$ the set

$$\mathcal{T}_\eta = \mathcal{T}_\eta(\mathbf{v}) := \{\lambda \,:\, |\tilde{v}_\lambda| > \eta\} \tag{6.5}$$

has a *tree structure*. Thus, *thresholding* with respect to the modified sequences $\tilde{\mathbf{u}}$ creates trees.

Recall that in the linear case we have been able to compare the performance of **SOLVE** with the *best (unconstrained) N-term approximation*. We want now to develop sparsity measures that respect tree structure. Once a suitable measure has been identified one can follow conceptually the lines of Section 3. The counterpart for the spaces $\ell^w_\tau(\mathcal{J})$, entering the characterization of best (unconstrained) N-term approximation rates (recall Proposition 5.1), are now the following spaces (see [9])

$$ {}_t\ell^w_\tau(\mathcal{J}) := \{\mathbf{v} \in \ell_2(\mathcal{J}) : \tilde{\mathbf{v}} \in \ell^w_\tau(\mathcal{J})\}, \quad \|\mathbf{v}\|_{{}_t\ell^w_\tau(\mathcal{J})} := \|\tilde{\mathbf{v}}\|_{\ell^w_\tau(\mathcal{J})}. \tag{6.6}$$

Thus, by (5.11) and (5.12), one has

$$\#\mathcal{T}_\eta(\mathbf{v}) \leq \eta^{-\tau}|\mathbf{v}|^\tau_{{}_t\ell^w_\tau(\mathcal{J})}, \quad \mathbf{v} \in {}_t\ell^w_\tau(\mathcal{J}). \tag{6.7}$$

In analogy to Proposition 5.1 one has, [9]

PROPOSITION 6.2. *Let $\mathbf{v}_\eta := \mathbf{v}|_{\mathcal{T}_\eta}$. Then $\mathbf{v} \in {}_t\ell^w_\tau(\mathcal{J})$ implies the error estimate*

$$\|\mathbf{v} - \mathbf{v}_\eta\| \lesssim \eta^{1-\tau/2}|\mathbf{v}|^{\tau/2}_{{}_t\ell^w_\tau(\mathcal{J})} \lesssim [\#(\mathcal{T}_\eta)]^{-s}|\mathbf{v}|_{{}_t\ell^w_\tau(\mathcal{J})}, \tag{6.8}$$

with $s = 1/\tau - 1/2$, see (5.13). Conversely, $\|\mathbf{v} - \mathbf{v}_\eta\| \leq C[\#(\mathcal{T}_\eta)]^{-s}$ for all $\eta > 0$ implies $\mathbf{v} \in {}_t\ell^w_\tau(\mathcal{J})$ with $\|\mathbf{v}\|_{{}_t\ell^w_\tau(\mathcal{J})} \lesssim C$.

One can again relate the membership of $\mathbf{v}$ to ${}_t\ell^w_\tau(\mathcal{J})$ to the regularity of the corresponding expansion v, [9].

REMARK 6.3. *Let $\mathcal{H} = H^t$ for some $t > 0$. If the wavelet expansion v with coefficient sequence $\mathbf{v}$ belongs to $B^{t+sd}_{\tau'}(L_{\tau'})$ for some τ' satisfying $\tau' > (s+1/2)^{-1/2}$, then $\mathbf{v} \in {}_t\ell^w_\tau(\mathcal{J})$ with $\tau = (s+1/2)^{-1/2}$. Thus, in terms of regularity, ${}_t\ell^w_\tau(\mathcal{J})$ differs from $\ell^w_\tau(\mathcal{J})$ by a little additional regularity imposed on the respective expansions, due to the stronger metric $L_{\tau'}$, $\tau' > \tau$. Thus a tree approximation rate N^{-s} can still be achieved for much larger spaces than H^{t+sd} which governs the corresponding rate for uniform refinements.*

6.2. Tree Coarsening. We shall specify next a coarsening routine **CCOARSE** that preserves *tree structures* and, as before, applies to finitely supported sequences. It will be referred to as **TCOARSE**. Its definition requires some preparation. Given $\mathbf{w}$, a tree $\mathcal{T} = \mathcal{T}^*(\eta, \mathbf{w})$ is called η-best for $\mathbf{w}$ if

$$\|\mathbf{w} - \mathbf{w}|_\mathcal{T}\|_{\ell_2(\mathcal{J})} \leq \eta \;\text{ and }\; \#\mathcal{T} = \min\{\#\mathcal{T}' : \|\mathbf{w} - \mathbf{w}|_{\mathcal{T}'}\|_{\ell_2(\mathcal{J})} \leq \eta,\ \mathcal{T}' \text{ a tree}\}.$$

Requiring best trees will be too stringent from a practical point of view. Therefore we shall be content with the following relaxed version. A tree $\mathcal{T} = \mathcal{T}(\eta, \mathbf{w})$ is called (η, C)-*near best* (or briefly near best when the parameters are clear from the context) if

$$\|\mathbf{w} - \mathbf{w}|_\mathcal{T}\|_{\ell_2(\mathcal{J})} \leq \eta \;\text{ and }\; \#\mathcal{T} \leq C\#\mathcal{T}^*(\eta/C, \mathbf{w}).$$

The action of **TCOARSE** can now be described as follows.

TCOARSE $[\eta, \mathbf{w}] \to \bar{\mathbf{w}}_\eta$ *determines for a fixed constant $C^* \geq 1$, any finitely supported input $\mathbf{w}$, and any tolerance $\eta > 0$ an (η, C^*)-near best tree $\mathcal{T}(\eta, \mathbf{w})$ and sets $\bar{\mathbf{w}}_\eta := \mathbf{w}|_{\mathcal{T}(\eta,\mathbf{w})}$.*

The realization of this routine can be based on the **second algorithm** for generating near best tree approximations developed in [3]. To apply the results from [3] in

the present situation, the role of the partition P associated in [3] to a tree $\mathcal{T}$ is played here by the set $\mathcal{L}^-(\mathcal{T})$ of outer leaves, while for $\lambda \in \mathcal{L}^-(\mathcal{T})$ the local error terms for a given $\mathbf{v} \in \ell_2(\mathcal{J})$ are here given by $e(\lambda) := \tilde{v}_\lambda^2$. Since, by Remark 6.1 b), at any given point the number of $\lambda \in \mathcal{L}^-$ that overlap that point is finite we see that the $e(\lambda)$ are subadditive in the sense of [3]. Hence the results from [3] apply. To use the algorithm from [3] we need to know the values $\tilde{w}_\lambda$, $\lambda \in \mathcal{T}(\operatorname{supp} \mathbf{w})$, the smallest tree containing the support of $\mathbf{w}$. Summing the squares of the entries of $\mathbf{w}$ starting from the leaves of $\mathcal{T}(\operatorname{supp} \mathbf{w})$ working towards the roots, provides these quantities at an expense of $\#\mathcal{T}(\operatorname{supp} \mathbf{w})$ operations. Combining this with Theorem 5.2 from [3] establishes the following fact.

PROPOSITION 6.4. *For any given finitely supported input* $\mathbf{w}$ *the computational cost of the output* $\bar{\mathbf{w}}_\eta$ *produced by* **TCOARSE**$[\eta, \mathbf{w}]$ *remains proportional to* $\#\mathcal{T}(\operatorname{supp} \mathbf{w})$. *The underlying tree* $\mathcal{T}(\eta, \mathbf{w})$ *is* (η, C^*)*-near best, where* C^* *is the constant appearing in the estimate (5.8) in Theorem 5.2 of [3].*

The routine **TCOARSE** will be used as **CCOARSE** in step (iii) of **SOLVE**. The constant C^* enters the definition of the number K of perturbed iterations in step (ii) of **SOLVE** (see (4.8)). As in the linear case its purpose is to control the ${}_t\ell_\tau^w(\mathcal{J})$-norms of the approximants. This is made precise by the following counterpart to Proposition 5.4.

PROPOSITION 6.5. *If* $\mathbf{v} \in {}_t\ell_\tau^w(\mathcal{J})$ *and* $\|\mathbf{v} - \mathbf{w}\|_{\ell_2(\mathcal{J})} \leq \eta$ *with* $\#\operatorname{supp} \mathbf{w} < \infty$. *Then* $\bar{\mathbf{w}}_\eta := \mathbf{TCOARSE}[\mathbf{w}, 3C^*\eta]$ *satisfies*

$$\#\operatorname{supp} \bar{\mathbf{w}}_\eta \lesssim \|\mathbf{v}\|_{{}_t\ell_\tau^w(\mathcal{J})}^{1/s} \eta^{-1/s}, \quad \|\mathbf{v} - \bar{\mathbf{w}}_\eta\|_{\ell_2(\mathcal{J})} \leq (1 + 3C^*)\eta, \tag{6.9}$$

and

$$\|\bar{\mathbf{w}}_\eta\|_{{}_t\ell_\tau^w(\mathcal{J})} \lesssim \|\mathbf{v}\|_{{}_t\ell_\tau^w(\mathcal{J})}, \tag{6.10}$$

where the constants depend only on τ *when* $\tau \to 0$, *on* C^* *in* **TCOARSE** *and the constants from Proposition 6.2.*

Proof: The second estimate in (6.9) follows from the triangle inequality. As for the first estimate in (6.9), assume that $\mathbf{v} \in {}_t\ell_\tau^w(\mathcal{J})$ and consider the thresholding tree $\mathcal{T}_\delta(\mathbf{v})$ for $\mathbf{v}$ defined by (6.5). We shall show that for a judiciously chosen δ the restriction of $\mathbf{w}$ to this tree provides a 3η-accurate approximation to $\mathbf{w}$, i.e. one has

$$\|\mathbf{w} - \mathbf{w}|_{\mathcal{T}_\delta(\mathbf{v})}\|_{\ell_2(\mathcal{J})} \leq 3\eta. \tag{6.11}$$

Therefore, by the definition of the (η, C^*)-near best tree $\mathcal{T}(3C^*\eta, \mathbf{w})$ and by (6.7), we have

$$\#\mathcal{T}(3C^*\eta, \mathbf{w}) \leq C^*\#\mathcal{T}^*(3\eta, \mathbf{w}) \leq C^*\#\mathcal{T}_\delta(\mathbf{v}) \leq C^*\delta^{-\tau}\|\mathbf{v}\|_{{}_t\ell_\tau^w(\mathcal{J})}^\tau. \tag{6.12}$$

To determine $\delta > 0$ for which (6.11) holds, note first that

$$\begin{aligned} \|\mathbf{w} - \mathbf{w}|_{\mathcal{T}_\delta(\mathbf{v})}\|_{\ell_2(\mathcal{J})} &\leq \|\mathbf{w} - \mathbf{v}\|_{\ell_2(\mathcal{J})} + \|\mathbf{v} - \mathbf{v}_{\mathcal{T}_\delta(\mathbf{v})}\|_{\ell_2(\mathcal{J})} + \|(\mathbf{v} - \mathbf{w})|_{\mathcal{T}_\delta(\mathbf{v})}\|_{\ell_2(\mathcal{J})} \\ &\leq 2\eta + C\|\mathbf{v}\|_{{}_t\ell_\tau^w(\mathcal{J})}^{\tau/2}\delta^{1-\tau/2}, \end{aligned} \tag{6.13}$$

where we have invoked Proposition 6.2. Now choose δ such that $\eta = C\|\mathbf{v}\|_{{}_t\ell_\tau^w(\mathcal{J})}^{\tau/2}\delta^{1-\tau/2}$ so that (6.11) is indeed valid. Recalling now that for s and τ related by $\tau = (s+1/2)^{-1}$, one has $1 - \tau/2 = s\tau$ and hence $\delta = C^{-1/(\tau s)}\|\mathbf{v}\|_{{}_t\ell_\tau^w(\mathcal{J})}^{-1/2s}\eta^{1/(s\tau)}$ which gives $\delta^{-\tau} = C^{1/s}\|\mathbf{v}\|^{\tau/2s}\eta^{-1/s}$. Substituting this into (6.12), and noting that $\frac{1}{2s}+1 = \frac{1}{s\tau}$, confirms the first estimate in (6.9).

It remains to prove (6.10). It suffices to show that $\|\mathbf{z}\|_{{}_t\ell_\tau^w(\mathcal{J})} \lesssim \|\mathbf{v}\|_{{}_t\ell_\tau^w(\mathcal{J})}$ with $\mathbf{z} := \mathbf{v} - \bar{\mathbf{w}}_\eta$. We shall show that

$$N^{1/\tau}\tilde{z}_N^* \lesssim \|\mathbf{v}\|_{{}_t\ell_\tau^w(\mathcal{J})}, \quad N = 1, 2, 3, \ldots, \tag{6.14}$$

which, in view of the definition of $\|\mathbf{z}\|_{{}_t\ell^w_\tau(\mathcal{J})}$, will complete the proof. We consider two cases.

Case 1) $N > 4\#\mathcal{T}(3C^*\eta, \mathbf{w}) := m$: First note that $\tilde{z}_\lambda = \tilde{v}_\lambda$ except for $\lambda \in \mathcal{T}(3C^*\eta, \mathbf{w}) = \operatorname{supp} \bar{\mathbf{w}}_\eta$. Thus, we can view the sequence $\tilde{z}^*_j$ as obtained from $\tilde{v}^*_j$ by changing at most m of these values and then rearranging to put them in their proper order. It follows that one of the $\tilde{v}^*_j$, $j = N-2m, \dots, N$ was not changed and when rearranged becomes $\tilde{z}^*_{j'} = \tilde{v}^*_j$ with $|j'-j| \le m$. Hence

$$N^{1/\tau}\tilde{z}^*_N \le N^{1/\tau}\tilde{z}^*_{j'} = N^{1/\tau}\tilde{v}^*_j \le 4^{1/\tau} j^{1/\tau}\tilde{v}^*_j \le 4^{1/\tau}\|\mathbf{v}\|_{{}_t\ell^w_\tau(\mathcal{J})},$$

which verifies (6.14) in this case.

Case 2) $N \le 4\#\mathcal{T}(3C^*\eta, \mathbf{w})$: In this case, let $\mathcal{T}^*$ be a subtree of $\mathcal{T}_{\tilde{z}^*_N}(\mathbf{z})$ obtained by possibly discarding some of the nodes λ with $\tilde{z}_\lambda = \tilde{z}^*_N$ so as to guarantee that $\#\mathcal{T}^* = N$. We let $\mathcal{L}^+ := \mathcal{L}^+(\mathcal{T}^*)$ be the set of inner leaves of $\mathcal{T}^*$ so that, by Remark 6.1 a), $cN \le \#\mathcal{L}^+ \le N$ holds for some constant $c > 0$ depending only on the supports S_λ of the wavelets $\psi_\lambda \in \Psi$. Moreover, by Remark 6.1 b), only a uniformly bounded finite number of supports S_ν, $\nu \in \mathcal{L}^+$, overlap any given point. Then, by Remark 6.1 a) and Hölder's inequality,

$$\begin{aligned} N|\tilde{z}^*_N|^\tau &\lesssim \sum_{\nu\in\mathcal{L}^+} |\tilde{z}_\nu|^\tau \le \left(\sum_{\nu\in\mathcal{L}^+} |\tilde{z}_\nu|^2\right)^{\tau/2} \left(\sum_{\nu\in\mathcal{L}^+} 1\right)^{1-\tau/2} \\ &\lesssim \left(\sum_{\nu\in\mathcal{L}^+} |\tilde{z}_\nu|^2\right)^{\tau/2} N^{s\tau} \lesssim \left(\|\mathbf{z}\|_{\ell_2(\mathcal{J})} N^s\right)^\tau, \end{aligned} \tag{6.15}$$

where we have used Remark 6.1 b) in the last step. Since by (6.11), $\|\mathbf{z}\|_{\ell_2(\mathcal{J})} = \|\mathbf{v} - \bar{\mathbf{w}}_\eta\|_{\ell_2(\mathcal{J})} \le 3C^*\eta$, and since by the first estimate in (6.9), $\eta \lesssim \|\mathbf{v}\|_{{}_t\ell^w_\tau(\mathcal{J})}(\#\mathcal{T}(\eta, \mathbf{w}))^{-s}$, we conclude from (6.15) that

$$N^{1/\tau}|\tilde{z}^*_N| \lesssim \eta N^s \lesssim \|\mathbf{v}\|_{{}_t\ell^w_\tau(\mathcal{J})},$$

which finishes the proof. □

6.3. The Key Requirement. Up to this point we have not imposed any conditions on the subroutine **RES** which is used to approximate the residual at each iteration. We will now introduce a condition, called τ^*-sparsity, motivated by the analysis of the previous section for the linear case, see Proposition 5.3. We will then show that whenever **RES** is τ^*-sparse then the algorithm **SOLVE** is optimal in its rate/complexity for a certain range of error decay rates depending on τ^*. The subsequent section will then show how to construct τ^*-sparse routines for nonlinear problems.

We say that the scheme **RES** used to approximate residuals is *τ^*-sparse* if the following property holds.

τ^*-Sparsity: *Whenever the exact solution* $\mathbf{u}$ *of (2.9) belongs to* ${}_t\ell^w_\tau(\mathcal{J})$ *for some* $\tau > \tau^*$, *then one has for any finitely supported input* $\mathbf{v}$ *and any tolerance* $\eta > 0$ *that the output* $\mathbf{w}_\eta := \mathbf{RES}[\eta, \mathbf{R}, \mathbf{v}]$ *satisfies*

$$\begin{aligned} \#\operatorname{supp}\mathbf{w}_\eta &\le C\eta^{-1/s}\left(\|\mathbf{v}\|^{1/s}_{{}_t\ell^w_\tau(\mathcal{J})} + \|\mathbf{u}\|^{1/s}_{{}_t\ell^w_\tau(\mathcal{J})} + 1\right), \\ \|\mathbf{w}_\eta\|_{{}_t\ell^w_\tau(\mathcal{J})} &\le C\left(\|\mathbf{v}\|_{{}_t\ell^w_\tau(\mathcal{J})} + \|\mathbf{u}\|_{{}_t\ell^w_\tau(\mathcal{J})} + 1\right), \end{aligned} \tag{6.16}$$

where C *depends only on* τ *when* $\tau \to \tau^*$. *Moreover, the number of operations needed to compute* $\mathbf{w}_\eta$ *stays proportional to* $\#\operatorname{supp}\mathbf{w}_\eta$.

The occurence of $\|\mathbf{u}\|_{{}_t\ell^w_\tau(\mathcal{J})}$ in the above estimates is already plausible from the linear case as expained in Remark 5.2.

It will be understood in the sequel that **TCOARSE** is used as **CCOARSE** and that a proper initialization is used that complies if necessary with the requirements on the quality of the initial guess (see Section 5) so that, in particular, the respective variant of the iteration (2.10) satisfies (3.7).

Under these premises we now show that τ^*-sparsity implies asymptotically optimal complexity of the scheme **SOLVE**.

THEOREM 6.1. *Assume that the scheme* **RES** *is τ^*-sparse for some $\tau^* > 0$. If the exact solution $\mathbf{u}$ of (2.9) belongs to ${}_t\ell^w_\tau(\mathcal{J})$ for some $\tau > \tau^*$, then the approximations $\bar{\mathbf{u}}(\epsilon)$ satisfy for every target accuracy $\epsilon > 0$*

$$\|\mathbf{u} - \bar{\mathbf{u}}(\epsilon)\|_{\ell_2(\mathcal{J})} \leq \epsilon, \tag{6.17}$$

while

$$\#\mathrm{supp}\,\bar{\mathbf{u}}(\epsilon) \leq C\epsilon^{-1/s}\|\mathbf{u}\|^{1/s}_{{}_t\ell^w_\tau(\mathcal{J})}, \quad \|\bar{\mathbf{u}}(\epsilon)\|_{{}_t\ell^w_\tau(\mathcal{J})} \leq C\|\mathbf{u}\|_{{}_t\ell^w_\tau(\mathcal{J})}, \tag{6.18}$$

where the constant C depends only on τ when $\tau \to \tau^*$. Moreover, the number of operations needed to compute $\bar{\mathbf{u}}(\epsilon)$ stays proportional to $\#\mathrm{supp}\,\mathbf{w}_\eta$.

Proof: The first part follows directly from Proposition 4.2. From (4.13) we know that the result $\mathbf{v}^K$ after K perturbed iterations in the $j+1$st block of step (ii) in **SOLVE** satisfies $\|\mathbf{u} - \mathbf{v}^K\|_{\ell_2(\mathcal{J})} \leq \epsilon_j/(2(1+3C^*))$. Now Proposition 6.5 ensures that then

$$\|\bar{\mathbf{u}}^{j+1}\|_{{}_t\ell^w_\tau(\mathcal{J})} \leq C\|\mathbf{u}\|_{{}_t\ell^w_\tau(\mathcal{J})}, \quad \#\mathrm{supp}\,\bar{\mathbf{u}}^{j+1} \leq C\epsilon_j^{-1/s}\|\mathbf{u}\|^{1/s}_{{}_t\ell^w_\tau(\mathcal{J})}, \tag{6.19}$$

while the computational work stays proportional to the support size of $\mathbf{v}^K$. Here it is important to note that the constant C is *independent* of the input $\mathbf{v}^K$ of **TCOARSE**. Thus for $j > 0$ the input of the first application of **RES** in step (ii) of **SOLVE** satisfies (6.19). Since there are only a uniformly bounded number K of applications of **RES** in each iteration block, $\mathbf{v}^K$ also satisfies (6.19) with a constant depending now on K. (Here we have tacitly assumed that the initial guess has been subjected to a **TCOARSE** so that it also satisfies (6.19). Otherwise, we would have to add $\#\mathrm{supp}\,\mathbf{u}^0$ to the above estimates.) The estimate in (6.18) now follows from these estimates for the terminal value of j. This also shows that the number of operations remains proportional to $\#\mathrm{supp}\,\bar{\mathbf{u}}(\epsilon)$. □

We shall discuss below how to obtain schemes **RES** that are τ^*-sparse for certain τ^* and what limits the value of τ^*.

7. Nonlinear Evaluation Schemes. Just as the efficient application of compressible matrices $\mathbf{A}$ was pivotal for the adaptive solution of (3.3), we need efficient evaluation schemes for $\mathbf{F}(\mathbf{v})$ that allow us to realize the residual approximation in **RES**. Such a scheme has been already proposed in [9] for a class of nonlinearities F that will be described next.

7.1. A Class of Nonlinear Mappings. We shall be concerned with nonlinear operators of the form

$$V = (v_1, \cdots, v_n) \mapsto w = F(D^{\alpha_1}v_1, \cdots, D^{\alpha_n}v_n), \tag{7.1}$$

acting from $\mathcal{H} \times \cdots \times \mathcal{H}$ to the dual $\mathcal{H}'$ (here $\alpha_i = (\alpha_{i,1}, \cdots, \alpha_{i,d})$ are multi-indices). This clearly covers our previous example of a single argument $n = 1$ but also further important cases like the nonlinearity appearing in the Navier-Stokes equations. Although we shall not address variational problems involving nonlinearities of several arguments in this paper we shall present the evaluation schemes in this somewhat

greater generality because they are important for such applications and because we shall apply the case of two arguments later in connection with Newton's scheme.

We shall first describe our requirements on F in the wavelet coordinate domain and point out later circumstances under which these requirements are met.

Denoting by $\mathbf{v}_i = (v_{i,\lambda})$ the arrays of the wavelet coefficients of the function v_i, $\mathbf{V} = (\mathbf{v}_1, \cdots, \mathbf{v}_n)$ and $\mathbf{F}$ the corresponding discrete mapping

$$\mathbf{F}(\mathbf{V}) := (\langle \psi_\lambda, F(D^{\alpha_1} v_1, \cdots, D^{\alpha_n} v_n) \rangle)_{\lambda \in \mathcal{J}}, \tag{7.2}$$

we make the following basic assumptions.

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

$$\|\mathbf{F}(\mathbf{U}) - \mathbf{F}(\mathbf{V})\|_{\ell_2(\mathcal{J})} \le C \sum_{i=1}^{n} \|\mathbf{u}_i - \mathbf{v}_i\|_{\ell_2(\mathcal{J})}, \tag{7.3}$$

with $C = C(\max_i\{\|\mathbf{u}_i\|_{\ell_2(\mathcal{J})}, \|\mathbf{v}_i\|_{\ell_2(\mathcal{J})}\})$*, where* $x \mapsto C(x)$ *is a positive non-decreasing function.*

Bearing the norm equivalences (2.5), (2.6) in mind, we see that property P1 from Section 3.2 is a special case of Assumption 1 for $n = 1$. For local operators one infers the following local version of this stability assumption (see [9])

$$\|(\mathbf{F}(\mathbf{U}) - \mathbf{F}(\mathbf{V}))_{\{\lambda : S_\lambda \subset D\}}\|_{\ell_2(\mathcal{J})} \le C \sum_{i=1}^{n} \|(\mathbf{u}_i - \mathbf{v}_i)_{\{\lambda : S_\lambda \cap D \neq \emptyset\}}\|_{\ell_2(\mathcal{J})}, \tag{7.4}$$

for any domain D.

Assumption 2. *There exists a constant* $\gamma > d/2$ *such that for any finitely supported* $\mathbf{V}$ *(i.e. with all* $\mathbf{v}_i$ *finitely supported) and* $\mathbf{w} = \mathbf{F}(\mathbf{V})$*, we have the estimate*

$$|w_\lambda| \le C \sup_{\mu \,:\, S_\lambda \cap S_\mu \neq \emptyset} \left(\sum_{i=1}^{n} |v_{i\mu}| \right) 2^{-\gamma(|\lambda| - |\mu|)}, \tag{7.5}$$

where $C = C(\max_i \|\mathbf{v}_i\|_{\ell_2(\mathcal{J})})$ *and* $x \mapsto C(x)$ *is a positive non-decreasing function.*

The parameter γ plays a similar role as the compressibility range s^* for the wavelet representation of linear operators.

Nonlinear mappings that satisfy the above assumptions are, for instance, those with polynomial growth. In the special case of a single argument a typical condition reads

$$|F^{(k)}(v)| \lesssim (1 + |v|)^{(p-k)_+}, \quad k = 0, \ldots, n^*, \qquad \text{for some } \; n^* \ge 1. \tag{7.6}$$

REMARK 7.1. *One can show that, for the special case* $n = 1$*, (7.6) implies Assumptions 1 and 2, and in particular property P1, with no condition on* p *when* $t \ge d/2$ *and otherwise provided that*

$$1 \le p < p^* := \frac{d + 2t}{d - 2t}, \tag{7.7}$$

see [9].

In the general case, when the nonlinear map F has the form $F(D^{\alpha_1} u_1, \cdots, D^{\alpha_n} u_n)$, we impose the growth condition

$$|D^\beta F(x_1, \cdots, x_n)| \le C \prod_{i=1}^{n} (1 + |x_i|)^{[p_i - \beta_i]_+}, \quad |\beta| = 0, 1, \cdots, n^*, \tag{7.8}$$

for some $p_i \geq 0$ and n^* a positive integer. The following fact (covering Remark 7.1) has been proven in [9].

THEOREM 7.1. *Suppose that $\mathcal{H} = H^t$ is (a closed subspace, determined e.g. by homogeneous boundary conditions, or equal to) $H^t(\Omega)$ for some $t \geq 0$. Assume that the growth assumptions (7.8) hold at least with $n^* = 0$. Then F maps $\mathcal{H} \times \cdots \times \mathcal{H}$ to $\mathcal{H}'$ whenever $t \geq 0$ satisfies*

$$[\frac{1}{2} - \frac{t}{d}]_+ + \sum_{i=1}^{n} p_i [\frac{1}{2} - \frac{t}{d} + \frac{|\alpha_i|}{d}]_+ < 1. \tag{7.9}$$

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

$$\|F(u) - F(v)\|_{\mathcal{H}'} \leq C \sum_{i=1}^{n} \|u_i - v_i\|_{\mathcal{H}}, \tag{7.10}$$

where $C = C(\max_i\{\|u_i\|_{\mathcal{H}}, \|v_i\|_{\mathcal{H}}\})$ and $x \to C(x)$ is nondecreasing, and therefore, on account of (2.5), Assumption 1 holds.

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

$$|D^\beta F(x_1, \cdots, x_n)| \leq C \prod_{i=1}^{n} (1 + |x_i|)^{p_i - \beta_i}, \quad \beta_i \leq p_i, \tag{7.11}$$

and $D^\beta F = 0$ if $\beta_i > p_i$ for some i, where the p_i are positive integers. We recall the following result from [9].

THEOREM 7.2. *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 F satisfies (7.8) 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 F satisfies (7.11) with p such that $\sum_{i=1}^{n} p_i[d/2 - t + |\alpha_i|]_+ < d/2 + t$, then $r = m$.

7.2. An Adaptive Evaluation Scheme. Given $\mathcal{T}_\eta(\mathbf{v}_i)$, $i = 1, \ldots, n$, our next objective is to *predict* the tree hull $\mathcal{T}_\eta(\mathbf{F}(\mathbf{V}))$ of significant coefficients of $\mathbf{w} = \mathbf{F}(\mathbf{V})$ for nonlinear mappings satisfying Assumptions 1 and 2. To this end, we follow [9] and fix $\eta > 0$. For the constant γ of (7.5), we define for all $\mu \in \mathcal{J}$ the number $n(\mu) = n(\mu, \mathbf{V})$ satisfying

$$\eta 2^{\gamma n(\mu, \mathbf{V})} \leq \max_{i=1,\ldots,n} |\tilde{v}_{i,\mu}| < \eta 2^{\gamma(n(\mu,\mathbf{V})+1)}, \tag{7.12}$$

where $\tilde{v}_{i,\mu}$ are the residuals for $\mathbf{v}_i$ (see (6.4)). Moreover, we define the *influence set*

$$\Lambda_{\eta,\mu}(\mathbf{V}) := \{\lambda \,:\, S_\lambda \cap S_\mu \neq \emptyset \text{ and } |\lambda| \leq |\mu| + [n(\mu, \mathbf{V})]_+\}. \tag{7.13}$$

To define a set of significant coefficients for the approximation of $\mathbf{F}(\mathbf{V})$ we need one further notion. Recall that $\mathcal{L}^-(\mathcal{T})$ denotes the set of *outer leaves* of $\mathcal{T}$, namely those $\lambda \in \mathcal{J}$ which do not belong to $\mathcal{T}$ but whose parents all belong to $\mathcal{T}$, see [9]. It can be shown that for any tree $\mathcal{T}$ there exists an expansion $\tilde{\mathcal{T}}$ such that for some constant C one has $\#\tilde{\mathcal{T}} \leq C\#\mathcal{T}$ while for any $\lambda \in \mathcal{L}^-(\tilde{\mathcal{T}})$ the number of $\mu \in \mathcal{L}^-(\tilde{\mathcal{T}})$ such that $S_\lambda \cap S_\mu \neq \emptyset$ is bounded by C, see Lemma 3.1 in [9]. Now let

$$\Lambda_\eta(\mathbf{V}) := \mathcal{J}_\phi \cup \bigcup_{\mu \in \tilde{\mathcal{T}}_\eta(\mathbf{V})} \Lambda_{\eta,\mu}(\mathbf{V}), \tag{7.14}$$

where $\tilde{\mathcal{T}}_\eta(\mathbf{V}) = \cup_{i=1}^n \tilde{\mathcal{T}}_\eta(\mathbf{v}_i)$, ($\tilde{\mathcal{T}}_\eta(\mathbf{v}_i)$ the expansion of $\mathcal{T}_\eta(\mathbf{v}_i)$) and as before $\mathcal{T}_\eta(\mathbf{v}_i)$ consists of all indices μ such that $\tilde{v}_{i,\mu} > \eta$. We notice that the sets $\Lambda_{\eta,\mu}(\mathbf{V})$ and hence $\Lambda_\eta(\mathbf{V})$ have tree structure.

The following fact has been shown in [9], Theorem 5.1.

PROPOSITION 7.2. *Assume that F satisfies Assumptions 1 and 2. Given any $\mathbf{V} \in (\ell_2(\mathcal{J}))^n$ let $\mathbf{w} = \mathbf{F}(\mathbf{V})$ and define $\tilde{\mathbf{w}}$ according to (6.4). Then there exists a constant C independent of $\mathbf{V}$ such that*

$$|\tilde{w}_\lambda| \leq C\eta \text{ if } \lambda \notin \Lambda_\eta(\mathbf{V}). \tag{7.15}$$

For any set $\Lambda \subset \mathcal{J}$ and any $\mathbf{v} \in \ell_2(\mathcal{J})$ we define the restriction

$$(\mathbf{v}|_\Lambda)_{\lambda\in\mathcal{J}} := \begin{cases} v_\lambda & \text{if} \quad \lambda \in \Lambda, \\ 0 & \text{if} \quad \lambda \notin \Lambda. \end{cases}$$

This suggests the following routine for approximating $\mathbf{F}(\mathbf{V})$ for any finitely supported vector $\mathbf{V}$:

EV $[\epsilon, \mathbf{F}, \mathbf{V}] \to \mathbf{w}_\epsilon$ *is defined by the following steps:*

(i) For $j = 0, 1, \ldots$, define the threshold $\eta_j = 2^{-j}$ and compute $\Lambda_j := \Lambda_{\eta_j}(\mathbf{V})$ by the strategy proposed above, see (7.14). Denote by $\mathcal{L}_j^-$ the set of outer leaves of Λ_j.

(ii) According to Proposition 7.2, we have $\|\mathbf{w} - \mathbf{w}|_{\Lambda_j}\|^2_{\ell_2(\mathcal{J})} \leq C\#(\mathcal{L}_j^-)\eta_j^2 := \epsilon_j^2$. Stop for the smallest j such that $\epsilon_j \leq \epsilon$ and define $\Lambda_\epsilon = \Lambda_j$.

(iii) Define the corresponding approximation $\mathbf{w}_\epsilon := \mathbf{w}|_{\Lambda_\epsilon}$.

Therefore the output $\mathbf{w}_\epsilon$ of **EV** $[\epsilon, \mathbf{F}, \mathbf{V}]$ satisfies

$$\|\mathbf{w}_\epsilon - \mathbf{F}(\mathbf{V})\|_{\ell_2(\mathcal{J})} \leq \epsilon. \tag{7.16}$$

A more detailed discussion of this scheme can be found in [9]. Here we mention only that the scheme indeed terminates after finitely many steps for any finitely supported input $\mathbf{V}$. In fact, finitely supported vectors belong to any space ${}_t\ell_\tau^w(\mathcal{J})$, $\tau < 2$. It will be seen below that $\mathbf{F}$ is stable on these spaces for a certain range of $\tau < 2$. Thus by definition $\mathcal{T}_\eta(\mathbf{F}(\mathbf{V}))$ grows at most like $\eta^{-\tau}$, so that the factor η_j^2 eventually forces the upper bounds $\#(\mathcal{L}_j^-)\eta_j^2 = \epsilon_j^2$ to become arbitrarily small.

This routine is still idealized at this point since nothing is said yet about the actual computation of the entries $w_\lambda = \mathbf{F}(\mathbf{V})_\lambda$. Once the significant coefficients have been identified though, one can resort to the techniques in [15] to compute them efficiently with sufficient accuracy. A more detailed treatment of this issue will be given elsewhere. In the sequel we shall work with the following

Assumption E: *The entries $w_\lambda = \mathbf{F}(\mathbf{v})_\lambda$ can be computed (with sufficient accuracy) on average at unit cost.*

We can now formulate concrete realizations of the scheme **SOLVE** that are suitable for nonlinear problems. We shall use **TCOARSE** as our version of **CCOARSE**. Moreover, for the semilinear elliptic problem (3.11) we can take

$$\mathbf{RES}_{\text{ell}}[\eta, \mathbf{A}, \mathbf{G}, \mathbf{v}] := \mathbf{APPLY}\,[\eta/3, \mathbf{A}, \mathbf{v}] + \mathbf{EV}\,[\eta/3, \mathbf{G}, \mathbf{v}] - \mathbf{RHS}\,[\eta/3, \mathbf{f}], \tag{7.17}$$

where **RHS** is defined here as in (5.2) but with **COARSE** replaced by **TCOARSE**.

For the general nonlinear problem (GNL), we shall now devise a residual approximation for the scheme from Section 3.3 with $\mathbf{B}_n := \mathbf{B}$. Suppose that $\|\mathbf{B}\|_{\ell_2(\mathcal{J})\to\ell_2(\mathcal{J})} \leq C_B$ and set

$$\mathbf{RES}_{\text{lc}}[\eta, \mathbf{B}, \mathbf{R}, \mathbf{v}] := \mathbf{APPLY}\,[\eta/2, \mathbf{B}, \mathbf{EV}\,[\eta/2C_B, \mathbf{R}, \mathbf{v}]]\,. \tag{7.18}$$

Of course, we have assumed here that the matrix $\mathbf{B}$ is compressible. In particular, for the stationary choice $\mathbf{B} = D\mathbf{R}(\mathbf{u}^0)$ this might be expected to be the case. We shall return to this issue later.

7.3. Complexity Estimates. We have now explained how to obtain concrete realizations of the scheme **SOLVE** in each of the three cases (L), (SL), (GNL). The remainder of this section will be devoted to giving a complexity analysis of these schemes. We begin with the following result from [9] (see Theorems 3.2 and 5.1).

THEOREM 7.3. *Assume that F satisfies Assumptions 1 and 2. Given any $\mathbf{V} \in ({}_t\ell^w_\tau(\mathcal{J}))^n$ for some $d/\gamma < \tau < 2$, then we have for the set $\Lambda_\eta(\mathbf{V})$ defined by (7.14)*

$$\#(\Lambda_\eta(\mathbf{V})) \lesssim \left(1 + \|\mathbf{V}\|^\tau_{({}_t\ell^w_\tau(\mathcal{J}))^n}\right)\eta^{-\tau}. \tag{7.19}$$

Therefore it follows that

$$\|\mathbf{F}(\mathbf{V})\|_{{}_t\ell^w_\tau(\mathcal{J})} \lesssim 1 + \|\mathbf{V}\|_{({}_t\ell^w_\tau(\mathcal{J}))^n}. \tag{7.20}$$

This can be used to derive the following estimate for the complexity of the routine **EV**, see [9].

PROPOSITION 7.3. *Assume that $\mathbf{V} \in ({}_t\ell^w_\tau(\mathcal{J}))^n$. Let $\mathbf{w}_\epsilon = \mathbf{EV}[\epsilon, \mathbf{F}, \mathbf{V}]$. Then one has for s related to τ by (5.13)*

$$\#\mathrm{supp}\,\mathbf{w}_\epsilon \lesssim \left(1 + \|\mathbf{V}\|^{1/s}_{({}_t\ell^w_\tau(\mathcal{J}))^n}\right)\epsilon^{-1/s}. \tag{7.21}$$

Moreover, (under Assumption E) the number of operations needed to compute $\mathbf{w}_\epsilon$ stays proportional to $\#\mathrm{supp}\,\mathbf{w}_\epsilon$.

Recall that in the case of semilinear equations, $\mathbf{R}$ involves a linear and a nonlinear operator as in (3.2) or (3.11). Also recall from Theorem 5.1 that when the wavelet representation $\mathbf{A}$ of the linear operator $\mathcal{A}$ defined by

$$\langle w, \mathcal{A}v\rangle = a(v,w), \quad v, w \in \mathcal{H} = H^t,$$

belongs to $\mathcal{C}_{s^*}$, then $\mathbf{A}$ is bounded on $\ell^w_\tau(\mathcal{J})$ with $s < s^*$ related to τ by (5.13). However, when also nonlinear operators are involved, Theorem 6.1 tells us that the spaces ${}_t\ell^w_\tau(\mathcal{J})$ should now play the role of $\ell^w_\tau(\mathcal{J})$. Thus, we shall prove in Proposition 7.6 below the boundedness of $\mathbf{A}$ with respect to this slightly stronger norm.

To prepare for Proposition 7.6, we make some remarks.

REMARK 7.4. *It has been shown in [11] that when $\mathcal{A}$ is a local operator, i.e, $\langle v, \mathcal{A}w\rangle = 0$ whenever $|\mathrm{supp}\, v \cap \mathrm{supp}\, w| = 0$ and when $\mathcal{A}$ is still bounded as a mapping from H^{t+a} to H^{-t+a} for $a \le m + d/2$, where m is less than or equal to the order of differentiabilty and vanishing moments of the wavelets ψ_λ, then one has*

$$|a(\psi_\lambda, \psi_\nu)| \lesssim 2^{-\sigma||\lambda|-|\nu||}, \quad \sigma = t + m + d/2. \tag{7.22}$$

Moreover, we know from Proposition 3.4 in [7] that in this case $\mathbf{A}$ belongs to $\mathcal{C}_{s^*}$ with

$$s^* = \frac{t+m}{d}. \tag{7.23}$$

Note that σ agrees with the value of γ in Assumption 2 or at least depends in an analogous way on the spatial dimension, the order of the operator and the order of the vanishing moments and the smoothness of the wavelets, see Theorem 7.2 (ii). Note that the condition $s < s^*$ means, in view of (5.13), that $\frac{1}{\tau} < s^* + \frac{1}{2} = \frac{t+r}{d} + \frac{1}{2}$ which is equivalent to

$$d/\sigma < \tau. \tag{7.24}$$

This is the same restriction on τ as given in Theorem 7.3.

REMARK 7.5. *One can show that for piecewise polynomial wavelets s^* can be chosen larger than in (7.23) resulting in a weaker restriction on τ than (7.24), see [2].* By definition one has ${}_t\ell^w_\tau(\mathcal{J}) \subset \ell^w_\tau(\mathcal{J})$. We shall need the following refinement of Theorem 5.1.

PROPOSITION 7.6. *Under the assumptions from Remark 7.4 on the linear part $\mathcal{A}$ let*

$$\sigma = m + t + d/2. \tag{7.25}$$

Then one has for $\tau > d/\sigma$

$$\|\mathbf{A}\mathbf{v}\|_{{}_t\ell^w_\tau(\mathcal{J})} \lesssim \|\mathbf{v}\|_{{}_t\ell^w_\tau(\mathcal{J})}, \quad \mathbf{v} \in {}_t\ell^w_\tau(\mathcal{J}), \tag{7.26}$$

that is $\mathbf{A}$ maps ${}_t\ell^w_\tau(\mathcal{J})$ boundedly into itself.

Proof: By assumption (3.9), $\mathcal{A}$ is a topological isomorphism from $\mathcal{H}$ onto $\mathcal{H}'$ and obviously satisfies Assumption 1. The validity of (7.4) for $\mathbf{F}(\mathbf{v}) := \mathbf{A}\mathbf{v}$ is an immediate consequence of the locality of $\mathcal{A}$. We need to show that Assumption 2 holds for all $\gamma' < \sigma$, defined by (7.25). To this end, note that (7.22) provides

$$\begin{aligned}
|(\mathbf{A}\mathbf{v})_\lambda| \le & \sum_{\substack{|\nu| \le |\lambda|,\\ S_\nu \cap S_\lambda \neq \emptyset}} |v_\nu| 2^{-\sigma(|\lambda|-|\nu|)} + \sum_{\substack{|\nu| > |\lambda|,\\ S_\nu \cap S_\lambda \neq \emptyset}} |v_\nu| 2^{-\sigma(|\lambda|-|\nu|)} 2^{-2\sigma(|\nu|-|\lambda|)} \\
\le & \sup_{\substack{|\nu| \le |\lambda|,\\ S_\nu \cap S_\lambda \neq \emptyset}} 2^{-\gamma'(|\lambda|-|\nu|)} |v_\nu| \sum_{\substack{|\nu| \le |\lambda|,\\ S_\nu \cap S_\lambda \neq \emptyset}} 2^{-(\sigma-\gamma')(|\lambda|-|\nu|)} \\
& + \sup_{\substack{|\nu| > |\lambda|,\\ S_\nu \cap S_\lambda \neq \emptyset}} |v_\nu| 2^{-\sigma(|\lambda|-|\nu|)} \sum_{\substack{|\nu| > |\lambda|,\\ S_\nu \cap S_\lambda \neq \emptyset}} 2^{-2\sigma(|\nu|-|\lambda|)}.
\end{aligned}$$

We now check that both sums appearing on the right hand side are bounded independently of λ provided that $\gamma' < \sigma$. Indeed, in the first sum for any $k > |\lambda|$, there are a bounded number C_0 of indices ν with $|\nu| = k$ such that $S_\nu \cap S_\lambda \neq \emptyset$. Hence this sum is bounded by $C_0 \sum_{j=0}^\infty 2^{-j(\sigma-\gamma')}$. For the second sum, note that for $|\nu| = |\lambda| + k$, there are at most $C_0 2^{kd}$ indices ν for which $S_\nu \cap S_\lambda \neq \emptyset$. Since, by (7.25), $2\sigma > d$ this sum can also be bounded by a geometric series. We have thus verified Assumption 2 for all $\gamma' < \sigma$. The assertion now follows from Theorem 7.3 and the restriction on τ given in Theorem 7.3. □

We shall make use of the following consequence of Proposition 7.6.

COROLLARY 7.7. *Under the same assumptions as in Proposition 7.6 let $\frac{1}{\tau} = s + \frac{1}{2}$, $s < s^*$. Then $\mathbf{w}_\eta = \mathbf{APPLY}\,[\eta, \mathbf{A}, \mathbf{v}]$ satisfies:*

(i) $\#\text{flops}, \#\text{supp}\,\mathbf{w}_\eta \lesssim \|\mathbf{v}\|^{1/s}_{{}_t\ell^w_\tau(\mathcal{J})} \eta^{-1/s}$;

(ii) $\|\mathbf{w}_\eta\|_{{}_t\ell^w_\tau(\mathcal{J})} \lesssim \|\mathbf{v}\|_{{}_t\ell^w_\tau(\mathcal{J})}$.

Proof: (i) follows directly from Theorem 5.1 (ii) and the fact that $\|\cdot\|_{\ell^w_\tau(\mathcal{J})} \lesssim \|\cdot\|_{{}_t\ell^w_\tau(\mathcal{J})}$. As for (ii), note that, since the chunks $(\mathbf{v}_{[j]} - \mathbf{v}_{[j-1]})$ have disjoint supports, for each j the vector $\mathbf{w}_j$, defined by (5.7), can be interpreted as $\mathbf{w}_j = \mathbf{C}^{(j)}\mathbf{v}$ where the matrix $\mathbf{C}^{(j)}$ is a compressed version of $\mathbf{A}$ defined as follows. All columns with indices outside $\text{supp}\,\mathbf{v}_{[j]}$ are zero. The columns of $\mathbf{C}^{(j)}$ whose indices belong to $\text{supp}\,(\mathbf{v}_{[k]} - \mathbf{v}_{[k-1]})$, $k \le j$, agree with the corresponding columns in the matrix $\mathbf{A}_{j-k}$. Therefore, since the $\mathbf{A}_j$ are derived from $\mathbf{A}$ by replacing certain entries by zero, we conclude that $\mathbf{C}^{(j)}$ is obtained by replacing certain entries in $\mathbf{A}$ by zero. Thus the $\mathbf{C}^{(j)}$ still statisfy (7.22) uniformly in j. Assumptions 1 and 2 remain also valid. Thus Proposition 7.6 can be applied to $\mathbf{C}^{(j)}$ with constants independent of j which finishes

the proof. □

We have now collected all the ingredients needed to confirm τ^*-sparsity of the residual approximations defined before. We start with (7.17).

COROLLARY 7.8. *Let γ be the parameter given in Theorem 7.2 for the respective nonlinear mapping F. Suppose that σ, defined by (7.25) satisfies $\sigma \geq \gamma$. Then* $\mathbf{RES}_{\mathrm{ell}}$, *defined by (7.17) is τ^* sparse with $\tau^* := d/\gamma$.*

Proof: We have to verify the validity of (6.16) for $\tau > \tau^*$. If $\mathbf{u} \in {}_t\ell^w_\tau(\mathcal{J})$ for some $\tau > \tau^*$, then Proposition 7.6 implies that $\mathbf{f} \in {}_t\ell^w_\tau(\mathcal{J})$ and $\|\mathbf{f}\|_{{}_t\ell^w_\tau(\mathcal{J})} \lesssim \|\mathbf{u}\|_{{}_t\ell^w_\tau(\mathcal{J})}$. Hence, since **TCOARSE** satisfies completely analogous properties with regard to ${}_t\ell^w_\tau(\mathcal{J})$ as **COARSE** with respect to $\ell^w_\tau(\mathcal{J})$ (see [7]), we conclude that the output $\mathbf{f}_\eta$ of $\mathbf{RHS}\,[\eta, \mathbf{f}]$ satisfies

$$\#\mathrm{supp}\,\mathbf{f}_\eta \lesssim \eta^{-1/s}\|\mathbf{u}\|^{1/s}_{{}_t\ell^w_\tau(\mathcal{J})}, \quad \|\mathbf{f}_\eta\|_{{}_t\ell^w_\tau(\mathcal{J})} \lesssim \|\mathbf{u}\|_{{}_t\ell^w_\tau(\mathcal{J})}. \tag{7.27}$$

Furthermore, Corollary 7.7 says that the output of **APPLY** remains bounded in ${}_t\ell^w_\tau(\mathcal{J})$, while Proposition 7.3 ensures that the same is true for the output of **EV**. Hence, by (7.17), one has

$$\|\mathbf{w}_\eta\|_{{}_t\ell^w_\tau(\mathcal{J})} \lesssim \left(\|\mathbf{v}\|_{{}_t\ell^w_\tau(\mathcal{J})} + \|\mathbf{u}\|_{{}_t\ell^w_\tau(\mathcal{J})} + 1\right), \quad \mathbf{w}_\eta := \mathbf{RES}_{\mathrm{ell}}[\eta, \mathbf{A}, \mathbf{G}, \mathbf{v}], \tag{7.28}$$

which is the second estimate in (6.16). The first estimate in (6.16) follows also from (7.27), Theorem 7.3, and Corollary 7.7 (i). This completes the proof. □

Combining Corollary 7.8 with Theorem 6.1, yields the first main result of this paper.

THEOREM 7.4. *Under the same assumptions as in Corollary 7.8 suppose that the solution $u = \sum_{\lambda\in\mathcal{J}} u_\lambda\psi_\lambda$ satisfies $\mathbf{u} \in {}_t\ell^w_\tau(\mathcal{J})$ for some $\tau > \tau^* := d/\gamma$. Then the approximate solution $u(\epsilon) = \sum_{\lambda\in\Lambda(\epsilon)} \bar{u}(\epsilon)_\lambda\psi_\lambda$ produced by* **SOLVE** *(with the initialization for the semilinear problem) after finitely many steps satisfies*

$$\|u - u(\epsilon)\|_{\mathcal{H}} \leq C_1\epsilon.$$

Moreover

$$\#(\mathrm{flops}),\ \#(\Lambda(\epsilon)) \lesssim \epsilon^{-1/s}\|\mathbf{u}\|^{1/s}_{{}_t\ell^w_\tau(\mathcal{J})}, \quad \|\bar{\mathbf{u}}(\epsilon)\|_{{}_t\ell^w_\tau(\mathcal{J})} \lesssim \|\mathbf{u}\|_{{}_t\ell^w_\tau(\mathcal{J})}, \tag{7.29}$$

where the constants depend only on τ when $\tau \to \tau^*$.

Let us briefly discuss now the locally convergent scheme from Section 2.3. We shall assume that for the general problem (2.2) the nonlinear map in (2.1) satisfies Assumptions 1 and 2. Moreover, we assume that a sufficiently good initial guess $\mathbf{u}^0$ is given, so that the error reduction (3.25) holds, and that **SOLVE** is initialized accordingly, see Section 5.

COROLLARY 7.9. *Let γ be the parameter given in Theorem 7.2 for the respective nonlinear mapping F. Moreover, assume that the matrix $\mathbf{B} = \mathbf{B}_n$ appearing in (2.10) satisfies decay estimates like (7.22) for some $\sigma \geq \gamma$. Then the scheme* $\mathbf{RES}_{\mathrm{lc}}$ *defined by (7.18) is τ^*-sparse for $\tau^* := d/\gamma$.*

Proof: The assertion follows again from Propositions 7.6 and 7.3. □

COROLLARY 7.10. *Under the assumptions of Corollary 7.9 the assertion of Theorem 7.4 remains valid for the locally convergent scheme based on* $\mathbf{RES}_{\mathrm{lc}}$.

We end this Section with analyzing the compressibility of the special choice $\mathbf{B} = D\mathbf{R}(\mathbf{u}^0)^T$. We consider $\mathcal{H} = H^t$ and only nonlinear maps of a single argument $n = 1$ and the subcritical case $t < d/2$, $p < p^*$, recall (7.7). Recall from (2.11) that the entries of $D\mathbf{R}(\mathbf{u}^0)^T = D\mathbf{R}(\mathbf{u}^0)$ have the form $w_{\lambda,\nu} := \langle\psi_\lambda, \psi_\nu R'(u^0)\rangle$. Since in

view of the $\mathcal{H} = H^t$-normalization of the ψ_λ, one has $\|\psi_\lambda\|_{L_\infty} \sim 2^{(\frac{d}{2}-t)|\lambda|}$ the same arguments as used in the proof of Theorem 4.2 in [9] yield

$$|w_{\lambda,\nu}| \lesssim \|\psi_\lambda\|_{L_\infty} \inf_{P\in\mathbb{P}_r} \|P - \psi_\nu R'(u^0)\|_{L_1(S_\lambda)} \lesssim 2^{-(r+\frac{d}{2}+t)|\lambda|} |\psi_\nu R'(u^0)|_{W^r(L_\infty(S_\lambda))},$$

where we assume without loss of generality that $|\lambda| \geq |\nu|$. Moreover, we obtain

$$|\psi_\nu R'(u^0)|_{W^r(L_\infty(S_\lambda)} \lesssim \max_{l\leq r} |\psi_\nu|_{W^{r-l}(L_\infty(S_\lambda))} |R'(u^0)|_{W^l(L_\infty(S_\lambda))}.$$

Abbreviating as before $\sigma = r + \frac{d}{2} + t$, we can estimate the first factor by

$$2^{(\frac{d}{2}-t)|\nu|} 2^{(r-l)|\nu|} = 2^{\sigma|\nu|} 2^{-(2t+l)|\nu|}, \tag{7.30}$$

while the second factor can be estimated along the lines of the proof of Theorem 4.2 in [9] as

$$\begin{aligned} |R'(u^0)|_{W^l(L_\infty(S_\lambda))} &\lesssim \max_{k=1,\dots,l} \|\mathbf{u}^0\|_{\ell_2(\mathcal{J})}^{m+k-1} \\ &\quad \times \sup_{\mu: S_\mu \cap S_\lambda \cap S_\nu \neq \emptyset} |u^0_\mu| 2^{(l+(p-1)\epsilon)|\mu|} 2^{(p-1)(\frac{d}{2}-t)|\mu|}, \end{aligned} \tag{7.31}$$

where $\epsilon > 0$, $j_i \in \mathbb{N}$ and

$$m := \begin{cases} (p-l-1)_+ & \text{if} \quad \|\mathbf{u}^0\|_{\ell_2(\mathcal{J})} \geq 1; \\ 0 & \text{if} \quad \|\mathbf{u}^0\|_{\ell_2(\mathcal{J})} < 1. \end{cases}$$

As in [9] we can choose ϵ so that

$$2^{(l+(p-1)\epsilon)|\mu|} 2^{(p-1)(\frac{d}{2}-t)|\mu|} \lesssim 2^{(l+\frac{d}{2}+t)|\mu|} 2^{-\epsilon|\mu|} 2^{-(\frac{d}{2}-t)|\mu|} = 2^{-\epsilon|\mu|} 2^{(l+2t)|\mu|}.$$

Thus, combining (7.30) and (7.31), we obtain

$$|w_{\lambda,\nu}| \lesssim C(\|\mathbf{u}^0\|_{\ell_2(\mathcal{J})}) 2^{-\sigma||\lambda|-|\nu||} \sup_{\mu: S_\mu \cap S_\lambda \cap S_\nu \neq \emptyset} |u^0_\mu| 2^{-\epsilon|\mu|} 2^{(r+2t)(|\mu|-|\nu|)}. \tag{7.32}$$

Note that the first factor $C(\|\mathbf{u}^0\|_{\ell_2(\mathcal{J})}) 2^{-\sigma||\lambda|-|\nu||}$ represents the same scalewise decay of the entries as in the matrix $\mathbf{A}$ in (7.22). This ensures that $D\mathbf{R}(\mathbf{u}^0)$ belongs to $\mathcal{C}_{s^*}$ with s^* given by (7.23). However, the entries are weighted by additional u^0-dependent factors, that could, in principle, become rather large when the finite expansion u^0 contains basis functions from high scales overlapping S_λ. Nevertheless, these factors depend only on u^0 (and hence on the accuracy δ of the initial guess) but not on the accuracy by which $D\mathbf{R}(\mathbf{u}^0)$ is applied through the scheme **APPLY**. Therefore, in principle, one obtains asymptotically optimal complexity, however, with possibly poor quantitative behavior.

8. Newton's Method. In concrete cases the error reduction ρ obtained in (3.25) or (3.19) may be so close to one that the number K of necessary updates in the perturbed scheme **SOLVE** may become fairly large. So in spite of its asymptotic optimality, the quantitative performance may be poor. We shall therefore address Newton's method, corresponding to $\mathbf{B}_n = (D\mathbf{R}(\mathbf{u}^n))^{-1}$ in (2.10), as an example where the ideal scheme permits a faster error decay. The adaptive realization of Newton's method, applied to the infinite dimensional problem (2.2) or better yet (2.9), does not quite fit into the format of **SOLVE**, explaining its separate treatment in this section.

Note that, for $\mathbf{B}_n = (D\mathbf{R}(\mathbf{u}^n))^{-1}$, (2.10) can be reformulated as follows. Given an initial guess $\mathbf{u}^n$ the next iterate $\mathbf{u}^{n+1}$ is determined by solving

$$D\mathbf{R}(\mathbf{u}^n)\mathbf{w}^n = -\mathbf{R}(\mathbf{u}^n) \tag{8.1}$$

and setting

$$\mathbf{u}^{n+1} := \mathbf{u}^n + \mathbf{w}^n. \tag{8.2}$$

We are not interested here in the weakest assumptions under which the iterative scheme (8.2) converges to a locally unique solution. We are instead content here with the following setting: Recall that the mapping R in the variational problem (2.2) has the form

$$R(v) = F(v) - f, \tag{8.3}$$

where throughout this section we shall confine the discussion again to nonlinear maps F of a single argument satisfying the growth condition (7.6) for some $n^* \geq 1$. (Of course, F can have a linear part as in (3.11).) Therefore we have, in particular, that R and F have the same Frechét derivative $DR(v) = DF(v)$. Moreover, we assume that for some open set $\mathcal{U} \subset \mathcal{H}$ one has

(N1) The Frechét derivative $DR(v) : w \mapsto DR(v)w$ is an isomorphism from $\mathcal{H}$ to $\mathcal{H}'$ and for all $v, y \in \mathcal{U}$ and some $\omega > 0$

$$\|(DR(v))^{-1}(DR(v+sy) - DR(v))y\|_{\mathcal{H}} \leq s\omega\|y\|_{\mathcal{H}} \tag{8.4}$$

holds for all $s \in [0,1]$, $v \in \mathcal{U}$ such that $v + y \in \mathcal{U}$.

(N2) There exists a solution $u \in \mathcal{U}$ and an initial guess u^0 in $\mathcal{U}$ such that

$$\|u - u^0\|_{\mathcal{H}} \leq \delta < 2/\omega \quad \text{and} \quad B_\delta(u) \subseteq \mathcal{U} \tag{8.5}$$

with ω from (N1).

Let us check that (N1) holds, for instance, in the semilinear case (SL) when F defined by (3.11) with monotone (scalar valued) G and when F satisfies (7.6) with $n^* \geq 2$. (Note that Lipschitz continuity of G' would suffice as well). To see this, recall that in this case one has $\langle z, DF(v)w\rangle = a(z,w) + \langle z, G'(v)w\rangle$, where G' is the (pointwise) derivative of G. Hence, we obtain

$$\|DR(v)w\|_{\mathcal{H}'} = \sup_{z\in\mathcal{H}} \frac{\langle z, DR(v)w\rangle}{\|z\|_{\mathcal{H}}} = \sup_{z\in\mathcal{H}} \frac{a(z,w) + \langle z, G'(v)w\rangle}{\|z\|_{\mathcal{H}}} \geq \frac{a(w,w)}{\|w\|_{\mathcal{H}}} \geq c\|w\|_{\mathcal{H}},$$

where c is the ellipticity constant from(3.9). Likewise, one has

$$\|DR(v)w\|_{\mathcal{H}'} \leq C\|w\|_{\mathcal{H}} + \|G'(v)w\|_{\mathcal{H}'}.$$

When F satisfies (7.6) then $Q(v,w) := G'(v)w$ satisfies (7.8) which, by Theorem 7.1, means that $\|G'(v)w\|_{\mathcal{H}'} \leq C(\|v\|_{\mathcal{H}})\|w\|_{\mathcal{H}}$. This shows that $DR(v)$ is an isomorphism from $\mathcal{H}$ to $\mathcal{H}'$. The verification of (8.4) is similar now using that G satisfies (7.6) for $n^* \geq 2$.

Given the validity of (N1) and (N2), standard arguments can be employed to prove that all iterates

$$u^{n+1} = u^n - DR(u^n)^{-1}R(u^n), \tag{8.6}$$

arising from Newton's method formulated in $\mathcal{H}$, remain in $\mathcal{U}$ and satisfy

$$\|u - u^n\|_{\mathcal{H}} < \delta \quad \text{for} \quad n \in \mathbb{N} \quad \text{and} \quad \lim_{n\to\infty} \|u - u^n\|_{\mathcal{H}} = 0. \tag{8.7}$$

In fact, one has superlinear convergence

$$\|u - u^{n+1}\|_{\mathcal{H}} \leq \frac{\omega}{2}\|u - u^n\|_{\mathcal{H}}^2, \quad n = 0, 1, 2, \ldots, \tag{8.8}$$

see e.g. [17, 18]. Finally, note that, by (2.5), the corresponding iterates in wavelet coordinates satisfy

$$\|\mathbf{u}-\mathbf{u}^{n+1}\|_{\ell_2(\mathcal{J})} \le \tilde{\omega}\|\mathbf{u}-\mathbf{u}^n\|^2_{\ell_2(\mathcal{J})}, \quad n\in\mathbb{N}_0, \quad \tilde{\omega} := \frac{C_1^2\omega}{2c_1}. \tag{8.9}$$

The purpose of this section is to show how the approximate solution of the linear problem (8.1) can be performed again by an iterative scheme along the lines of [8]. By our assumption (N1), we know that $DR(z)$ is an isomorphism from $\mathcal{H}$ to $\mathcal{H}'$ provided that $z\in\mathcal{U}$. Thus

$$c_2\|\mathbf{v}\|_{\ell_2(\mathcal{J})} \le \|D\mathbf{R}(\mathbf{z})\mathbf{v}\|_{\ell_2(\mathcal{J})} \le C_2\|\mathbf{v}\|_{\ell_2(\mathcal{J})}, \quad \mathbf{v}\in\ell_2(\mathcal{J}), \tag{8.10}$$

holds for all $\mathbf{z}$ such that $z=\sum_{\lambda\in\mathcal{J}} z_\lambda\psi_\lambda\in\mathcal{U}$. Given this mapping property (8.10), the adaptive scheme from [8] can actually be applied under fairly general assumptions on the linear isomorphism. For the sake of simplicity we shall assume that $D\mathbf{R}(\mathbf{z})$ is symmetric positive definite. This is for example true in the case (SL) when G is monotone. From positive definiteness and (8.10), we can find a damping parameter $\alpha>0$ such that for some $\rho<1$

$$\|\mathbf{I}-\alpha D\mathbf{R}(\mathbf{v})\| \le \rho, \quad \forall\ \ v\in\mathcal{U}. \tag{8.11}$$

The heart of the envisaged adaptive Newton scheme will be to solve the linear problem (8.1) approximately with the aid of a variant, which will be called $\mathbf{SOLVE}_{\mathrm{N}}$, of the scheme $\mathbf{SOLVE}_{\mathrm{lin}}$ discussed in Section 3. Before we proceed to describing the ingredients of $\mathbf{SOLVE}_{\mathrm{N}}$ let us point out two issues to be addressed when designing these ingredients and analyzing their complexity.

(a) The first point concerns the application of the Jacobian. Approximating at each stage the Jacobian $D\mathbf{R}(\mathbf{u}^n)$ in order to use the **APPLY** scheme based on (5.7) might be computationally very expensive. (b) The second point concerns the complexity of approximately solving the linear problem (8.1). Recall from Theorem 6.1 that the logic of complexity estimates is to infer from a certain compressibility (or regularity) of the solution a corresponding convergence rate of the adaptive scheme. In the context of the Newton iteration such a property will be assumed about the solution $\mathbf{u}$ of the original nonlinear problem (2.9) which, however, does not necessarily imply the same property for the solutions of the subproblems (8.1). So it is initially not clear how to derive complexity estimates for the resolution of these subproblems. It will be seen though that the solutions to these subproblems become increasingly closer to elements having the necessary properties, a fact that, as it turns out, can be exploited as long as the subproblems are not solved too accurately. In particular, the question then arises whether the quadratic convergence of the Newton scheme can be preserved.

We now turn to collecting the ingredients of the adaptive Newton scheme. First of all, the coarsening will be again done by **TCOARSE** even though the problem is linear. More importantly, the **RES** scheme will be of the form $\mathbf{RES}_{\mathrm{lin}}$ from (5.15) but with different schemes playing the roles of **APPLY** and **RHS**.

In view of the above issue (a), we shall pursue here the following approach. Recall from (2.11) that for any $\mathbf{v},\mathbf{z}\in\ell_2(\mathcal{J})$ and corresponding $v=\sum_{\lambda\in\mathcal{J}} v_\lambda\psi_\lambda$, $z=\sum_{\lambda\in\mathcal{J}} z_\lambda\psi_\lambda\in\mathcal{H}$,

$$D\mathbf{R}(z)\mathbf{v} = (\langle\psi_\lambda, DR(z)v\rangle : \lambda\in\mathcal{J}), \tag{8.12}$$

where $DR(z)$ is the Frechét derivative of R at z. This suggests employing the scheme **EV** with $Q(z,v):=DR(z)v=DF(z)v$. Note that, under the above assumptions on F, we have

$$Q:(z,v)\mapsto Q(z,v), \quad Q:\mathcal{H}\times\mathcal{H}\mapsto\mathcal{H}'. \tag{8.13}$$

In fact, since F satisfies (7.6) for some $n^* \geq 1$, one readily verifies that

$$|D^\beta Q(x_1,x_2)| \lesssim (1+|x_1|)^{p-1-\beta_1}(1+|x_2|)^{(1-\beta_2)_+}, \quad 0 \leq \beta_1 \leq n^*-1,\ 0 \leq \beta_2.$$

We shall make use of the following consequences of this observation.

REMARK 8.1. *The mapping Q satisfies (7.8) with $p_1 := p-1, p_2 := 1$, $\alpha_1 = \alpha_2 = 0$. Moreover, (7.9) and (8.13) are seen to be implied by the condition (7.6) on F. Hence Theorems 7.1 and 7.2 ensure that Q satisfies Assumptions 1 and 2 in Section 7.1.*

This suggests the following routine:

APPLY$_{\rm N}[\eta, D\mathbf{R}(\mathbf{z}), \mathbf{v}] \to \mathbf{w}_\eta$: *determines for any tolerance $\eta > 0$ and any finitely supported input vectors $\mathbf{v}$ and $\mathbf{z}$ a finitely supported output vector $\mathbf{w}_\eta$ such that*

$$\|D\mathbf{R}(\mathbf{z})\mathbf{v} - \mathbf{w}_\eta\|_{\ell_2(\mathcal{J})} \leq \eta, \tag{8.14}$$

through

$$\mathbf{APPLY}_{\rm N}[\eta, D\mathbf{R}(\mathbf{z}), \mathbf{v}] := \mathbf{EV}[\eta, \mathbf{Q}, (\mathbf{z}, \mathbf{v})], \tag{8.15}$$

where the routine **EV** *was introduced in Section 7.2*

It remains to specify the routine **RHS**. Here it is issue (b) that calls for some further preparations. The main point is that if the current right hand sides in (8.1) are not approximated too accurately then one actually approximates a nearby right hand side of a problem whose solution is known to be sufficiently sparse and thus can be approximated efficiently by a linear version of **SOLVE**.

REMARK 8.2. *Suppose that $R \in C^2$ and let for any $z \in \mathcal{U}$ and $z = \sum_{\lambda\in\mathcal{J}} z_\lambda \psi_\lambda$*

$$\mathbf{G}(\mathbf{z}) := D\mathbf{R}(\mathbf{z})(\mathbf{u}-\mathbf{z}), \tag{8.16}$$

where $\mathbf{u}$ is the exact solution of (2.9). Then there exists a constant $\hat{C}$ such that

$$\|\mathbf{G}(\mathbf{z}) + \mathbf{R}(\mathbf{z})\|_{\ell_2(\mathcal{J})} \leq \hat{C}\|\mathbf{u}-\mathbf{z}\|^2_{\ell_2(\mathcal{J})}. \tag{8.17}$$

Proof: One has

$$-\mathbf{R}(\mathbf{z}) = \mathbf{R}(\mathbf{u}) - \mathbf{R}(\mathbf{z}) = D\mathbf{R}(\mathbf{z})(\mathbf{u}-\mathbf{z}) + \mathcal{O}\left(\|\mathbf{u}-\mathbf{z}\|^2_{\ell_2(\mathcal{J})}\right),$$

which confirms the claim. □

We shall employ the following routine in which $\hat{C}$ is the constant of Remark 8.2:

RHS$_{\rm N}[\eta, \mathbf{R}, \mathbf{z}] \to \mathbf{r}_\eta(\mathbf{z})$ *which is defined for any finitely supported $\mathbf{z}$ with $z \in \mathcal{U}$ such that $\|\mathbf{u}-\mathbf{z}\|_{\ell_2(\mathcal{J})} \leq \xi$ and for any $\eta/2 > \hat{C}\xi^2$ by*

$$\mathbf{RHS}_{\rm N}[\eta, \mathbf{R}, \mathbf{z}] := -\left(\mathbf{EV}[\frac{\eta}{2} - \hat{C}\xi^2, \mathbf{F}, \mathbf{z}] - \mathbf{RHS}[\eta/2, \mathbf{f}]\right), \tag{8.18}$$

where **RHS** *is defined by (5.2) but with* **TCOARSE** *used as* **CCOARSE**.

The role of the above conditions on $\mathbf{z}$ and η will become clear later.

We are now prepared to describe the version of $\mathbf{SOLVE}_{\rm lin}$ to be used for the approximate solution of the Newton systems (8.1) as follows:

SOLVE$_{\rm N}[\eta, \mathbf{R}, \mathbf{z}] \to \mathbf{w}_\eta$ *determines for a given $\mathbf{z} \in \ell_2(\mathcal{J})$, such that $z \in \mathcal{U}$, an approximate solution $\mathbf{w}_\eta$ of the system $D\mathbf{R}(\mathbf{z})\mathbf{w} = -\mathbf{R}(\mathbf{z})$ satisfying*

$$\|\mathbf{w} - \mathbf{w}_\eta\|_{\ell_2(\mathcal{J})} \leq \eta, \tag{8.19}$$

by invoking $\mathbf{SOLVE}_{\rm lin}[\eta, D\mathbf{R}(\mathbf{z}), -\mathbf{R}(\mathbf{z})] \to \mathbf{w}_\eta$, *where, under the above assumptions on* $\mathbf{z}$ *and* η, *in step (ii) of* **SOLVE** *the residual approximation*

$$\text{(8.20)}\quad \mathbf{RES}_{\rm N}[\eta, \mathbf{R}, \mathbf{z}, \mathbf{v}] := \alpha\left(\mathbf{APPLY}_{\rm N}\,[\frac{\eta}{2\alpha}, D\mathbf{R}(\mathbf{z}), \mathbf{v}] - \mathbf{RHS}_{\rm N}\,[\frac{\eta}{2\alpha}, -\mathbf{R}(\mathbf{z})]\right),$$

and in step (iii) **TCOARSE** *is used.*

Note that, in view of (8.3), the evaluation of $\mathbf{R}$ also requires the approximation of the data $\mathbf{f}$ as stated explicitly in (8.18). From Theorem 7.3, Proposition 7.3 and Proposition 6.5 we infer, as in Remark 5.2, that $\mathbf{u} \in {}_t\ell^w_\tau(\mathcal{J})$ implies $\mathbf{f} \in {}_t\ell^w_\tau(\mathcal{J})$, and its η-accurate tree approximation satisfies estimates of the form $\#\mathrm{supp}\,\mathbf{f}_\eta \lesssim \eta^{-1/s}\|\mathbf{u}\|^{1/s}_{{}_t\ell^w_\tau(\mathcal{J})}$, $\|\mathbf{f}_\eta\|_{{}_t\ell^w_\tau(\mathcal{J})} \lesssim \|\mathbf{u}\|_{{}_t\ell^w_\tau(\mathcal{J})}$.

Moreover, by Remark 8.1 we can apply Theorem 7.3 and Proposition 7.3 to conclude that the output $\mathbf{w}_\eta$ of $\mathbf{APPLY}_{\rm N}\,[\eta, D\mathbf{R}(\mathbf{z}), \mathbf{v}]$ satisfies

$$\text{(8.21)}\quad \begin{aligned} \#\mathrm{supp}\,\mathbf{w}_\eta &\lesssim \eta^{-1/s}\left(1 + \|\mathbf{z}\|^{1/s}_{{}_t\ell^w_\tau(\mathcal{J})} + \|\mathbf{v}\|^{1/s}_{{}_t\ell^w_\tau(\mathcal{J})} + \|\mathbf{u}\|^{1/s}_{{}_t\ell^w_\tau(\mathcal{J})}\right), \\ \|\mathbf{w}_\eta\|_{{}_t\ell^w_\tau(\mathcal{J})} &\lesssim 1 + \|\mathbf{z}\|_{{}_t\ell^w_\tau(\mathcal{J})} + \|\mathbf{v}\|_{{}_t\ell^w_\tau(\mathcal{J})} + \|\mathbf{u}\|_{{}_t\ell^w_\tau(\mathcal{J})}. \end{aligned}$$

Likewise the output $\mathbf{r}_\eta(\mathbf{z})$ of $\mathbf{RHS}_{\rm N}\,[\eta, \mathbf{R}, \mathbf{z}]$ satisfies

$$\text{(8.22)}\quad \#\mathrm{supp}\,\mathbf{r}_\eta(\mathbf{z}) \lesssim \eta^{-1/s}\left(1 + \|\mathbf{z}\|^{1/s}_{{}_t\ell^w_\tau(\mathcal{J})}\right), \quad \|\mathbf{r}_\eta(\mathbf{z})\|_{{}_t\ell^w_\tau(\mathcal{J})} \lesssim 1 + \|\mathbf{z}\|_{{}_t\ell^w_\tau(\mathcal{J})}.$$

Recalling from (6.16) the definition of τ^*-sparseness, we can infer from (8.21) the following consequence.

REMARK 8.3. *Let* $\tau^* := d/\gamma$ *where* γ *is the parameter associated with* F *by Theorem 7.2. Then the scheme* $\mathbf{RES}_{\rm N}$, *defined by (8.20), is* τ^**-sparse whenever* $\|\mathbf{z}\|_{{}_t\ell^w_\tau(\mathcal{J})} \lesssim 1$.

We can now formulate an adaptive Newton iteration as follows.

$\mathbf{NEWTON}\,[\epsilon, R, \bar{\mathbf{u}}^0] \to \bar{\mathbf{u}}(\epsilon)$ *determines for any finitely supported initial guess* $\bar{\mathbf{u}}^0$ *whose corresponding expansion* u_0 *satisfies (8.5), an approximate solution* $\bar{\mathbf{u}}(\epsilon)$ *satsifying*

$$\text{(8.23)}\qquad \|\mathbf{u} - \bar{\mathbf{u}}(\epsilon)\|_{\ell_2(\mathcal{J})} \le \epsilon,$$

by the following steps:

(i) Set $\epsilon_0 := c_1^{-1}\delta$, $j = 0$.

(ii) If $\epsilon_j \le \epsilon$ *stop and output* $\bar{\mathbf{u}}(\epsilon) := \bar{\mathbf{u}}^j$. *Otherwise choose some* $\eta_j > 0$ *(see (8.26) below) and perform*

$$\mathbf{SOLVE}_{\rm N}\,[\eta_j, \mathbf{R}, \bar{\mathbf{u}}^j] \to \bar{\mathbf{w}}^j.$$

(iii) Let (see (8.9))

$$\hat{\mathbf{u}} := \bar{\mathbf{u}}^j + \bar{\mathbf{w}}^j, \quad \hat{\eta}_j := \tilde{\omega}\epsilon_j^2 + \eta_j, \quad \bar{\mathbf{u}}^{j+1} := \mathbf{TCOARSE}\,[3C^*\hat{\eta}_j, \hat{\mathbf{u}}],$$

(where C^* *is the constant from Section 6.2) and set* $\epsilon_{j+1} := (1 + 3C^*)\hat{\eta}_j$, $j+1 \to j$ *and go to (ii).*

The choice of the dynamic tolerance η_j in step (ii) is yet to be specified. The first requirement is to keep the iterates $\bar{\mathbf{u}}^j$ in the right neighborhood of the solution, which means that the corresponding expansions $\bar{u}^j$ lie in $B_\delta(u)$. For this we shall use the following lemma.

LEMMA 8.1. *Fix a positive number* $\beta < 1$ *and assume that* $\delta > 0$ *is chosen sufficiently small to ensure that, in addition to (8.5),*

$$\text{(8.24)}\qquad \delta < \frac{c_1^3}{(1 + 3C^*)C_1^3\omega},$$

and

$$\frac{(1+3C^*)\tilde{\omega}\delta}{c_1} < \beta. \tag{8.25}$$

Then the condition $\eta_j \leq \eta_0 < \delta/(2(1+3C^*)C_1)$ *implies that* $\bar{u}_j \in B_\delta(u)$ *for all subsequent approximate iterates. Moreover, if*

$$\eta_j \leq \frac{\epsilon_j\,(\beta - (1+3C^*)\tilde{\omega}\epsilon_j)}{1+3C^*}, \quad j = 0, 1, \ldots, \tag{8.26}$$

one has for $\hat{\eta}_j$ *defined in step (iii) of* **NEWTON**

$$\epsilon_{j+1} = (1+3C^*)\hat{\eta}_j \leq \beta\epsilon_j, \quad j = 0, 1, \ldots. \tag{8.27}$$

Proof: Denoting by $\mathbf{u}^1$ the exact solution of $D\mathbf{R}(\bar{\mathbf{u}}^0)\mathbf{u}^1 = -\mathbf{R}(\bar{\mathbf{u}}^0)$ and recalling from step (i) that $\|\mathbf{u} - \bar{\mathbf{u}}^0\|_{\ell_2(\mathcal{J})} \leq c_1^{-1}\delta = \epsilon_0$, the vector $\hat{\mathbf{u}}$ produced in steps (ii), (iii) satisfies, by (8.9) and (8.19),

$$\|\mathbf{u} - \hat{\mathbf{u}}\|_{\ell_2(\mathcal{J})} \leq \|\mathbf{u} - \mathbf{u}^1\|_{\ell_2(\mathcal{J})} + \|\mathbf{u}^1 - \hat{\mathbf{u}}\|_{\ell_2(\mathcal{J})} \leq \tilde{\omega}c_1^{-2}\delta^2 + \eta_0.$$

Thus, taking the coarsening step into account, we infer from (2.5) and (8.9) that

$$\begin{aligned}\|u - \bar{u}^1\|_{\mathcal{H}} &\leq C_1\|\mathbf{u} - \bar{\mathbf{u}}^1\|_{\ell_2(\mathcal{J})} \leq (1+3C^*)C_1(\tilde{\omega}c_1^{-2}\delta^2 + \eta_0) \\ &= \frac{(1+3C^*)C_1^3\omega\delta^2}{2c_1^3} + (1+3C^*)C_1\eta.\end{aligned}$$

Thus, when e.g. $\frac{(1+3C^*)C_1^3\omega\delta^2}{2c_1^3} < \delta/2$, which is (8.24), it suffices to take $\eta_0 < \delta/(2(1+3C^*)C_1)$ at the initial stage $j = 0$, to ensure that

$$\|\bar{u}^1 - u\|_{\mathcal{H}} < \|u - u^0\|_{\mathcal{H}},$$

which verifies that $\bar{u}^1 \in B_\delta(u)$. We can now iterate this result, e.g. using $\bar{u}^1$ in place of $\bar{u}^0$, we obtain that $\bar{u}^2 \in B_\delta(u)$, and so on. Now when (8.25) holds we have $\beta > (1+3C^*)\tilde{\omega}\epsilon_0$ so that the condition (8.26) on η_j is feasible for $j = 0$. Moreover, (8.26) implies that $\epsilon_{j+1} = (1+3C^*)\hat{\eta}_j \leq \beta\epsilon_j$ which is (8.27) which ensures that the error bounds decay, so that (8.26) remains feasible for all j. This completes the proof. □

PROPOSITION 8.4. *Assume that* δ *and* η *satisfy (8.24), (8.25) and (8.26), respectively. Then the scheme* **NEWTON** *terminates after finitely many steps and produces a finitely supported vector* $\bar{\mathbf{u}}(\epsilon)$ *satisfying*

$$\|\mathbf{u} - \bar{\mathbf{u}}(\epsilon)\|_{\ell_2(\mathcal{J})} \leq \epsilon. \tag{8.28}$$

Thus, by (2.5),

$$\|u - \sum_{\lambda \in \operatorname{supp}\bar{\mathbf{u}}(\epsilon)} u(\epsilon)_\lambda \psi_\lambda\|_{\mathcal{H}} \leq C_1\epsilon.$$

Proof: We employ a simple perturbation argument as in the proof of Lemma 8.1. Let $\mathbf{u}^{j+1}$ denote the exact Newton iteration $\mathbf{u}^{j+1} = \bar{\mathbf{u}}^j - D\mathbf{R}(\bar{\mathbf{u}}^j)^{-1}\mathbf{R}(\bar{\mathbf{u}}^j)$. By step (i) we know that $\|\mathbf{u} - \bar{\mathbf{u}}^0\|_{\ell_2(\mathcal{J})} \leq \epsilon_0$. Then, supposing that $\|\mathbf{u} - \bar{\mathbf{u}}^j\|_{\ell_2(\mathcal{J})} \leq \epsilon_j$, we infer from (8.9) that

$$\|\mathbf{u} - \mathbf{u}^{j+1}\|_{\ell_2(\mathcal{J})} \leq \tilde{\omega}\epsilon_j^2. \tag{8.29}$$

Hence, denoting by $\mathbf{w}^j := \mathbf{u}^{j+1} - \bar{\mathbf{u}}^j$ the exact solution of $D\mathbf{R}(\bar{\mathbf{u}}^j)\mathbf{w} = -\mathbf{R}(\bar{\mathbf{u}}^j)$, we obtain according to step (iii)

$$\begin{aligned}\|\mathbf{u} - \bar{\mathbf{u}}^{j+1}\|_{\ell_2(\mathcal{J})} &\le \|\mathbf{u} - \mathbf{u}^{j+1}\|_{\ell_2(\mathcal{J})} + \|\mathbf{u}^{j+1} - \hat{\mathbf{u}}\|_{\ell_2(\mathcal{J})} + \|\hat{\mathbf{u}} - \bar{\mathbf{u}}^{j+1}\|_{\ell_2(\mathcal{J})} \\ &\le \tilde{\omega}\epsilon_j^2 + \|\mathbf{w}^j - \bar{\mathbf{w}}^j\|_{\ell_2(\mathcal{J})} + 3\hat{\eta}_j \le \tilde{\omega}\epsilon_j^2 + \eta + 3C^*\hat{\eta}_j \\ &= (1 + 3C^*)\hat{\eta}_j = \epsilon_{j+1}, \end{aligned} \tag{8.30}$$

which advances the induction assumption. By (8.27), this finishes the proof. □

It remains to analyze the work/accuracy rate of **NEWTON**. So far the only condition on the tolerances η_j in step (ii) of **NEWTON** is (8.26) which ensures that the error bounds ϵ_j decay at all. This would allow us to keep η_j proportional to ϵ_j which would result in an overall first order error reduction rate. On the other hand, choosing η_j proportional to ϵ_j^2, the error bounds ϵ_j decay, by step (iii), quadratically. However, according to the earlier discussion of issue b), the subproblem (8.1) should not be resolved too accurately, as reflected by the above right hand side scheme $\mathbf{RHS}_{\mathrm{N}}$, see (8.18). The following main result of this section says that within these constraints on the tolerances η_j one can still realize an outer convergence rate, ranging from first to second order, in such a way that the overall scheme exhibits optimal rate/complexity.

THEOREM 8.2. *Suppose that (N1), (N2) and the above hypotheses on F hold. Assume that δ satisfies (8.24) and (8.25) for some fixed $\beta < 1$. Moreover, assume that at the jth stage of* **NEWTON** *the tolerance η_j is in addition to (8.26) subjected to the condition*

$$\eta_j \rho^K \ge \zeta \hat{C} \epsilon_j^2 \quad \text{for some fixed } \zeta > 1, \tag{8.31}$$

where ρ, K are the constants from (3.7) and (4.8). Then, for any target accuracy $\epsilon > 0$, **NEWTON** *outputs after finitely many steps a finitely supported vector $\bar{\mathbf{u}}(\epsilon)$ satisfying $\|\mathbf{u} - \bar{\mathbf{u}}(\epsilon)\|_{\ell_2(\mathcal{J})}$ and hence*

$$\|u - \sum_{\lambda \in \operatorname{supp} \bar{\mathbf{u}}(\epsilon)} \bar{u}(\epsilon)_\lambda \psi_\lambda\|_{\mathcal{H}} \le C_1 \epsilon.$$

Moreover, if the solution $\mathbf{u}$ of (2.9) belongs to ${}_t\ell_\tau^w(\mathcal{J})$ for some $d/\gamma < \tau < 2$, the output $\mathbf{u}(\epsilon)$ of **NEWTON** *has the following properties:*

$$\|\mathbf{u}(\epsilon)\|_{{}_t\ell_\tau^w(\mathcal{J})} \lesssim \|\mathbf{u}\|_{{}_t\ell_\tau^w(\mathcal{J})}, \quad \#\operatorname{supp} \mathbf{u}(\epsilon) \lesssim \|\mathbf{u}\|^{1/s}_{{}_t\ell_\tau^w(\mathcal{J})} \epsilon^{-1/s}, \tag{8.32}$$

where as before $\tau^{-1} = s + 1/2$. Under the assumption E the number of floating point operations needed to compute $\bar{\mathbf{u}}(\epsilon)$ remains proportional to $\#\operatorname{supp} \mathbf{u}(\epsilon)$.

It is understood that in the final step η_j is chosen within the above constraints as large as possible so as to attain the target accuracy. Note that, as indicated before, within the above constraints on η_j (see (8.26), (8.31)) the convergence rate of the outer inexact Newton iteration can be chosen to vary between linear and quadratic convergence.

Proof of Theorem 8.2: First, observe that in step (iii) one has at the jth stage, by the the first part of (8.30),

$$\|\mathbf{u} - \hat{\mathbf{u}}\|_{\ell_2(\mathcal{J})} \le \|\mathbf{u} - \mathbf{u}^{j+1}\|_{\ell_2(\mathcal{J})} + \|\mathbf{u}^{j+1} - \hat{\mathbf{u}}\|_{\ell_2(\mathcal{J})} \le \tilde{\omega}\epsilon_j^2 + \eta = \hat{\eta}_j. \tag{8.33}$$

Hence, by Proposition 6.5 and step (iii), we obtain

$$\|\bar{\mathbf{u}}^{j+1}\|_{{}_t\ell_\tau^w(\mathcal{J})} \lesssim \|\mathbf{u}\|_{{}_t\ell_\tau^w(\mathcal{J})}, \quad \#\operatorname{supp} \bar{\mathbf{u}}^{j+1} \lesssim \|\mathbf{u}\|^{1/s} \epsilon_{j+1}^{-1/s}, \tag{8.34}$$

where, as before, $\frac{1}{\tau} = s + \frac{1}{2}$. Moreover, the computational work remains proportional to $\#\mathrm{supp}\, \bar{\mathbf{u}}^{j+1}$.

Thus the only question concerns the intermediate complexity in step (ii). Here it is important that the Newton updates (8.1) are, in view of (8.31), not computed too accurately. In fact, under this constraint the approximation of $\mathbf{R}(\bar{\mathbf{u}}^j)$, computed in $\mathbf{SOLVE}_{\mathrm{N}}$, is incidentally also a sufficiently accurate approximation to $\mathbf{G}(\bar{\mathbf{u}}^j)$, see Remark 8.2. To explain this, recall (8.18) and set

$$\mathbf{Y}^j := \mathbf{EV}\,[\eta_j - \hat{C}\epsilon_j^2, -\mathbf{R}(\bar{\mathbf{u}}^j)].$$

Then one has, by Remark 8.2,

$$\begin{aligned}\|\mathbf{G}(\bar{\mathbf{u}}^j) - \mathbf{Y}^j\|_{\ell_2(\mathcal{J})} &\le \|\mathbf{G}(\bar{\mathbf{u}}^j) + \mathbf{R}(\bar{\mathbf{u}}^j)\|_{\ell_2(\mathcal{J})} + \|\mathbf{R}(\bar{\mathbf{u}}^j) + \mathbf{Y}^j\|_{\ell_2(\mathcal{J})} \\ &\le \hat{C}\epsilon_j^2 + \eta_j - \hat{C}\epsilon_j^2 = \eta_j.\end{aligned}$$

Thus, within the accuracy range permitted by (8.31), the routine $\mathbf{RHS}_{\mathrm{N}}$ invoked by $\mathbf{SOLVE}_{\mathrm{N}}$ satisfies the accuracy requirements for the perturbed equation $D\mathbf{R}(\bar{\mathbf{u}}^j)\hat{\mathbf{w}} = \mathbf{G}(\bar{\mathbf{u}}^j)$, whose solution is, by definition of $\mathbf{G}(\bar{\mathbf{u}}^j)$, just $\hat{\mathbf{w}}^j = \mathbf{u} - \bar{\mathbf{u}}^j$. Now the ${}_t\ell_\tau^w(\mathcal{J})$-norm of $\hat{\mathbf{w}}^j$ is, by assumption and the reasoning at the beginning of the proof, uniformly bounded. Therefore Theorem 6.1, combined with Remark 8.3, bounds the complexity of the $\bar{\mathbf{w}}^j$, which completes the proof. □

REFERENCES

[1] A. Barinka, Fast evaluation tools for adaptive wavelet schemes, PhD. Dissertation, RWTH Aachen, 2002, in preparation.

[2] A. Barinka, T. Barsch, P. Charton, A. Cohen, S. Dahlke, W. Dahmen, K. Urban, Adaptive wavelet schemes for elliptic problems – Implementation and numerical experiments, SIAM J. Scient. Comput., 23, (2001), 910–939.

[3] P. Binev, and R. DeVore, Fast computation in adaptive tree approximation, preprint, June 2002,

[4] P. Binev, W. Dahmen, R. DeVore, Adaptive Finite Element Methods with Convergence Rates, IGPM Report, RWTH Aachen, June 2002.

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

[6] A. Canuto, A. Tabacco, K. Urban, The wavelet element method, part II: Realization and additional features, Preprint # 1052, Istituto del Analisi Numerica del C.N.R. Pavia, 1997, to appear in Appl. Comp. Harm. Anal.

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

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

[9] A. Cohen, W. Dahmen, R. DeVore, Sparse evaluation of nonlinear functionals of multiscale expansions, manuscript, 2001.

[10] A. Cohen, R. Masson, Wavelet adaptive methods for second order elliptic problems, boundary conditions and domain decomposition, Numer. Math. 86 (2000), 193–238.

[11] S. Dahlke, W. Dahmen, R. Hochmuth, R. Schneider, Stable multiscale bases and local error estimation for elliptic problems, Appl. Numer. Math., 23 (1997), 21–47.

[12] S. Dahlke, W. Dahmen, K. Urban, Adaptive wavelet methods for saddle point problems – Convergence rates, IGPM Report # 204, RWTH Aachen, July 2001, to appear in SIAM J. Numer. Anal.

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

[14] W. Dahmen and R. Schneider, Wavelets on Manifolds I: Construction and Domain Decomposition, SIAM J. Math. Anal., 31 (1999), 184–230.

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

[16] W. Dahmen, K. Urban, J. Vorloeper, Adaptive wavelet methods – basic consepts and applications, IGPM Report # 215, RWTH Aachen, May 2002.

[17] P. Deuflhard, M. Weiser, Local inexact Newton multilevel FEM for nonlinear elliptic problems, in: M-O. Bristeau, G. Etgen, W. Fitzigibbon, J-L. Lion, J. Periaux, M. Wheeler (eds.): Computational Science for the 21st Century, Tours, France. Wiley-Interscience-Europe, pp.129-138 (1997).

[18] P. Deuflhard, M. Weiser, Global inexact Newton multilevel FEM for nonlinear elliptic problems, in: W. Hackbusch, G. Wittum (eds.), Multigrid Methods, Lecture Notes in Computational Science and Engineering 3, Springer–Verlag, pp. 71-89 (1998).

[19] R. DeVore, Nonlinear approximation, Acta Numerica, **7**, Cambridge University Press, 1998, 51-150.

[20] J. Pousin, J. Rappaz, Consistency, stability, a-priori and a-posteriori errors for Petrov-Galerkin methods applied to nonlinear problems, Numer. Math., 69(1994), 213–231.

[21] M. Renardy, R.C. Rogers, *An Introduction to Partial Differential Equations*, Texts in Applied Mathematics, 13, Springer-Verlag, New York, 1993.

[22] R. Verfürth, *A Review of A Posteriori Error Estimation and Adaptive Mesh-Refinement Techniques*, Wiley-Teubner, New-York, 1996.

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

[24] E. Zeidler, *Nonlinear Functional Analysis and its Applications, III Variational Methods and Optimization*, Springer-Verlag, New York, 1985.