

Adaptive numerical methods for PDEs ¹

Ronald A. DeVore

Résumé. Les méthodes adaptatives sont d'usage courant pour la résolution numérique des EDP. Il n'existe cependant pas de théorie bien établie analysant leur performance et justifiant leur utilisation. L'objet de cet exposé est de présenter les premiers éléments d'une telle théorie, dont les pierres angulaires sont l'approximation non-linéaire et les théorèmes de régularité pour les EDP. Une méthode adaptative numérique peut être assimilée à une forme d'approximation non-linéaire, la solution de l'équation n'étant pas approchée par une suite d'espaces fixée à l'avance, mais par les éléments d'un ensemble non-linéaire de fonctions décrit par N paramètres (éléments finis adaptatifs, ondelettes). La théorie de l'approximation non-linéaire relie l'efficacité de ce type d'approximation à la régularité de la solution mesurée dans certains espaces de Besov. L'analyse de cette régularité pour une équation donnée détermine ainsi le potentiel d'une méthode adaptative par comparaison avec les méthodes linéaires pour lesquelles une théorie similaire existe et fait intervenir d'autres classes de régularité. Une telle approche est illustrée dans le cadre elliptique et hyperbolique. Nous présentons finalement un algorithme adaptatif utilisant les ondelettes pour les équations elliptiques développé par Albert Cohen, Wolfgang Dahmen et l'auteur, algorithme dont les performances sont optimales au sens où elles reflètent les taux d'approximation prédits par la théorie.

Mots clés. Equations aux dérivées partielles, méthodes numériques, méthodes adaptatives, régularité, vitesse de convergence

Abstract. While adaptive numerical methods are often used in solving partial differential equations, there is not yet a cohesive theory which justifies their use or analyzes their performance. The purpose of this talk is to put forward the first building blocks of such a theory, the cornerstones of which are nonlinear approximation and regularity theorems for PDEs. Any adaptive numerical method can be viewed as a form of nonlinear approximation: the solution u of the PDE is approximated by elements from a nonlinear manifold of functions. The theory of nonlinear approximation relates the efficiency of this type of approximation to the regularity of u in a certain family of Besov spaces. Regularity for PDEs are needed to determine the smoothness of u in this new Besov scale. Together, the approximation theory and regularity theory determine the efficiency of approximation that is possible using adaptive methods. A similar analysis gives the efficiency of linear algorithms. The two can then be compared to predict whether nonlinear methods would result in better performance. Examples will be given in the setting of both elliptic and hyperbolic problems. A wavelet based algorithm

¹This work has been supported in part by the Office of Naval Research Contract N0014-91-J1343 and the Army Research Office Contract DAAG55-98-1-0002

for elliptic equations developed by Albert Cohen, Wolfgang Dahmen, and the author will be presented as one of the successes of this theory.

Key words. *Partial differential equations, numerical methods, adaptive methods, regularity, rate of convergence*

AMS subject classification. 41A25, 41A46, 65N99, 65N12, 65N55 35J, 65N, 73C

1 Overview

Adaptive methods are an important tool for numerically solving Partial Differential Equations (PDEs). Many adaptive numerical algorithms have been suggested for both elliptic equations and time varying problems. Yet, it is rare indeed when there is any accompanying theory which shows that a specific algorithm converges. More so, we are not aware of any work which bounds the decrease of error in terms of the number of parameters or better yet the number of arithmetic operations. Then, how can we tell whether adaptive methods are effective or even necessary in resolving a given PDE? The purpose of this talk is to put down the foundations of a theory which can provide answers to this question. The theory is built on two pillars: approximation theory and regularity for PDEs.

Approximation theory enters the picture in a natural way since any numerical method generates a sequence u_n of approximations to the solution u with n denoting the number of parameters used in the approximation. One can view the nonadaptive methods for solving PDEs, such as standard Finite Element Methods (FEM), as a form of linear approximation. Namely, a sequence of linear spaces \mathcal{S}_n , $n = 1, 2, \dots$, is selected in advance, independent of the solution u . The functions in \mathcal{S}_n are typically piecewise polynomials subordinate to a fixed partition Δ_n . The numerical method generates an approximant $u_n \in \mathcal{S}_n$. The word linear refers to the fact that the spaces \mathcal{S}_n are linear spaces and are fixed in advance independent of u .

Adaptive methods, on the other hand, generate their approximants in a nonlinear fashion since the form of the approximant u_n depends on u and the previous approximants u_k , $k < n$. Usually, the complete process can be described as approximating u from a nonlinear space Σ_n (which depends on n parameters). For example in adaptive finite elements, the space Σ_n would consist of piecewise polynomials (of a fixed degree and prescribed smoothness across boundaries of the partition) on an adaptively generated partition into n simplices.

One of the many successes of approximation theory during the last decades has been to characterize, for a specified approximation process, the class of functions with a prescribed rate of approximation. For example, in the case of linear approximation, this theory tells us necessary and sufficient conditions on a function v in order that v can be approximated in a given norm $\|\cdot\|_X$ by the elements of \mathcal{S}_n with an accuracy $O(n^{-s})$, $n \rightarrow \infty$. If the spaces \mathcal{S}_n have dimension of order n ($\dim \mathcal{S}_n \approx n$), the characterization is that v should have a certain smoothness (depending on s) measured in X . For example if $X = L_p(\Omega)$, with $\Omega \subset \mathbb{R}^d$, then v should have smoothness of order sd in $L_p(\Omega)$, roughly speaking, v should be in the Sobolev space $W^{sd} = W^{sd}(L_p(\Omega))$. We shall recall this linear theory of approximation in §4.

There is a similar theory for nonlinear approximation which characterizes the functions v which can be approximated with an order like $O(n^{-s})$ by the elements of Σ_n . These characterizations are usually of the form that v should be in a certain Besov space B^{sd} (see §5). This Besov space is significantly larger than the Sobolev space W^{sd} showing the potential advantage of nonlinear methods. But whether this is actually so for a specific function v depends on the smoothness of v in these two scales (the Sobolev scale for linear and Besov scale for nonlinear).

To clarify this, let s_L be the largest value of s such that v is in W^s and let s_{NL} be the largest value of s such that v is in the Besov space B^s . If $s_{NL} > s_L$, then we can approximate v better by the elements of Σ_n than by the elements of \mathcal{S}_n . In other words, nonlinear is definitely better than linear. But it could happen that $s_L = s_{NL}$ and there is no gain in using nonlinear methods.

Armed with this knowledge, to ascertain the potential performance of adaptive methods, i.e. to determine the value of s , we need to know the regularity of the solution u in the two scales of smoothness spaces corresponding to linear and nonlinear approximation. That is, we need to determine s_L and s_{NL} for the solution u and decide whether s_{NL} is larger than s_L . It is the role of regularity theorems for PDEs to provide the answer to this question.

In other words, regularity theorems for PDEs, coupled with the existing theory of linear and nonlinear approximation can decide in advance whether adaptive algorithms have any potential to perform better than the simpler linear algorithms for numerically resolving PDEs.

Establishing the fact that $s_{NL} > s_L$ does not, in and of itself, tell us how a particular adaptive algorithm performs. Indeed, each specific adaptive algorithm has to be analyzed individually for its performance. However, the above tells us what form such an analysis should take. Namely, the error produced by the algorithm should be bounded in terms of Besov regularity of the solution u . It does remain, however, a substantial problem to establish the performance of any proposed numerical algorithm.

There is only one case in which we know of a satisfactory error analysis for a specific adaptive method. This is the analysis of the wavelet based adaptive method for elliptic equations which has recently been proposed by Albert Cohen, Wolfgang Dahmen, and the speaker [4]. This paper constructs an algorithm which for each prescribed $\epsilon > 0$ approximates u (in the energy norm) to tolerance ϵ using at most $C\epsilon^{-1/s_{NL}}$ wavelet functions and at most this many arithmetic operations. In other words, it operates at the maximal efficiency possible for nonlinear wavelet based algorithms. Later in this presentation (§7); we shall touch upon the essential ingredients of this method which allow such an error analysis.

The remainder of this talk will try to put flesh on the theory outlined above. It consists of the following: (i) a review of the existing theory of linear and nonlinear approximation as they pertain to numerical methods for PDEs, (ii) a discussion of known regularity theorems for PDEs in the two scales of smoothness spaces relating to linear and nonlinear methods, (iii) the introduction and analysis of the adaptive wavelet based algorithm for elliptic operator equations introduced in [4].

When discussing adaptive strategies, our emphasis in this talk will be on wavelet based methods rather than Finite Element Methods. The results in the wavelet case are cleaner and theoretically more far reaching than for FEM. Also, the intuition behind adaptive wavelet based methods is more easily explained. We shall indicate the nature of the results in the complimentary FEM theory as well. We begin in the next section with a rudimentary discussion of the aspects of

wavelet theory that we shall need.

2 Wavelets

As noted above, much of the following discussion of linear and nonlinear methods will be formulated in the context of wavelets. We shall say only a few words about wavelet bases. This will also allow us to set some notation to which we adhere in this talk. There are many excellent accounts of wavelets and we recommend either of the books [15], [9]. Also, the finer aspects of wavelets as they relate to numerical methods for PDEs can be found in the excellent accounts [8] and [3].

Given a domain (or manifold) $\Omega \subset \mathbb{R}^d$, a wavelet basis can be described on one level as simply a collection of functions that form either an orthogonal or Riesz basis for $L_2(\Omega)$. However, there are several issues that separate the wavelet bases from others. One is the space/frequency character of these bases; wavelet bases possess local support in space and frequency to the extent that this is possible given the Heisenberg uncertainty principle. Secondly, wavelet bases are efficient for computation: the first n discrete wavelet coefficients can be computed with $O(n)$ arithmetic operations and reconstruction of the function from these coefficients is equally efficient. Wavelet bases also have a homogeneous structure. In the ideal case, each function in the wavelet bases is a shifted dilate of a finite number ($2^d - 1$ in the case of \mathbb{R}^d) functions (called the mother wavelets). The wavelet bases also give complete characterizations of most smoothness spaces in terms of sequence norms applied to the coefficients in wavelet decompositions. In solving elliptic equations, the stiffness matrix has a sparse well conditioned representation in terms of the wavelet basis.

To obtain a feeling for wavelet bases, it suffices to understand the simplest wavelet, the Haar function

$$(2.1) \quad H(x) := \chi_{[0,1/2)} - \chi_{[1/2,1)} = \begin{cases} 1, & 0 \leq x < 1/2 \\ -1, & 1/2 \leq x < 1. \end{cases}$$

Let $\mathcal{D} = \mathcal{D}(\mathbb{R})$ denote the collection of all dyadic intervals and $\mathcal{D}_j = \mathcal{D}_j(\mathbb{R})$ the collection of dyadic intervals of side length 2^{-j} . If $I = 2^{-j}[k, k+1)$ is a dyadic interval then the Haar function

$$(2.2) \quad H_I := H_{j,k}(x) := 2^{j/2} H(2^j x - k), \quad j, k \in \mathbb{Z},$$

is a shifted dilate of H scaled to I and normalized in $L_2(\mathbb{R}^d)$. The collection $(H_I)_{I \in \mathcal{D}} = (H_{j,k})_{j,k \in \mathbb{Z}}$ of all of these shifted dilates form an orthonormal basis for $L_2(\mathbb{R})$. We speak of j being the dyadic level associated to $H_I = H_{j,k}$ and 2^j as being its frequency localization and I as its support cube.

We also have Haar bases associated to finite domains. In the case, the domain $\Omega = [0, 1]$, the characteristic function χ_Ω (called the scaling function in the case of the Haar basis) together with the Haar functions $H_I, I \subset \Omega$, form an orthonormal basis for $L_2(\Omega)$. Each function $f \in L_2(\Omega)$ has the representation

$$(2.3) \quad f = \langle f, \chi_\Omega \rangle \chi_\Omega + \sum_{I \subset \Omega} \langle f, H_I \rangle H_I$$

with $\langle \cdot, \cdot \rangle$ the usual L_2 inner product. We can compute the $L_2(\Omega)$ norm of f simply as

$$(2.4) \quad \|f\|_{L_2(\Omega)}^2 = |\langle f, \chi_\Omega \rangle|^2 + \sum_{I \subset \Omega} |\langle f, H_I \rangle|^2.$$

We also have a Haar basis for $L_2(\mathbb{R}^d)$. We introduce the following notation which is also utilized for more general constructions. Let $\mathcal{D} = \mathcal{D}(\mathbb{R}^d)$ denote the collection of dyadic cubes in \mathbb{R}^d and let $\mathcal{D}_j = \mathcal{D}_j(\mathbb{R}^d)$ denote those dyadic cubes in \mathcal{D} which have side length 2^{-j} . To each function $g \in L_2(\mathbb{R}^d)$ and each dyadic cube $I = 2^{-j}[\mathbf{k}, \mathbf{k} + \mathbf{1}]$, $\mathbf{k} \in \mathbb{Z}^d$, we introduce the function

$$(2.5) \quad g_I := 2^{jd/2} g(2^j \cdot -\mathbf{k})$$

which is a copy of g scaled and shifted relative to I and normalized in $L_2(\mathbb{R}^d)$.

Let $\psi^0 := \phi := \chi_{[0,1]}$ denote the scaling function and $\psi^1 := \psi := H$ denote the univariate Haar wavelet and let E' be the set of vertices of the cube $[0, 1]^d$ and $E \subset E'$ the set of nonzero vertices. For each vertex $e = (e_1, \dots, e_d)$ from E' , we denote the function

$$(2.6) \quad \psi^e(x_1, \dots, x_d) := \psi^{e_1}(x_1) \cdots \psi^{e_d}(x_d).$$

The collection of functions $\{\psi_I^e\}_{I \in \mathcal{D}, e \in E}$ are an orthonormal basis for $L_2(\mathbb{R}^d)$.

We can also obtain an orthogonal basis for $L_2(\mathbb{R}^d)$ by starting the wavelet decomposition at a fixed dyadic level. In this case, we have to adjoin a layer of scaling functions. For example, let Φ be the collection of all scaling functions χ_I , $I \in \mathcal{D}_0$, and Ψ be the collection of all Haar wavelets ψ_I^e , $I \in \mathcal{D}_+ := \cup_{j \geq 0} \mathcal{D}_j$, $e \in E$. Then, $\Phi \cup \Psi$ is an orthonormal basis for $L_2(\mathbb{R}^d)$: each $f \in L_2(\mathbb{R}^d)$ has the Haar decomposition

$$(2.7) \quad f = \sum_{I \in \mathcal{D}_0} \langle f, \chi_I \rangle \chi_I + \sum_{e \in E} \sum_{I \in \mathcal{D}_+} \langle f, \psi_I^e \rangle \psi_I^e.$$

The Haar decomposition (2.3) is just one example of a univariate wavelet decomposition. There are many generalizations. First, there are many univariate scaling functions ϕ which have an associated wavelet ψ for which the collection of shifted dilates $\{\psi_I\}_{I \in \mathcal{D}}$ form an orthonormal basis for $L_2(\mathbb{R})$. We call such functions ψ orthogonal wavelets. The most notable examples of orthogonal wavelets are the Meyer wavelets and the family of Daubechies wavelets $\psi = D_k$, $k = 1, \dots$. Here $D_1 = H$ and in general the D_k have compact support and by increasing the parameter k , we increase the smoothness of ψ (at the expense of increasing its support). Another generalization is to drop the requirement of orthogonality and replace this by biorthogonality as will be discussed below.

We can obtain multivariate wavelet bases for $L_2(\mathbb{R}^d)$ in exactly the same way as in the Haar case (see (2.6) and (2.7)). There are also wavelet decompositions for domains $\Omega \subset \mathbb{R}^d$. There are special constructions for domains such as cubes or polyhedral domains which utilize the wavelet ψ_I^e in the interior of the domain and some altered wavelets near the boundary. There are also constructions for more general domains and even manifolds. We refer the reader to the paper [8] for a discussion of these possibilities.

Let us now summarize the nature of wavelet decompositions and introduce the notation we shall utilize in this talk. We shall restrict ourselves to the setting of biorthogonal wavelet basis of compactly supported functions which is general enough for our purposes.

Given a domain $\Omega \subset \mathbb{R}^d$, we shall denote a wavelet basis for Ω by $\{\psi_\lambda\}_{\lambda \in \nabla}$. This notation incorporates the three parameters j, k, e (or equivalently I, e) into the one λ . We use $|\lambda| := j$ to denote the dyadic level of the wavelet. We let $\Psi_j = \{\psi_\lambda : \lambda \in \nabla_j\}$, $\nabla_j := \{\lambda \in \nabla : |\lambda| = j\}$, consist of the wavelets at level j .

There are several constructions of wavelet bases of compactly supported functions for domains (and manifolds) Ω . We refer the reader to [8] for a discussion of this topic but remark that roughly speaking there are such constructions for Lipschitz domains.

We shall limit ourselves to the case where the wavelet decomposition begins at some fixed level j_0 . For notational convenience only, we assume $j_0 = 1$. We define Ψ_0 to be the set of scaling functions in the wavelet basis. We shall assume that Ω is a domain or manifold which admits two sets of functions:

$$(2.8) \quad \Psi = \{\psi_\lambda : \lambda \in \nabla\} \subset L_2(\Omega), \quad \tilde{\Psi} = \{\tilde{\psi}_\lambda : \lambda \in \nabla\} \subset L_2(\Omega)$$

that form a *biorthogonal* wavelet bases on Ω : writing $\langle \Theta, \Phi \rangle := (\langle \theta, \phi \rangle_{L_2(\Omega)})_{\theta \in \Theta, \phi \in \Phi}$ for any two collections Θ, Φ of functions in $L_2(\Omega)$, one has

$$(2.9) \quad \langle \Psi, \tilde{\Psi} \rangle = \mathbf{I},$$

where \mathbf{I} is the identity matrix.

A typical feature in the theory of biorthogonal bases is that the sequences $\Psi, \tilde{\Psi}$ are Riesz-bases. That is, using the shorthand notation $\mathbf{d}^T \Psi := \sum_{\lambda \in \nabla} d_\lambda \psi_\lambda$, one has

$$(2.10) \quad \|\mathbf{d}\|_{\ell_2(\nabla)} \sim \|\mathbf{d}^T \Psi\|_{L_2(\Omega)} \sim \|\mathbf{d}^T \tilde{\Psi}\|_{L_2(\Omega)}.$$

3 Smoothness spaces

There are many ways to measure the smoothness of functions resulting in smoothness spaces such as the Lipschitz spaces, Sobolev spaces and potential spaces. The basic idea is to introduce a notion of α orders of smoothness (α derivatives) in L_p . In describing the approximation properties of functions, the Besov spaces are the most important of these different possibilities and we shall therefore restrict our discussion to this family.

The Besov spaces were originally described using moduli of smoothness and this remains the most flexible approach. However, there is a simpler description of these spaces in terms of wavelet decompositions which is equivalent to the moduli of smoothness definition in all cases of interest to us in this talk. We shall therefore present the Besov spaces in the wavelet context

Besov spaces depend on three parameters α, p , and q . The first parameter α gives the order of smoothness which can be thought of as the number of derivatives (although we allow noninteger values of α). The second parameter p indicates the L_p space in which the smoothness is measured. Thus the α and p say roughly that the function should have α derivatives in L_p . The third index q is a fine tuning index and serves to make subtle distinctions in the spaces once α and p are fixed.

In most discussions it is sufficient to think in terms of the primary indices α and p . We invoke q only in certain cases when we want a precise result.

To define the Besov spaces, we return to wavelet decompositions as discussed above. In this talk, Ω will always denote a domain in \mathbb{R}^d which admits a wavelet basis $\Psi := \{\psi_\lambda\}_{\lambda \in \Delta}$ and $\tilde{\Psi}$ will be its dual basis. For each f which is locally in $L_1(\Omega)$, we have the wavelet coefficients

$$(3.1) \quad c_\lambda(f) := \langle f, \tilde{\psi}_\lambda \rangle.$$

The wavelet coefficients can also be defined for distributions f provided the order of the distribution is commensurate with the smoothness of $\tilde{\psi}_\lambda$.

The wavelets and wavelet coefficients given above are normalized for $L_2(\Omega)$. It will also be convenient to have normalizations for $L_p(\Omega)$ when $p \neq 2$. For any $0 < p \leq \infty$, we define

$$(3.2) \quad \psi_{\lambda,p} := 2^{jd(1/p-1/2)} \psi_\lambda$$

which are wavelets normalized for $L_p(\Omega)$ and the wavelet coefficients

$$(3.3) \quad c_{\lambda,p}(f) := \langle f, \tilde{\psi}_{\lambda,p'} \rangle = 2^{jd(1/2-1/p)} c_\lambda(f),$$

where $1/p' + 1/p = 1$ (this serves to define p' even when $p < 1$ in which case $p' < 0$).

The Besov spaces $B_q^\alpha(L_p(\Omega))$ for $0 < q, p \leq \infty$, $\alpha > 0$ is the collection of all distributions f for which

$$(3.4) \quad \|f\|_{B_q^\alpha(L_p(\Omega))} \asymp \begin{cases} \left(\sum_{j=0}^\infty 2^{j\alpha q} \left(\sum_{\lambda \in \nabla_j} c_{\lambda,p}(f)^p \right)^{q/p} \right)^{1/q}, & 0 < q < \infty, \\ \sup_{j \geq 0} 2^{j\alpha} \left(\sum_{\lambda \in \nabla_j} c_{\lambda,p}(f)^p \right)^{1/p}, & q = \infty. \end{cases}$$

is finite. The right side of (3.4) defines a quasi-norm for this space. Several remarks are in order to explain (3.4)

Remark (i): We can define spaces of functions for all $\alpha > 0$ by using the right side of (3.4). However, these spaces will coincide with Besov spaces only for a certain range of α and p that depend on the wavelet ψ . For example, in the case $1 \leq p \leq \infty$, we need that (a) $\psi \in B_q^\beta(L_p(\mathbb{R}^d))$, for some $\beta > \alpha$, (b) ψ has r vanishing moments with $r > \alpha$. When $p < 1$, we also need that $r > d/p - d$ (see the following remark).

Remark (ii) When $p < 1$ and $\Omega = \mathbb{R}^d$, (3.4) characterizes the space $B_q^\alpha(H_p(\mathbb{R}^d))$ (with the correct range of parameters) where this latter Besov space can be defined by replacing the L_p modulus of smoothness by the modulus of smoothness in the Hardy space H_p (see Kyriazis (1996)). However, if $\alpha > d/p - d$, this space is the same as $B_q^\alpha(L_p(\mathbb{R}^d))$. The last statement also holds if Ω is a domain.

Remark (iii) For a fixed value of $0 \leq p < \infty$, the spaces $B_\tau^\alpha(L_\tau(\Omega))$, $1/\tau = \alpha/d + 1/p$, play a prominent role in nonlinear approximation. If we chose the wavelets normalized in L_p , then the characterization (3.4) becomes simply

$$(3.5) \quad \|f\|_{B_\tau^\alpha(L_\tau(\Omega))} \asymp \left(\sum_{I \in \Delta} c_{\lambda,p}(f)^\tau \right)^{1/\tau}.$$

Let us also say a few words about the connection of Besov spaces to Sobolev spaces since the latter may be more familiar. In the case $p = 2$, the Sobolev space $W^\alpha(L_2(\Omega))$ (frequently denoted by H^α in the FEM literature) is the same as the Besov space $B_2^\alpha(L_2(\Omega))$ and their norms are equivalent. For other values of $p \neq 2$, the Sobolev space $W^k(L_p(\Omega))$, $k = 1, 2, \dots$, is not a Besov space but we have the embeddings

$$(3.6) \quad B_p^k(L_p(\Omega)) \subset W^k(L_p(\Omega)) \subset B_\infty^k(L_p(\Omega))$$

For α not an integer, the Sobolev space $W^\alpha(L_p(\Omega))$ (by its very definition) coincides with $B_p^\alpha(L_p(\Omega))$.

It is useful to have a pictorial description of smoothness spaces. We shall correspond smoothness spaces with points in the upper right quadrant of \mathbb{R}^2 . Namely, a smoothness space consisting of functions of smoothness order α in L_p will be identified with the point $(1/p, \alpha)$ (see Figure 3.1). This identification is coarse in the sense that several spaces are identified with the same point. For example all space $B_q^\alpha(L_p(\Omega))$ are identified with $(1/p, \alpha)$ irrespective of q . We will come back to this picture often but at this stage let us just point out, as an example, how to interpret the Sobolev embedding theorem in this picture. The line with slope d passing through $(1/p, 0)$ is the demarkation line for embeddings of Besov spaces into $L_p(\Omega)$ (see Figure 3.1). Any Besov space with primary indices corresponding to a point above that line is embedded into $L_p(\Omega)$ (regardless of the secondary index q). Besov spaces corresponding to points on the demarkation line may or may not be embedded in $L_p(\Omega)$. For example the Besov spaces $B_\tau^\alpha(L_\tau(\Omega))$ with $1/\tau = \alpha/d + 1/p$ correspond to points on the demarkation line and they are embedded in $L_p(\Omega)$. Points below the demarkation line are never embedded in $L_p(\Omega)$.

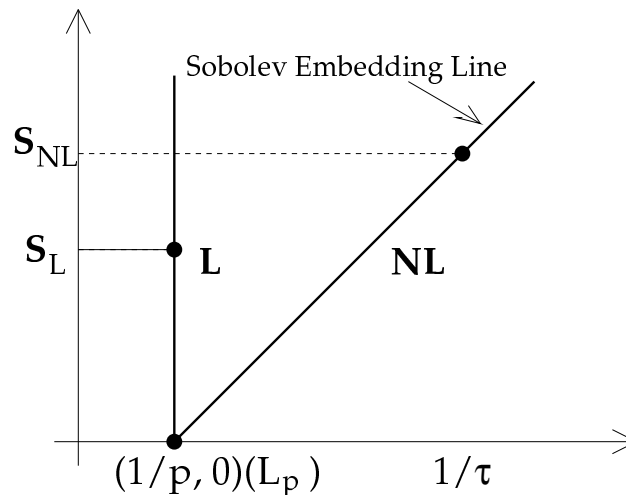


Figure 3.1: Graphical interpretation of linear and nonlinear approximation ($d = 1$).

4 Linear methods

After this short discourse concerning wavelets and smoothness spaces, let us return to our main thread of thought which is the understanding of when the solution to our PDE can be approximated

with a prescribed efficiency by linear or nonlinear methods. We begin with linear methods.

Let us consider a function u defined on a domain $\Omega \subset \mathbb{R}^d$ which is the solution to a PDE which we wish to numerically resolve. We shall call u the target function. We shall consider numerical methods which fix a sequence X_n of linear spaces whose dimension is of order n and approximate u by an element $A_n(u)$ of X_n . This type of numerical algorithm is said to be linear because the approximants $A_n(u)$ come from the linear space X_n which is fixed in advance (does not change with u).

To assess the performance of such a numerical algorithm, we would choose a norm $\|\cdot\|$ in which we want to measure error. Typical choices are the energy norm or the L_2 norm for elliptic problems, the L_1 norm for conservation laws, and the L_∞ norm for Hamilton-Jacobi equations.

The error for approximating u with this algorithm is measured by

$$(4.1) \quad e_n(u) := \|u - A_n(u)\|.$$

As a benchmark for the performance of the numerical algorithm, it would be useful to compare this error with the ideal error

$$(4.2) \quad E_n(u) := E(u, X_n) := \inf_{g \in X_n} \|u - g\|.$$

In some ideal cases, this is made with great success. For example, for elliptic problems in which $A_n(u)$ is the Galerkin approximation to u from X_n , we have that $A_n(u)$ is the best approximation to u in the energy norm and so

$$(4.3) \quad E_n(u) = \|u - A_n(u)\| = e_n(u).$$

Let us consider our two main examples. Standard Finite Element Methods would begin with a sequence (Δ_n) of partitions of Ω and a corresponding spaces $\mathcal{S}_n := \mathcal{S}(\Delta_n)$ of piecewise polynomials on that partition. Typical assumptions are that Ω is a polyhedral domain and the elements of the partition are simplices. To be useful, the space \mathcal{S}_n should admit a nice basis (N_k) with an accessible dual basis; usually coefficients in the representation $S = \sum c_k N_k$, for a given $S \in \mathcal{S}_n$, are determined by nodal values of S or its derivatives (their values at the vertices of the partition). We assume that the dimension of \mathcal{S}_n is of order n .

In linear wavelet methods, we would fix a sequence $\Lambda_n \subset \Delta$, $n = 1, 2, \dots$, of indexing sets and consider the spaces

$$\mathcal{W}_n := \text{span}\{\psi_\lambda : \lambda \in \Lambda_n\}.$$

We shall always assume that $\Lambda_n \subset \Lambda_{n+1}$, $n \geq 1$, so that the spaces \mathcal{W}_n are likewise nested. A typical choice is Λ_n to be the first n wavelets in their natural order. The numerical method would choose a value of n then create an approximant $W_n(u)$ to u from \mathcal{W}_n .

Given one of these linear methods of approximation and given our target function u , we introduce the real number $s_L > 0$ defined by the properties that for each $s < s_L$,

$$(4.4) \quad E_n(u) \leq M n^{-s},$$

and further for each $s > s_L$,

$$(4.5) \quad \limsup_{n \rightarrow \infty} n^s E_n(u) = \infty.$$

In other words, s_L is the supremum over all s satisfying (4.4). No numerical algorithm which generates approximations from X_n can provide approximations u_n which provide accuracy better than $O(n^{-s_L})$.

How can we determine the value of s_L for our target function u ? We call on our two pillars: approximation theory and regularity theorems for PDEs. Let us first consider the role of approximation theory. For the statement of the following theorem, we fix a domain $\Omega \subset \mathbb{R}^d$ on which the PDE is posed and fix an $L_p(\Omega)$ norm $\|\cdot\|_p$ with $1 \leq p \leq \infty$ in which we shall measure error. Similar results hold when the error is measured in a Sobolev or Besov norm.

The following generic theorem holds for a variety of settings which we shall delineate in a moment:

Theorem 4.1 *There is a real number $r > 0$, such that for any $0 < s < r$, we have that a function $v \in L_p(\Omega)$ satisfies*

$$(4.6) \quad E(v, X_n)_p \leq M n^{-s/d}, \quad n = 1, 2, \dots,$$

if and only if v is in the Besov space $B_\infty^s(L_p(\Omega))$.

When it applies, this theorem completely characterizes the functions which can be approximated with order $O(n^{-s})$. It says that to achieve this error it is necessary and sufficient that v has s orders of smoothness in L_p .

Theorem 4.1 holds in a variety of settings. We discuss the two main settings of interest to us in this talk. If \mathcal{W}_n denotes the wavelet space spanned by the first n wavelets in the wavelet basis for $L_2(\Omega)$ (see §2), then this theorem holds with $r = \min(m, \rho)$ with m the number of vanishing moments of the wavelets and ρ the smoothness of the wavelets as measured in $L_p(\Omega)$ (the wavelet should be in $B_\infty^\rho(L_p(\Omega))$).

The situation for approximation using piecewise polynomials is a little less clean. In fact, spaces $\mathcal{S}(\Delta_n)$ of piecewise polynomials of fixed degree which are defined by continuity assumptions across the boundaries of the simplices T are not completely understood from the standpoint of their dimension or approximation properties. On the other hand, the spaces used in FEM all have stable bases which form good partitions of unity and the following remarks apply to approximation from these spaces (see [2])

We assume that Δ_n is a partition of a fixed polyhedral domain into a collection of n simplices T . We assume that this simplicial decomposition is more or less uniform and that each simplex $T \in \Delta_n$ satisfies the shape condition. This means that there are balls B_ρ and B_R of radius ρ and R respectively such that

$$(4.7) \quad B_\rho \subset T \subset B_R$$

and

$$(4.8) \quad c_0 \leq \frac{R}{\rho} \leq c_1 c_0 n^{-1/d} \leq r \leq R \leq c_1 n^{-1/d}$$

with absolute constants $c_0, c_1 > 0$. Let $\mathcal{S}_n := \mathcal{S}(\Delta_n)$ be a linear space of piecewise polynomials of fixed degree m subordinate to Δ_n which admits a good partition of unity in the sense of [2]. Then, Theorem 4.1 holds for $X_n = \mathcal{S}_n$ and $r := \min(m + 1, \rho)$ where ρ is the smoothness of the elements $S \in \mathcal{S}_n$ as measured in L_p (each S is assumed to be in $B_\infty^\rho(L_p(\Omega))$). The generic

theorem holds for a larger range of s if the partitions Δ_n are not nested but rather satisfy certain mixing conditions (see [10] for a discussion of this).

One should not underestimate the power of Theorem 4.1. It is an if and only if theorem. Not only does it give a sufficient condition ($v \in B_\infty^s(L_p(\Omega))$) for approximation of v to be bounded by Mn^{-s} , it also says that if v does not satisfy this smoothness condition then there is no hope in achieving this approximation order.

How do we utilize this theorem in our search for the number s_L for our solution u to the PDE. What we need to determine is the maximum value of s for which u lies in the Besov space $B_\infty^s(L_p(\Omega))$. This has a simple interpretation in our picture of smoothness spaces. We fix the point $(1/p, 0)$ (which corresponds to L_p) and consider the vertical line passing through this point which therefore consists of all points of the form $(1/p, s)$. We search along this line for the maximum value of s , such that u is in the corresponding Besov space. This determines s_L (see Figure 4.2).

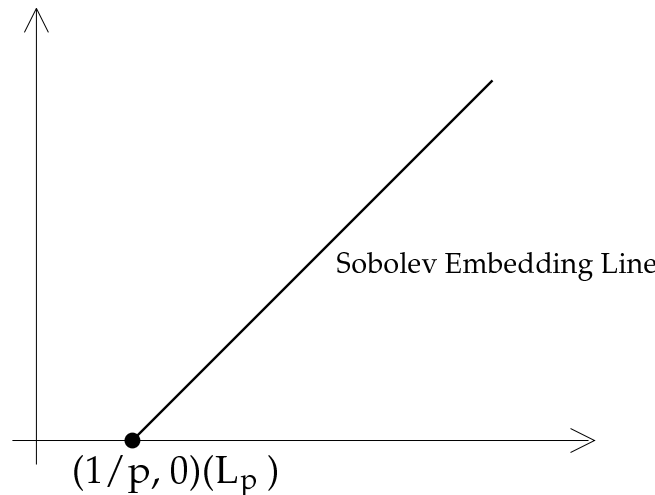


Figure 4.2: Graphical interpretation of linear and nonlinear approximation.

It is the role of regularity theorems for PDEs to provide us with the answer to this question. We shall return to this topic later in this talk after we have introduced the concepts needed to determine s_{NL} for nonlinear approximation.

5 Nonlinear methods

Let us now consider nonlinear numerical methods for recovering the solution u . In this case, the numerical method no longer generates an approximation from a linear space (prescribed in advance) but rather a nonlinear manifold Σ_n where the dimension (number of parameters) of Σ_n is of order n . Similar to the linear case described above we can define the ideal error

$$(5.1) \quad \sigma_n(v) := \inf_{g \in \Sigma_n} \|v - g\|$$

where $\|\cdot\|$ is the norm we have chosen to measure error.

Given our target function u , we define $s_{NL} > 0$ to be the real number such that

$$(5.2) \quad \sigma_n(u) \leq Mn^{-s}, \quad n \rightarrow \infty,$$

for all $s < s_{NL}$ and

$$(5.3) \quad \limsup_{n \rightarrow \infty} n^s \sigma_n(u) = \infty$$

for all $s > s_{NL}$. Approximation theory will tell us the conditions on u which determine s_{NL} .

Let us begin with the wavelet case in which the results are very precise. An adaptive wavelet method generates an approximation to u from the set Σ_n which consists of all functions S of the form

$$(5.4) \quad S = \sum_{\lambda \in \Lambda} c_\lambda \psi_\lambda$$

where the cardinality $\#\Lambda$ of Λ is $\leq n$. This is called *n-term approximation* and recent results [11] (see also [5]) characterize its approximation properties. To describe these, we introduce the sequence $(c_n^*(v))_{n=1}^\infty$ which is the nondecreasing rearrangement of $(|c_\lambda(v)|)_{\lambda \in \Delta}$. In other words, $c_n^*(v)$ is the n -th largest of the $|c_\lambda(v)|$, $\lambda \in \Delta$. The following theorem characterizes the functions which can be approximated to rate $O(n^{-s})$ by n -term wavelet approximation.

Theorem 5.1 *Let $0 < p < \infty$. A function $v \in L_p(\Omega)$ ($v \in H_p(\Omega)$ if $p \leq 1$) satisfies*

$$(5.5) \quad \sigma_n(v)_p \leq M(v)n^{-s}$$

if and only if the rearranged wavelet coefficients of v satisfy

$$(5.6) \quad c_n^*(v) \leq M'(v)n^{-s-1/p}, \quad n = 1, 2, \dots,$$

for some constant $M'(v)$. Moreover, the smallest constant $M(v)$ in (5.5) is equivalent to the smallest constant $M'(v)$ satisfying (5.6).

Let us denote by X^s the space of functions v satisfying (5.6). This is not a classical smoothness space (Besov or Sobolev space) but it is very close to requiring that v has s orders of smoothness in $L_{\tau(s)}$ with $\tau(s) := (s/d + 1/p)^{-1}$. In fact, we have the embeddings

$$(5.7) \quad B_{\tau(s)}^s(L_{\tau(s)}(\Omega)) \subset X^s \subset B_{\tau(s-\epsilon)}^{s-\epsilon}(L_{\tau(s-\epsilon)}(\Omega)),$$

where $\epsilon > 0$ is arbitrary. There are also more precise connections between approximation rates and Besov spaces (see [10]).

Theorem 5.1 is enough for us to describe how to determine s_{NL} for our target function u . We look at the scale of Besov spaces $B_{\tau(s)}^s(L_{\tau(s)}(\Omega))$, $\tau(s) = (s/d + 1/p)^{-1}$, $s > 0$. These spaces are nested and strictly decrease as s increases. Then s_{NL} is the supremum of all the s for which u is in $B_{\tau(s)}^s(L_{\tau(s)}(\Omega))$.

This all has a nice interpretation in our picture of smoothness spaces (see Figure 4.2). We fix the L_p norm in which we shall measure the error. This identifies the point $(1/p, 0)$. The spaces $B_{\tau(s)}^s(L_{\tau(s)}(\Omega))$ all live on the line emanating from $(1/p, 0)$ with slope d . Thus, starting at

$(1/p, 0)$ we move on this line as long as u is in the corresponding Besov space and stop when u is not in this space. This identifies the number s_{NL} .

We shall not give a precise formulation of the corresponding results for adaptive Finite Element Methods except to say that the story is roughly the same as in the wavelet case. Such adaptive methods begin with a polyhedral domain Ω and an initial decomposition \mathcal{P}_0 of Ω into simplicial cells T . The adaptive procedure begins with the initial triangulation \mathcal{P}_0 and iteratively refines simplicies. Thus, at the first iteration we generate a partition \mathcal{P}_1 which is obtained from \mathcal{P}_0 by refining some of the simplices in \mathcal{P}_0 and not others. In general \mathcal{P}_{j+1} is gotten from \mathcal{P}_j in the same way. If the adaptive strategy (i.e. the selection of simplicies which are to be subdivided) is chosen correctly and if the resulting space of piecewise polynomials allows for good local bases then it is possible to prove that whenever u is in the Besov space $B_{\tau(s)}^s(L_{\tau(s)})$, it will be approximated with the efficiency $O(n^{-s/d})$ with n the number of simplices in the resulting partition. Inverse estimates can be proven if the refinement strategy guarantees the shape preserving property of the simplicies in \mathcal{P}_j for each j . For precise formulations of the above and for details we refer to the forthcoming paper [1].

6 The theory in action

We have seen that to determine whether it is beneficial to use nonlinear methods to approximate our target function u we need to determine the two numbers s_L and s_{NL} associated to u and check whether $s_{NL} > s_L$. We do this by checking the regularity of u in the two scales of smoothness spaces associated to linear and nonlinear approximation. A result which determines the regularity of u in one of these scales is called a regularity theorem for PDE's. A typical regularity theorem infers the smoothness of the solution u to a PDE from information in the PDE such as the coefficients, inhomogeneous term, initial conditions, or boundary conditions.

To illustrate how this theory plays out in specific settings, we shall consider two model problems; one hyperbolic and the other elliptic.

6.1 Conservation laws

Consider the scalar univariate conservation law

$$(6.1) \quad \begin{cases} u_t + f(u)_x = 0 & x \in \mathbb{R}, t > 0, \\ u(x, 0) = u_0(x) & x \in \mathbb{R}, \end{cases}$$

where f is a given flux, u_0 a given initial condition which will assume is of compact support, and u is the sought after solution. This is a well-studied nonlinear transport equation with transport velocity $a(u) = f'(u)$. We shall assume that the flux is strictly convex which means the transport velocity is strictly increasing. The important fact for us is that, even when the initial condition u_0 is smooth, the solution $u(\cdot, t)$ will develop spontaneous shock discontinuities at later times t .

The proper setting for the analysis of conservation laws is in L_1 and in particular the error of numerical methods should be measured in this space. Thus, concerning the performance of linear numerical methods, the question arises as to the possible values of the smoothness parameter s_L

of $u(\cdot, t)$ as measured in L_1 . It is known that if the initial condition u_0 is in BV, then the solution u remains in this space for all later time $t > 0$ (note that $\text{BV} \subset B_\infty^1(L_1(\mathbb{R}))$). However, since, for any initial condition, this solution develops discontinuities, the Sobolev embedding theorem precludes u being in any Besov space $B_\infty^s(L_1)$ for any $s > 1$. This means that the largest value we can expect for s_L is $s_L = 1$ and we get this value whenever $u_0 \in \text{BV}$. Thus, the optimal performance, we can expect from linear methods of approximation is $O(n^{-1})$ with n the dimension of the linear spaces used in the approximation. Typical numerical methods utilize spaces of piecewise polynomials on a uniform mesh with mesh length h and the above remarks mean that the maximum efficiency we can expect for such numerical methods is $O(h)$, $h \rightarrow 0$. In reality, the best proven estimates are $O(\sqrt{h})$ under the assumption that $u_0 \in \text{BV}$. This discrepancy between the possible performance of numerical algorithms and the actual performance is not unusual. The solution is known to have sufficient regularity to be approximated, for example, by piecewise constants with uniform mesh h to accuracy $O(h)$ but algorithms which capture this accuracy are generally not known.

To understand the possible performance of nonlinear methods such as moving grid methods, we should estimate the smoothness of the solution in the nonlinear Besov scale $B_{\tau(s)}^s(L_{\tau(s)})$, $\tau(s) := (s + 1)^{-1}$, corresponding to approximation in the L_1 -norm. A rather surprising result of DeVore and Lucier [12] shows that starting with any initial condition u_0 of bounded variation which is in this space, the solution u will remain in this Besov space for all later time $t > 0$. In particular, if u_0 is C^∞ with compact support then this means that nonlinear methods such as moving grid methods could provide arbitrarily high efficiency. In fact, such algorithms, based on piecewise polynomial approximation, can be constructed using the method of characteristics (see Lucier [14] for the case of piecewise linear approximation).

In summary, whenever the initial condition u_0 is of bounded variation and in the smoothness space $B_{\tau(s)}^s(L_{\tau(s)})$ with $s > 1$, then the use of adaptive methods is justified since $s_{NL} > s_L$. In particular, if u_0 is of bounded variation and in C^∞ then $s_{NL} = \infty$ while $s_L = 1$.

6.2 Elliptic equations

An extensive accounting of the role of linear and nonlinear approximation in the solution of elliptic problems is given in Dahmen [8] and Dahlke, Dahmen, and DeVore [6]. We shall therefore limit ourselves to reiterating a couple of important points about the role of regularity theorems and the form of nonlinear estimates. We consider the model problem

$$(6.2) \quad \begin{aligned} \Delta u &= f & \text{on } \Omega \subset \mathbb{R}^d, \\ u &= 0 & \text{on } \partial\Omega. \end{aligned}$$

of Laplace's equation on a domain $\Omega \subset \mathbb{R}^d$ with zero boundary conditions. This equation is closely related to the Dirichlet problem for harmonic functions on Ω :

$$(6.3) \quad \begin{aligned} \Delta v &= 0 & \text{on } \Omega \subset \mathbb{R}^d, \\ v &= g & \text{on } \partial\Omega. \end{aligned}$$

We shall also limit our discussion to estimating error in the L_2 -norm. These results extend trivially to approximation in the Sobolev space $H^\alpha = W^\alpha(L_2(\Omega))$ and in particular to the case $\alpha = 1$

which is equivalent to the energy norm for (6.2). There are also various results known for general p [13].

Consider first the case where $f \in L_2(\Omega)$ and Ω has a smooth boundary. Then, the solution u to (6.2) has smoothness $W^2(L_2(\Omega))$ and can therefore be approximated by linear spaces of piecewise polynomials of dimension n to accuracy $O(n^{-2/d})$. This accuracy can be obtained by using standard FEM with uniformly refined partitions.

If the boundary $\partial\Omega$ of Ω is not smooth then the solution u to (6.2) has singularities due to corners or other nonsmoothness of the boundary $\partial\Omega$. For example for Laplace's equation on a general Lipschitz domain, we can only expect that the solution u is in the Sobolev space $W^{3/2}(L_2(\Omega))$. Thus, in general, we can at most expect $s_L = 3/2$.

Because of the appearance of singularities due to the boundary, adaptive numerical techniques are suggested for numerically recovering the solution u . We understand that to justify the use of such methods, we should determine the regularity of the solution in the scale of Besov spaces $B_{\tau(s)}^s(L_{\tau(s)}(\Omega))$, $\tau(s) := (s + 1/2)^{-1}$. Such regularity has been studied by Dahlke and DeVore [7]. They prove, among other things, that for any Lipschitz domain the nonlinear smoothness s_{NL} associated to u always exceeds the linear smoothness. Namely, $s_{NL} \geq \frac{3d}{2(d-1)}s_L$. In other words, the use of nonlinear adaptive methods for numerically recovering the solution u to (6.2) is theoretically justified.

7 An adaptive algorithm for elliptic problems

Up to this point, we have not discussed the properties of any specific numerical algorithm but rather have addressed the question of whether nonlinear or adaptive algorithms could possibly be of benefit in numerically approximating the solution of a PDE. Even if we have decided that an adaptive method should be of use, there remains the problem of constructing an adaptive algorithm which exhibits the expected performance. This is indeed a nontrivial task. We shall close this talk by discussing the recent wavelet based adaptive algorithm given in [4] which has been proven to exhibit optimal performance in the sense of providing the best allowable rate of approximation to u .

7.1 The setting

Let Ω be a domain (or manifold) in \mathbb{R}^d and let A be a linear operator mapping H into H^* where H is a subspace with the property that either H or its dual H^* is embedded in $L_2(\Omega)$. The operator A induces the bilinear form a defined on $H \times H$ by

$$(7.1) \quad a(u, v) := \langle Au, v \rangle,$$

where $\langle \cdot, \cdot \rangle$ denotes the (H^*, H) duality product. We assume that the bilinear form a is symmetric positive definite and *elliptic* in the sense that

$$(7.2) \quad a(v, v) \sim \|v\|_H^2, \quad v \in H.$$

It follows that H is a pre-Hilbert space with respect to the inner product a and that this inner product induces a norm (called the energy norm) on H by

$$(7.3) \quad \|\cdot\|_a^2 := a(\cdot, \cdot).$$

The energy norm is equivalent to $\|\cdot\|_H$. By duality, A thus defines an isomorphism from H onto H^* .

We are interested in numerically recovering the solution u to the elliptic equation

$$(7.4) \quad Au = f$$

with $f \in H^*$. It follows that u is also the unique solution of the variational equation

$$(7.5) \quad a(u, v) = \langle f, v \rangle, \quad \text{for all } v \in H.$$

The typical examples included in the above assumptions are the Poisson or the biharmonic equations on bounded domains in \mathbb{R}^d ; single or double layer potentials and hypersingular operators on closed surfaces arising in the context of boundary integral equations. In these examples H is a Sobolev space, e.g. $H = H_0^1(\Omega)$, $H_0^2(\Omega)$, or $H = H^{-1/2}(\Omega)$ (see [8]).

The numerical methods developed in [4] require the existence of a biorthogonal wavelet basis Ψ for Ω . The wavelets in Ψ are in H , whereas those in the dual basis $\tilde{\Psi}$ are in H^* . Thus, each $v \in H$ has a wavelet expansion $v = \mathbf{d}^T \Psi$ (with coordinates $d_\lambda = \langle v, \tilde{\psi}_\lambda \rangle$). We assume that

$$(7.6) \quad \|\mathbf{D}^{-1} \mathbf{d}\|_{\ell_2(\nabla)} \sim \|\mathbf{d}^T \Psi\|_H.$$

with \mathbf{D} a fixed positive diagonal matrix. Observe that (7.6) implies that $\mathbf{D}_{\lambda, \lambda} \sim \|\psi_\lambda\|_H^{-1}$, and that Ψ (resp. $\mathbf{D}^{-1} \Psi$) is an unconditional (resp. Riesz) basis for H . By duality, one easily obtains that each $v \in H^*$ has a wavelet expansion $v = \mathbf{d}^T \tilde{\Psi}$ (with coordinates $d_\lambda = \langle v, \psi_\lambda \rangle$) that satisfies

$$(7.7) \quad \|\mathbf{D} \mathbf{d}\|_{\ell_2(\nabla)} \sim \|\mathbf{d}^T \tilde{\Psi}\|_{H^*}.$$

We also assume that the wavelet bases Ψ and $\tilde{\Psi}$ provide characterizations of Besov and Sobolev spaces (as described earlier) for a suitable range of the smoothness parameter. In the context of elliptic equations, H is typically some Sobolev space $H^t = W^t(L_2(\Omega))$. In this case the above assumptions are satisfied whenever the wavelets are sufficiently smooth, with $\mathbf{D}_{\lambda, \lambda} = 2^{-|\lambda|t}$. For instance, when $A = -\Delta$, one has $t = 1$.

If we write the unknown solution u and the right hand side f in terms of their wavelet bases we obtain an infinite system of equations. After preconditioning using the matrix \mathcal{D} , we obtain from (7.5) the system of equations:

$$(7.8) \quad \mathbf{D} \langle A \Psi, \Psi \rangle^T \mathbf{D} \mathbf{D}^{-1} \mathbf{d} = \mathbf{D} \langle f, \Psi \rangle^T,$$

or more compactly,

$$(7.9) \quad \mathbf{A} \mathbf{u} = \mathbf{f},$$

where

$$(7.10) \quad \mathbf{A} := \mathbf{D} \langle A \Psi, \Psi \rangle^T \mathbf{D}, \quad \mathbf{u} := \mathbf{D}^{-1} \mathbf{d}, \quad \mathbf{f} := \mathbf{D} \langle f, \Psi \rangle^T \in \ell_2(\nabla).$$

The matrix \mathbf{A} is symmetric positive definite.

One can show that in all classical settings for elliptic problems, the matrix \mathbf{A} satisfies certain sparsity conditions. These arise from the fact that

$$(7.11) \quad 2^{-(|\lambda'|+|\lambda|)t} |\langle A\psi_{\lambda'}, \psi_{\lambda} \rangle| \lesssim 2^{-\|\lambda\|-|\lambda'|\sigma} (1 + d(\lambda, \lambda'))^{-\beta},$$

with $\sigma > d/2$ and $\beta > d$ and

$$(7.12) \quad d(\lambda, \lambda') := 2^{\min(|\lambda|, |\lambda'|)} \text{dist}(\text{supp}(\psi_{\lambda}), \text{supp}(\psi_{\lambda'})).$$

We refer the reader to [8] for a discussion of the various settings in which (7.11) is known to be valid.

7.2 The numerical method

To numerically resolve (7.4), we use the Galerkin method. We fix a finite set Λ of wavelet indices and approximate u from the space $S_{\Lambda} := \text{span}(\psi_{\lambda} : \lambda \in \Lambda)$. The approximate Galerkin solution u_{Λ} from S_{Λ} is defined by the conditions

$$(7.13) \quad a(u_{\Lambda}, v) = \langle f, v \rangle_{L_2(\Omega)}, \quad v \in S_{\Lambda}.$$

In matrix form, this is equivalent to solving the finite matrix problem

$$(7.14) \quad \mathbf{A}_{\Lambda} \mathbf{u}_{\Lambda} = \mathbf{f}_{\Lambda}$$

where \mathbf{A}_{Λ} is the finite section of \mathbf{A} gotten by choosing the rows and columns of \mathbf{A} corresponding to Λ , \mathbf{u}_{Λ} is the unknown vector (which determines the wavelet coefficients of u_{Λ}), and \mathbf{f}_{Λ} is the vector obtained by restricting \mathbf{f} to Λ .

The numerical method studied in [4] proceeds as follows. It starts with an initial set Λ_0 of wavelet indices (one may take $\Lambda_0 = \emptyset$) and given that a set Λ_j has been chosen, it generates a new set Λ_{j+1} with hopefully better approximation properties.

Let us describe the two main steps for determining the set $\Lambda' = \Lambda_{j+1}$ from $\Lambda := \Lambda_j$ which work from the discrete equations (7.9). Let \mathbf{u}_{Λ} be the current vector solution to (7.14). We can view \mathbf{u}_{Λ} as a vector defined for all $\lambda \in \nabla$ by defining $\mathbf{u}_{\lambda} = 0$, $\lambda \in \nabla \setminus \Lambda$. Then the residual $\mathbf{r}_{\Lambda} := \mathbf{A}(\mathbf{u} - \mathbf{u}_{\Lambda}) = \mathbf{f} - \mathbf{A}\mathbf{u}_{\Lambda}$ has norm

$$(7.15) \quad \|\mathbf{r}_{\Lambda}\|_{\ell_2(\nabla)}$$

that can, by the ellipticity assumptions, be related to the function error $\|u - u_{\Lambda}\|_H$. Note that \mathbf{r}_{Λ} vanishes on Λ . The set Λ' is obtained as follows. First, we enlarge Λ to a set $\tilde{\Lambda}$ containing Λ by adjoining the wavelet indices where \mathbf{r}_{Λ} is large. We adjoin a finite set of vertices $\tilde{\Lambda} \setminus \Lambda$ so that this set captures at least half of the energy of \mathbf{r}_{Λ} . We next solve the Galerkin problem on the new set $\tilde{\Lambda}$ resulting in the new vector $\mathbf{u}_{\tilde{\Lambda}}$. We examine the entries in $\mathbf{u}_{\tilde{\Lambda}}$ and put into Λ' only those indices whose keep are sufficiently large. This step can be viewed as thresholding the entries in \mathbf{u}_{Λ} .

In practice, the algorithm is implemented by choosing an $\epsilon > 0$ and an initial set Λ_0 and generating sets Λ_j , $j = 1, 2, \dots$, until the error tolerance ϵ is guaranteed. Note that the error at any given stage is upper bounded by a fixed multiple of the norm of the residual.

7.3 Performance of the algorithm

The results in [4] show that the above algorithm has optimal performance in the following sense. Suppose that the solution u to (7.4) can be approximated (in the energy norm) with n wavelets terms to accuracy

$$(7.16) \quad \sigma_n(u) \leq Mn^{-s}, \quad n = 1, 2, \dots,$$

with M an absolute constant. Then, for each $j \geq 1$, the above numerical algorithm will generate an approximation u_{Λ_j} with

$$(7.17) \quad \|u - u_{\Lambda_j}\|_a \leq CMN_j^{-s}$$

with $N_j := \#\Lambda_j$ the cardinality of Λ_j . Moreover, the number of arithmetic operations necessary to find Λ_j and to compute u_{Λ_j} will not exceed CN_j . The number of sorting operations necessary in the thresholding portion of the above algorithm does not exceed $CN_j \log N_j$.

The proof of this result is nontrivial and we shall only mention a few of the key ingredients in the proof in the following remarks.

Remark 1. Capturing at least half of the energy in the residual \mathbf{r}_Λ guarantees that the new Galerkin solution on $\tilde{\Lambda}$ reduces the error by a fixed factor $\eta < 1$:

$$(7.18) \quad \|u - u_{\tilde{\Lambda}}\|_H \leq \eta \|u - u_\Lambda\|_H.$$

This result would lead, in and of itself (without thresholding), to a convergent algorithm but would not sufficiently control the number of entries in the sets $\tilde{\Lambda}_j$.

Remark 2. In numerical implementation of the algorithm, it is necessary to limit the search for the entries which need to be adjoined to Λ in order to obtain $\tilde{\Lambda}$. Here the sparseness of the matrix \mathbf{A} plays a crucial role.

Remark 3. The thresholding step when employed with the correct threshold reduces the number of elements used in the approximation without seriously effecting the error. This is proved by establishing a general result on thresholding.

Remark 4. To bound the number of arithmetic operations requires fast methods for multiplying a sparse matrix \mathbf{B} (in our case $\mathbf{B} = \mathbf{A}$) with a sparse vector \mathbf{v} (in our case $\mathbf{v} = \mathbf{u}_\Lambda$) (see [4] for the interesting method to do this).

References

- [1] P. Binev, R. DeVore, and P. Petrushev, Adaptive approximation using piecewise polynomials, in preparation
- [2] C. de Boor and R. DeVore, Partitions of unity and approximation PAMS, **93** (1985), 705-708.
- [3] A. Cohen, Wavelet methods in Numerical Analysis, to appear in the Handbook of Numerical Analysis, vol. VII, 1998.
- [4] A. Cohen, W. Dahmen, and R. DeVore, Adaptive wavelet methods for elliptic operator equations: convergence rates, to appear in Math. Comp.
- [5] A. Cohen, R. DeVore, and R. Hochmuth Restricted approximation, Constructive Approximation, **16** (2000), 85–113.

- [6] S. Dahlke, W. Dahmen, and R. DeVore, Nonlinear approximation and adaptive techniques for solving elliptic equations, in: *Multiscale Techniques for PDEs*, W. Dahmen, A. Kurdila, and P. Oswald (eds), Academic Press, 1997, San Diego, 237–284.
- [7] S. Dahlke and R. DeVore, Besov regularity for elliptic boundary value problems, *Communications in PDEs*, **22**(1997), 1–16.
- [8] W. Dahmen, Wavelet and multiscale methods for operator equations, *Acta Numerica*, **6**(1997, Cambridge University Press, , 55–228.
- [9] I. Daubechies, *Ten Lectures on Wavelets*, CBMS-NSF Regional Conference Series in Applied Mathematics, **61**, SIAM Philadelphia, 1988.
- [10] R. DeVore, Nonlinear approximation, *Acta Numerica* **7** (1998), 51-150.
- [11] R. DeVore, B. Jawerth and V. Popov, Compression of wavelet decompositions, *Amer. J. Math.*, **114** (1992), 737–785.
- [12] R. DeVore and B. Lucier, High order regularity for conservation laws, *Indiana Journal of Math.*, **39** (1990), 413–430 .
- [13] Jerison and Kenig, The inhomogeneous Dirichlet problem in Lipschitz domains, *J. of Functional Analysis*, **130**(1995), 161–219.
- [14] B. Lucier, Regularity through approximation for scalar conservation laws, *SIAM J. Math. Analysis*, **19**(1998), 763–773.
- [15] Y. Meyer, *Ondelettes et Operateurs, Vol 1 and 2*, Hermann, Paris, 1990

Ronald A. DeVore

Department of Mathematics

University of South Carolina

Columbia, SC 29208

USA

e-mail: devore@math.sc.edu

<http://www.math.sc.edu/~devore/>