Adaptive wavelet techniques in Numerical Simulation

Albert Cohen[1], Wolfgang Dahmen[2], Ronald DeVore[3]

[1] *Laboratoire d'Analyse Numerique, Universite Pierre et Marie Curie, 175 Rue du Chevaleret, 75013 Paris, France*
[2] *Institut für Geometrie und Praktische Mathematik, RWTH Aachen, 52056 Aachen, Germany*
[3] *Department of Mathematics, University of South Carolina, Columbia, SC 29208, U.S.A.*

ABSTRACT

This chapter highlights recent developments concerning adaptive wavelet methods for time dependent and stationary problems. The first problem class focusses on hyperbolic conservation laws where wavelet concepts exploit sparse representations of the conserved variables. Regarding the second problem class, we begin with matrix compression in the context of boundary integral equations where the key issue is now to obtain sparse representations of (global) operators like singular integral operators in wavelet coordinates. In the remainder of the chapter a new fully adaptive algorithmic paradigm along with some analysis concepts are outlined which, in particular, works for nonlinear problems and where the sparsity of both, functions and operators, is exploited.

KEY WORDS: *Conservation laws, boundary integral equations, elliptic problems, saddle point problems, mixed formulations, nonlinear problems, matrix compression, adaptive application of operators, best N-term approximation, tree approximation, convergence rates, complexity estimates*

1. Introduction

Increasingly realistic models in computational mechanics and the search for more and more accurate simulations place continuously growing demands on computation that surpass the ongoing increase of computing power. Thus, paradoxically, these finer models might be of limited use in the absence of new computational strategies. One promising emerging strategy is to dynamically adapt discretizations in the course of the computational solution process. Adaptive strategies of this type have been observed to reduce the complexity of computational problems arising in large scale numerical simulation. Therefore adaptivity provides an enormeous potential for advancing the frontiers of computability. By bringing more and more complex tasks into reach it offers in the long run better and better access to physical phenomena through a powerful numerical microscope. On the other hand, to advance these techniques to their natural fruition requires an understanding of the power of adaptivity vis a vis traditional methods of computation. This includes clarifying the optimal performance that can be expected from adaptive methods and how this compares with the performance using non-adaptive techniques.

This chapter describes adaptive numerical strategies in the context of multiscale decompositions using wavelet bases. In addition to formulating adaptive strategies to be used

Encyclopedia of Computational Mechanics. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

in a variety of settings, the chapter will provide an a priori analysis of the computational efficiency of adaptive methods. This will dilineate the advantages of adaptive strategies versus standard computational methods.

Adaptivity takes a variety of forms distinguished by their principal goals. In many applications one is not interested in the full solution of given problem but only in certain local functionals of an object that may be globally defined like the solution of a boundary value problem. In this case an adaptive discretization reflects how much has to be paid to the global character of the object when trying to recover local information about it. However, this is not the direction of the present chapter. Instead, this chapter focuses on recovering the *whole* object in question. In the context of fluid mechanics, this may mean recovering the vortices in the wake of an airfoil, or the interaction of shocks even at some distance of the airfoil, or the recovery of a full stress field or, eventually, to understand more about developing turbulence. The objective is to develop numerical techniques that are able to extract information within desired error tolerances at minimal cost. This means that searched for quantities like pressure or velocity are to be recovered within some accuracy tolerance, e.g. with respect to some norm. This should be done at the expense of a number of degrees of freedom that remains proportional to the minimal number of degrees of freedom (in a certain discretization framework) needed to approximate the object based on full information within the desired target accuracy. From the mathematical point of view, it is not clear beforehand at all whether this is possible solely based on a–posteriori information acquired during a solution process. We shall indicate in this chapter an affirmative answer for a wide range of problems arising in engineering applications, see Cohen, Dahmen and DeVore, 2001; Cohen, Dahmen and DeVore, 2002a; Cohen, Dahmen and DeVore, 2002b; Cohen, Dahmen and DeVore, 2002c.

Our approach involves expansions of functions into *wavelet bases*. In such expansions, the wavelet coefficients encode *detail information* that has to be added when progressing to higher levels of resolution of the underlying function. These coefficients convey local structural information such as the regularity of the expanded function. The decomposition naturally breaks the function into different characteristic length scales. A central question in many dynamical simulation tasks concerns the interaction of these different scales. As we shall show, wavelet analysis offers a promising way to describe the behavior of contributions from different length scales under *nonlinear mappings*. We shall see that wavelet expansions offer quantitative ways of estimating nonlinear effects that appear for example in the Navier Stokes equations. Moreover, we shall point out how such an analysis aids the adaptive solution process. This already indicates the close marriage between the analysis and numerical resolution of a problem facilitated by wavelet concepts. Therefore the understanding of these concepts and their potential requires a certain amount of functional analysis as will be described in this chapter.

We do not attempt to give an exhaustive overview of wavelet analysis pertinent to computational mechanics issues. Nor will the topics presented here be treated in a selfcontained way. Both would be far beyond the scope of this chapter. Rather we shall focus on presenting some concepts and ideas which in our opinion best reflect the potential of wavelets, thereby offering some orientation that could be complemented by the extensive list of references. The following surveys and text books are recommended as sources of more detailed expositions Cohen, 2000; Cohen, 2003; Dahmen, 1997; Dahmen, 2001; DeVore, 1998.

The organization of the material is in some sense "two dimensional". Most simulation tasks are based on continuous mathematical models formulated in terms of integral or (partial)

Encyclopedia of Computational Mechanics. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

differential equations. The "first dimension" is to group the different concepts with respect to the following two major problem classes. The first one concerns *evolution problems*

$$\partial_t u = \mathcal{E}(u) \tag{1}$$

together with initial and boundary conditions. The second class concerns stationary problems

$$\mathcal{R}(u) = 0, \tag{2}$$

which are usually given in variational form. The scope of problems covered by (2) will be illustrated by a list of examples including mixed formulations and nonlinear problems. Of course, there is no clear dividing line. For instance, an implicit time discretization of a parabolic evolution problem leads to a family of problems of the type (2). The problems grouped under (2) are typically elliptic (in the sense of Agmon–Douglis–Nirenberg) for which *Hilbert space* methods are appropriate. In contrast, we focus under (1) on nonlinear *hyperbolic* problems. It will be seen that the respective concepts are quite different in nature. The nature of the relevant function spaces, e.g. L_1 which admits no unconditional basis, causes an impediment to exploiting the full potential of wavelets.

The "second dimension" of organization concerns the way wavelet features are exploited. In Section 2 we review briefly the main features that drive wavelets as analysis and discretization tools. Aside from transform mechanisms these are the *locality* (in physical and frequency domain), the *cancellation properties* and the *norm equivalences* between function and sequence spaces. The latter facilitate a stable coupling of the continuous and the discrete world. Together with the first two features this is also fundamental for a fast numerical processing.

In Section 3 these features are applied to (1). The primary focus of adaptivity here is the *sparse approximation* of the unknown solution, mainly thanks to the cancellation properties. In this context wavelets are *not* used as stand alone tools but are rather combined with conventional finite volume discretizations. The numerical approximation represented by arrays of cell averages is compressed in a manner similar to image compression. This amounts to a *perturbation* analysis where one seeks a significant data compression while preserving essentially the accuracy of the underlying reference discretization for a *fixed* level of resolution. The approach and the performance of such schemes are illustrated by some numerical examples concerning aerodynamical applications.

The remainder of the chapter is concerned with the problem class (2). Section 4 deals with *global operators* represented here by the classical boundary integral equations. Now the above mentioned main features of wavelets are used mainly to obtain *sparse approximations of operators*. This time the elliptic nature of the problem allows one to formulate stable Galerkin discretizations. When using finite elements or boundary elements the resulting stiffness matrices are densely populated and, depending on the operator, are increasingly ill–conditioned when the mesh size decreases. In wavelet formulations, the norm equivalences and cancellation properties are used to show that the stiffness matrices can be replaced by sparse well–conditioned ones without sacrificing discretization error accuracy. This allows one to solve such problems in linear time. Again this is essentially a perturbation approach where this time sparse approximations apply to the operator not to the function. Adaptivity refers here primarily to the quadrature used to compute the compressed stiffness matrices with a computational effort that stays proportional to the problem size. As for the current state of the art we refer to Dahmen, Harbrecht and Schneider, 2002; Harbrecht, 2001 and the references cited there.

Encyclopedia of Computational Mechanics. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

In Section 5 we introduce a *new algorithmic paradigm* that emerges form exploiting both, the sparse approximation of functions and the sparse representation of operators together. It aims at intertwining in some sense the analysis and resolution aspects of wavelet concepts as much as possible. Here are the main conceptual pillars of this approach:

A transform point of view: Many studies of wavelet methods for the numerical solution of PDEs are very similar in spirit to classical finite element discretizations where the trial spaces are spanned by *finite* collections of wavelets. This has so far dominated the use of wavelets in the context of boundary integral equations and is the point of view taken in Section 4. However, this does not yet fully exploit the potential of wavelets. In fact, similar to classical Fourier methods, wavelets can be used to formulate *transform methods* that are best explained in the context of variational formulations of (linear or nonlinear) operator equations like boundary value problems or boundary integral equations. Unlike finite volume or finite element schemes wavelet bases can be used to transform the original variational problem into an *equivalent* problem over ℓ_2 the space of square summable sequences indexed by the wavelet basis. Moreover, when the wavelets are correctly chosen in accordance with the underlying problem, the transformed (still infinite dimensional) problem is now well-posed in a sense to be made precise later. We shall point out now the main principles along these lines.

Staying with the infinite dimensional problem: In many cases, the underlying infinite dimensional problem, e.g. a PDE, is fairly well understood. In mathematical terms this means, when formulated as an operator equation, the operator is known to be boundedly invertible as a mapping from a certain function space into its dual, which is another way of saying that the problem is well posed in a certain topology - there exists a unique solution which depends continuously on the data in the topology given by that function space. When transforming to the wavelet domain, the properties of the operator are inherited by the transformed operator which now acts on sequence spaces. The main point we wish to stress is that the original infinite dimensional problem is often better understood than specific discretized finite dimensional versions and therefore there is an advantage in delaying as long as possible the movement to finite discretizations. A classical example of this is the Stokes problem where a positive definite quadratic functional is minimized under the divergence constraint and thus has *saddle point* character. The Stokes problem is well-posed in the above sense for the right pairs of functions spaces for the velocity and pressure component, see e.g. Brezzi and Fortin, 1991; Girault and Raviart, 1986. It is well-known that Galerkin discretizations, however, may very well become unstable unless the trial spaces for velocity and pressure satisfy a compatibility condition called the Ladyšhenskaya-Babuška-Brezzi (LBB) condition. For the Stokes problem this is well understood but in other situations, as in many physically very appropriate *mixed formulations*, finding stable pairs of trial spaces is a more delicate task. So in some sense one may run into self-inflicted difficulties when turning to finite discretizations even though the original problem is well behaved. Is there an alternative?

Stabilizing effects of adaptivity: The very fact that, unlike conventional schemes, a suitable wavelet basis captures the complete infinite dimensional problem and puts it into a well conditioned format over ℓ_2, can be used to avoid fixing any finite dimensional dicretization. Instead the well-posedness offers ways of formulating an iterative scheme for the *full infinite dimensional problem* that converges (conceptually) with a fixed error reduction per step. Only after this infinite dimensional analysis is complete do we enter the numerical stage by applying the involved infinite dimensional (linear and also nonlinear) operators *adaptively*

Encyclopedia of Computational Mechanics. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

within suitable stage dependent dynamically updated error tolerances. Roughly speaking, this numerical approach inherits the well-posedness of the original problem and allows us to avoid imposing compatibility conditions such as LBB.

Adaptive evaluation of operators: A central issue is then to actually realize concrete adaptive evaluation schemes for relevant operators and to analyze their computational complexity. We shall engage this issue for both linear and nonlinear examples. While at the first glance nonlinear operators seem to interfere in an essential way with wavelet concepts (as they do with regard to the Fourier techniques) we claim that they offer particularly promising perspectives in this regard. In conventional discretizations the image of a current adaptive approximation under a nonlinear mapping is usually discretized on the same mesh. However, a singularity, which may cause an adaptive local refinement, is often severely affected by a nonlinearity so that this mesh might no longer be optimal. In the present framework the adaptive evaluation will be seen to generate at each stage the right choice of degrees of freedom for the image of the nonlinearity, see 6.2. This is based on quantitative estimates on the interaction of different length scales under nonlinear mappings.

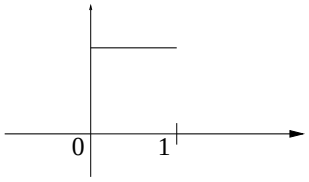
It would be far beyond the scope of this chapter to address any of the above issues in complete detail. Instead, our presentation will be more of an overview of this subject which should serve to orient the reader to the essential concepts and point of views. The interested reader will then find extensive references for further reading.

2. Wavelets

In this section we give a brief overview of those features of wavelets and multiresolution that are important for our presentation. There are many different ways of viewing and motivating wavelet expansions, see e.g. Daubechies, 1992. Our point of view in the present context is conveniently conveyed by the following example.

2.1. The Haar Basis

The starting point is the box function $\phi(x) = \chi_{[0,1)}(x)$, which takes the value one on $[0,1)$ and zero outside. The normalized dilates and translates $\phi_{j,k} = 2^{j/2}\phi\left(2^j \cdot -k\right)$, $k = 0, \ldots, 2^j - 1$, of ϕ are readily seen to be *orthonormal*, i.e., $\langle \phi_{j,k}, \phi_{j,l} \rangle_{[0,1]} := \int_0^1 \phi_{j,k}(x)\phi_{j,l}(x)dx = \delta_{k,l}$. Hence



$$P_j(f) := \sum_{k=0}^{2^j-1} \langle f, \phi_{j,k} \rangle \phi_{j,k}$$

is for each $j \in \mathbb{N}_0$ a simple orthogonal projector from $L_2([0,1])$ onto the space S_j of piecewise constant functions subordinate to the dyadic mesh of size 2^{-j}. This projection resolves the function f up to scale j while finer details are averaged out, see Figure 1

Encyclopedia of Computational Mechanics. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

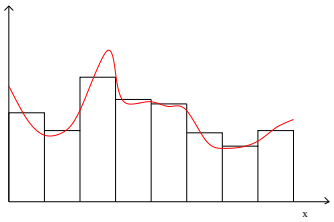


Figure 1. Piecewise constant approximation

If the resolution is found to be insufficient, one has to discard previous efforts and recompute with a larger j.

In contrast

$$f = \sum_{j=0}^{\infty} (P_j - P_{j-1}) f, \quad (P_{-1} := 0) \tag{3}$$

is a *multiscale representation* of the function f. Each term $(P_j - P_{j-1})f$ represents the detail in f at the scale 2^{-j}. Moreover, the layers of detail at each scale are mutually orthogonal.

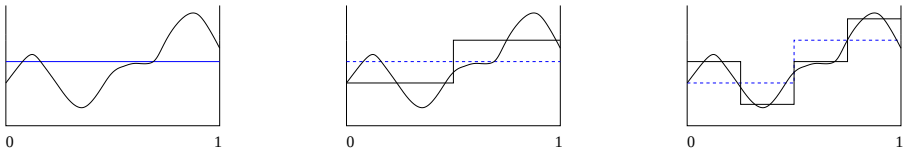
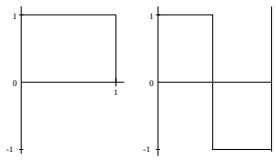


Figure 2. different levels of resolution of f

As can be seen from Figure 2, to encode the difference information one can use in place of the averaging profile $\phi(x)$ an oscillatory profile $\psi(x)$ - the *Haar wavelet* - given by



$$\psi(x) := \phi(2x) - \phi(2x-1) \text{ which implies}$$

$$\phi(2x) = (\phi(x) + \psi(x))/2 \quad \text{and therefore}$$

$$\phi(2x-1) = (\phi(x) - \psi(x))/2.$$

Hence, the fine scale averaging profiles can be recovered from a coarse scale average and an oscillatory profile. Thus, defining again $\phi_{j,k} := 2^{j/2}\phi(2^j \cdot -k)$, $\psi_{j,k} := 2^{j/2}\psi(2^j \cdot -k)$ one easily verifies the *two–scale relations*

$$\phi_{j,k} = \frac{1}{\sqrt{2}}(\phi_{j+1,2k} + \phi_{j+1,2k+1}), \qquad \psi_{j,k} := \frac{1}{\sqrt{2}}(\phi_{j+1,2k} - \phi_{j+1,2k+1})$$

$$\phi_{j+1,2k} = \frac{1}{\sqrt{2}}(\phi_{j,k} + \psi_{j,k}), \qquad \phi_{j+1,2k+1} = \frac{1}{\sqrt{2}}(\phi_{j,k} - \psi_{j,k}), \tag{4}$$

Encyclopedia of Computational Mechanics. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

which give rise to a *change of basis*

$$\sum_{k=0}^{2^{j+1}-1} c_{j+1,k}\phi_{j+1,k} = \sum_{k=0}^{2^j-1} c_{j,k}\phi_{j,k} + \sum_{k=0}^{2^j-1} d_{j,k}\psi_{j,k}, \tag{5}$$

where

$$\begin{aligned} c_{j,k} &= \frac{1}{\sqrt{2}}(c_{j+1,2k} + c_{j+1,2k+1}), \qquad d_{j,k} = \frac{1}{\sqrt{2}}(c_{j+1,2k} - c_{j+1,2k+1}) \\ c_{j+1,2k} &= \frac{1}{\sqrt{2}}(c_{j,k} + d_{j,k}), \qquad c_{j+1,2k+1} = \frac{1}{\sqrt{2}}(c_{j,k} - d_{j,k}). \end{aligned} \tag{6}$$

Thus the representation in terms of fine scale averages can be obtained from the oarse scale averages already in hand by simply adding the detail (lost through the coarse projection) encoded in terms of the oscillatory profiles. The translated dilations $\psi_{j,k}(x) := 2^{j/2}\psi(2^j \cdot -k)$ are easily seen to be pairwise orthonormal

$$\langle \psi_{j,k}, \psi_{j',k'} \rangle_{\mathbb{R}} = \delta_{(j,k),(j',k')}, \quad j, j', k, k' \in \mathbb{Z}. \tag{7}$$

Here and in the following we use the notation $\langle f, g \rangle_\Omega = \int\limits_\Omega fg dx$ but suppress at times the subscript Ω when the reference to the domain is clear from the context.

Obviously, the above change of basis (5) can be repeated which gives rise to a cascadic transform - the *fast wavelet transform*. It transforms a linear combination of fine scale box functions with an array of averages $\mathbf{c}_J$ into a linear combination of coarse scale box functions with coefficient array $\mathbf{c}_0$ and Haar wavelets with arrays of detail coeffcients $\boldsymbol{d}_j$ for each dyadic level $j < J$. This decomposition transform $\mathbf{T} : \mathbf{c}_J \to \boldsymbol{d}^J := (\mathbf{c}_0, \boldsymbol{d}_0, \boldsymbol{d}_1, \ldots, \boldsymbol{d}_{J-1})$ looks schematically as follows:

$$\begin{array}{ccccccccccc} \mathbf{c}_J & \to & \mathbf{c}_{J-1} & \to & \mathbf{c}_{J-2} & \to & \cdots & \to & \mathbf{c}_1 & \to & \mathbf{c}_0 \\ & \searrow & & \searrow & & & & \searrow & & \searrow & \\ & & \boldsymbol{d}_{J-1} & & \boldsymbol{d}_{J-2} & & \cdots & & \boldsymbol{d}_1 & & \boldsymbol{d}_0. \end{array} \tag{8}$$

In other words, from $\mathbf{c}_J$ we determine $\mathbf{c}_{J-1}$, and $\boldsymbol{d}_{J-1}$ by using (6), and so on. By (7) $\mathbf{T}$ is represented by a unitary matrix whose inverse is given by its transpose. Therefore the transform $\mathbf{T}^{-1} : \boldsymbol{d}^J := (\mathbf{c}_0, \boldsymbol{d}_0, \ldots, \boldsymbol{d}_{J-1}) \to \mathbf{c}_J$, that takes the detail coefficients into the single scale average coefficients, has a similar structure that can also be read off from the relations (6):

$$\begin{array}{ccccccccccc} \mathbf{c}_0 & \to & \mathbf{c}_1 & \to & \mathbf{c}_2 & \to & \cdots & \to & \mathbf{c}_{J-1} & \to & \mathbf{c}_J \\ & \nearrow & & \nearrow & & & & \nearrow & & \nearrow & \\ \boldsymbol{d}_0 & & \boldsymbol{d}_1 & & \boldsymbol{d}_2 & & \cdots & & \boldsymbol{d}_{J-1} & & . \end{array} \tag{9}$$

Thus, starting with $\mathbf{c}_0$ and the wavelet coefficients $\boldsymbol{d}_0, \ldots \boldsymbol{d}_{J-1}$, we can use (9) to find $\mathbf{c}_J$. Due to the cascadic structure and the fact that the relations in (6) involve only *finite filters* or *masks*, the number of operations required by both transforms is $O(2^J)$, i.e. stays proportional to the size of the arrays $\mathbf{c}_J$.

Encyclopedia of Computational Mechanics. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

In summary, there is a convenient and fast way of switching between different representations of a given projection $P_J(f) = \sum_k c_{J,k}\phi_{j,k}$, each having its advantages, as we shall see in subsequent discussions. From the theoretical point of view, since the wavelets allow us to encode the jth dyadic level of detail of a function f as

$$(P_{j+1} - P_j)f = \sum_{k=0}^{2^j-1} d_{j,k}(f)\psi_{j,k}, \qquad d_{j,k}(f), := \langle f, \psi_{j,k}\rangle,$$

the telescoping expansion (3) yields

$$f = P_0 f + \sum_{j=1}^{\infty}(P_j - P_{j-1})f = \sum_{j=-1}^{\infty}\sum_{k=0}^{2^j-1} d_{j,k}(f)\psi_{j,k} =: \boldsymbol{d}(f)^T\Psi. \tag{10}$$

The convergence of this expansion in L_2 follows from the convergence of the orthogonal projections in L_2.

A wavelet decomposition of a function $f \in L_2$ is analogous in spirit to the decimal representations of a real number. The wavelet coefficients play the role of *digits*; receiving more wavelet coefficients gives us progressively better accuracy in representing f. Of course, the classical Fourier transform of periodic functions and also Taylor expansions do, in principle, the same. The particular advantages of wavelet representations rely to a large extent on the following fact. First of all, the orthonormality of the $\psi_{j,k}$ gives

$$\|f\|_{L_2} = \left(\sum_{j=0}^{\infty}\|(P_j - P_{j-1})f\|_{L_2}^2\right)^{1/2} = \|\boldsymbol{d}(f)\|_{\ell_2}, \tag{11}$$

i.e. there is a tight relation between the function and coefficient norm. Thus perturbing the digits, which will happen in every computation, in particular, discarding small digits, will change the function norm only by the same small amount. Clearly, the convergence of the series implies that the digits will eventually have to become arbitrarily small. However, which digits become how small can easily be inferred from *local properties* of f. In fact, since the $\psi_{j,k}$ are orthogonal to constants - they have *first order vanishing moments* - one has for $S_{j,k} := \operatorname{supp}\psi_{j,k} = 2^{-j}[k, k+1]$

$$|d_{j,k}(f)| = \inf_{c\in\mathbb{R}}|\langle f - c, \psi_{j,k}\rangle| \le \inf_{c\in\mathbb{R}}\|f - c\|_{L_2(S_{j,k})} \le 2^{-j}\|f'\|_{L_2(S_{j,k})}, \tag{12}$$

where the last estimate follows e.g. from Taylor's expansion. Thus $d_{j,k}(f)$ is small when $f|_{S_{j,k}}$ is smooth.

2.2. Biorthogonal Wavelets on $\mathbb{R}$

The Haar basis is, of course, not suitable when, for instance, higher regularity of the approximation system is required. The discovery by I. Daubechies, Daubechies, 1992, of a family of compactly supported orthonormal wavelets in $L_2(\mathbb{R})$ of arbitrary high regularity opened the door to a wide range of applications. Of perhaps even more practical relevance was the subsequent construction of *biorthogonal wavelets* put forward by Cohen, Daubechies and Feauveau, 1992. The biorthogonal approach sacrifices L_2-orthogonality in favor of other properties such as symmetry of the basis functions and better localization of their supports.

Encyclopedia of Computational Mechanics. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

The construction of biorthogonal wavelets starts with a *dual pair* of compactly supported scaling functions $\phi, \tilde{\phi}$, i.e.

$$\langle \phi, \tilde{\phi}(\cdot - k)\rangle = \delta_{0,k}, \quad k \in \mathbb{Z}, \tag{13}$$

that satisfy the two scale relations

$$\phi(x) = \sum_{k\in\mathbb{Z}} a_k \phi(2x-k), \quad \tilde{\phi}(x) = \sum_{k\in\mathbb{Z}} \tilde{a}_k \tilde{\phi}(2x-k) \tag{14}$$

with finitely supported *masks* $(a_k)_{k\in\mathbb{Z}}$, $(\tilde{a}_k)_{k\in\mathbb{Z}}$, see (4). Each of the functions

$$\psi(x) = \sum_{k\in\mathbb{Z}} (-1)^k \tilde{a}_{1-k} \phi(2x-k), \quad \tilde{\psi}(x) = \sum_{k\in\mathbb{Z}} (-1)^k a_{1-k} \tilde{\phi}(2x-k) \tag{15}$$

generates by means of shifts and dilates a *biorthogonal basis* for $L_2(\mathbb{R})$. Each $f \in L_2(\mathbb{R})$ has the following unique expansions:

$$f \;=\; \sum_{j=-1}^{\infty} \sum_{k\in\mathbb{Z}} \langle f, \tilde{\psi}_{j,k}\rangle_{\mathbb{R}} \psi_{j,k} = \sum_{j=-1}^{\infty} \sum_{k\in\mathbb{Z}} \langle f, \psi_{j,k}\rangle_{\mathbb{R}} \tilde{\psi}_{j,k}, \tag{16}$$

where we have used the notation $\psi_{-1,k};= \phi_{0,k}$, $\tilde{\psi}_{-1,k};= \tilde{\phi}_{0,k}$. One thus has $\langle \psi_{j,k}, \tilde{\psi}_{l,m}\rangle_{\mathbb{R}} = \delta_{(j,l),(k,m)}$.

Each of these systems is a *Riesz basis*, which means *

$$\|f\|^2_{L_2(\mathbb{R})} \sim \sum_{j=-1}^{\infty} \sum_{k\in\mathbb{Z}} |\langle f, \tilde{\psi}_{j,k}\rangle|^2 \sim \sum_{j=-1}^{\infty} \sum_{k\in\mathbb{Z}} |\langle f, \psi_{j,k}\rangle|^2. \tag{17}$$

The inequalities (17) ensure a tight relation between the function norm and the coefficient norm.

Cohen, Daubechies and Feauveau, 1992 construct a family of biorthogonal pairs with each of $\psi, \tilde{\psi}$ of compact support. Given any desired order r of differentiability, one can find a biorthogonal pair in this family with ψ having r continous derivatives. Moreover, one can also require that a suitable linear combination of $(\phi(\cdot - k))_{k\in\mathbb{Z}}$ (respectively $(\tilde{\phi}(\cdot - k))_{k\in\mathbb{Z}}$) will represent any given polynomial of order $\leq m$, (respectively $\tilde{m}$). The biorthogonality relations then imply that the wavelets $\psi_{j,k}$, $\tilde{\psi}_{j,k}$ (for $j \geq 0$) are orthogonal to all polynomials of order $\tilde{m}$, m respectively. An analogous argument to (12) then shows that the coefficients $\langle f, \tilde{\psi}_{j,k}\rangle$, $\langle f, \psi_{j,k}\rangle$ decay like 2^{-mj}, $2^{-\tilde{m}j}$ when f has bounded derivatives on the supports of $\tilde{\psi}_{j,k}, \psi_{j,k}$ of order $m, \tilde{m}$, respectively, in L_2. Thus higher local smoothness results in a stronger size reduction of corresponding wavelet coefficients.

The setting of biorthogonal wavelets is particularly appealing from a practical point of view since the *primal generator* ϕ can be chosen as any B-spline, and in turn the primal wavelet generator ψ is also a spline function with an explicit - piecewise polynomial - analytical expression.

---

*Throughout this chapter we sometimes write $A \lesssim B$ to indicate the existence of a constant c such that $A \leq cB$ independent of any parameters on which A and B may depend. Moreover $A \sim B$ means that $A \lesssim B$ and $B \lesssim A$.

Encyclopedia of Computational Mechanics. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

2.3. Wavelets on Domains

Biorthogonal wavelets provide a beautiful and conceptually simple multiscale decomposition of functions. They offer great versatility in the choice of the basis and dual elements including compact support, smoothness, and even piecewise polynomial structure. Meanwhile they maintain the essential features of orthogonal wavelet decompositions such as norm equivalences and cancellation properties. Moreover, in connection with differential equations the space L_2 often plays only a secondary or artificial role. Therefore, dispensing with L_2-orthogonality is, in general, not even a quantitative drawback.

That is the good news. The bad news is that the above wavelet constructions are inherently made on $\mathbb{R}$, $\mathbb{R}^d$ or a torus. In numerical applications, the setting is typically on a finite domain or manifold Ω. Such domains and manifolds do not maintain the dilation and translation structure of the full Euclidean space or the torus. [†] Fortunately, there are constructions of multiscale bases tailored to general domains and manifolds. Albeit, these come at some expense of a certain level of technicality. In order not to destroy the main flow of this chapter, we shall only give an overview of some of the ideas used for these constructions. The reader can consult Dahmen, 1997 and the references quoted there for a more detailed description of the construction of multiscale bases on (bounded) domains and manifolds.

The starting point of these constructions is again *multiresolution*. By this we mean a hierarchy of (now finite dimensional) subspaces S_j of some function space $\mathcal{X}$

$$S_0 \subset S_1 \subset S_2 \subset \ldots \mathcal{X}, \quad \overline{\bigcup_j S_j} = \mathcal{X},$$

that are spanned by *single-scale bases* $S_j = \operatorname{span} \Phi_j =: S(\Phi_j)$, $\Phi_j = \{\phi_\gamma : \gamma \in \mathcal{I}_j\}$. The space $\mathcal{X}$ is typically an L_p or Sobolev space. It is important that the bases Φ_j are *scalewise stable* with respect to some discrete norm $\|\cdot\|$ in the sense that

$$\|(\|c_\gamma \phi_\gamma\|_{\mathcal{X}})_{\gamma \in \mathcal{I}_j}\| \sim \|\sum_{\gamma \in \mathcal{I}_j} c_\gamma \phi_\gamma\|_{\mathcal{X}}, \tag{18}$$

with constants that do not depend on the level j. In the case where $\mathcal{X} = L_p$ or W_p^s, the discrete norm $\|\cdot\|$ is typically the ℓ_p norm. For instance, Φ_j could be a finite element nodal basis on a j-fold refinement of some initial mesh for Ω. In this example the indices γ may represent the vertices in the mesh. One then looks for *decompositions*

$$S_{j+1} = S_j \bigoplus W_j, \quad W_j = \operatorname{span} \Psi_j, \quad \Psi_j = \{\psi_\lambda : \lambda \in \mathcal{J}_j\}.$$

The *multi-scale basis*

$$\Psi := \bigcup_{j=-1}^{\infty} \Psi_j =: \{\psi_\lambda : \lambda \in \mathcal{J}\} \qquad (\Psi_{-1} := \Phi_0)$$

is then a candidate for a wavelet basis. At this point a word on notation is in order. The index set has two component subsets: $\mathcal{J} = \mathcal{J}_\phi \cup \mathcal{J}_\psi$. The index set $\mathcal{J}_\phi$ has a finite cardinality and

[†]There is at least one strategy for maintaining the Euclidean structure by employing fictitious domain techniques; appending, for instance, essential boundary conditions by Lagrange multipliers, Dahmen and Kunoth, 2001; Kunoth, 1995.

Encyclopedia of Computational Mechanics. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

labels the basis functions in S_0 of "scaling function" type. The true wavelets correspond to the indices in $\mathcal{J}_\psi$. These indices absorb different types of information such as the scale $j = |\lambda|$, the spatial location $k(\lambda)$ or, when dealing with a spatial dimension $d > 1$, the type $e(\lambda)$ of ψ_λ. An example is $\psi_\lambda(x,y) = 2^j\psi^{1,0}(2^j(x,y) - (k,l)) = 2^{j/2}\psi(2^jx - k)2^{j/2}\phi(2^jy - l)$ where $\lambda \leftrightarrow (j,(k,l),(1,0))$.

Of course, there is a continuum of possible complements W_j and the question arises as to what are "good complements". The previous section already indicates the role of biorthogonality in this context. So a typical strategy is to split the multiresolution spaces S_j in such a way that there *exists* a biorthogonal or *dual* collection $\tilde{\Psi}$, corresponding to a *dual multiresolution sequence* $(\tilde{S}_j)$, that belongs to the dual $\mathcal{X}'$ in such a way that

$$\langle \psi_\lambda, \tilde{\psi}_\nu \rangle = \delta_{\lambda,\nu}, \quad \lambda, \nu \in \mathcal{J} \quad \text{and hence} \quad f = \sum_{\lambda \in \mathcal{J}} \langle f, \tilde{\psi}_\lambda \rangle \psi_\lambda. \tag{19}$$

The classical situation is $\mathcal{X} = \mathcal{X}' = L_2(\Omega)$ so that one has in this case the alternate representation $f = \sum_{\lambda\in\mathcal{J}}\langle f, \psi_\lambda\rangle\tilde{\psi}_\lambda$, see Carnicer, Dahmen and Peña, 1996; Cohen, Daubechies and Feauveau, 1992; Dahmen, 1994; Dahmen, 1996.

When no global smoothness is required the concept of *multiwavelets* (Alpert, 1993) offers a convenient way of generalizing the Haar basis to higher order accuracy and cancellation properties, see e.g. von Petersdorff, Schneider and Schwab, 1997 for an application to second kind integral equations.

We describe now one concrete approach (see Canuto, Tabacco and Urban, 1999; Canuto, Tabacco and Urban, 2000; Cohen and Masson, 1997; Dahmen and Schneider, 1999a; Dahmen and Schneider, 1999b) based on domain decomposition that works for $\mathcal{X} = L_2(\Omega)$ and realizes at least global continuity. Suppose that

$$\Omega = \dot{\bigcup}_{1\le m\le M} \kappa_m(\square), \quad \kappa_m : \square \to \Omega_m,$$

where each κ_i is a regular mapping from a parameter domain such as the unit d-cube, into a subdomain Ω_i of Ω. A wavelet basis Ψ^Ω is then constructed along the following lines:

- Start with biorthogonal wavelet bases $\Psi^{\mathbb{R}}, \tilde{\Psi}^{\mathbb{R}}$ on $\mathbb{R}$ and adapt them to biorthogonal wavelet bases $\Psi^I, \tilde{\Psi}^I$ on $I = [0,1]$.
- Use tensor products to obtain bases $\Psi^\square$ on the unit cube $\square = [0,1]^d$.
- Use parametric liftings to derive bases $\Psi^{\Omega_i} = \Psi^\square \circ \kappa_i^{-1}$ on $\Omega_i = \kappa_m(\square)$ which then have to be glued together to produce e.g. globally continuous bases $\Psi = \Psi^\Omega$ on Ω, see e.g. Canuto, Tabacco and Urban, 1999; Canuto, Tabacco and Urban, 2000; Cohen and Masson, 1997; Dahmen and Schneider, 1999a. An alternative approach leads to wavelets of arbitrary regularity permitted by the regularity of the domain, Dahmen and Schneider, 1999b.

The following Figure 3 has been provided as a courtesy by H. Harbrecht. It displays an example of a globally continuous primal and dual wavelet on a two dimensional patchwise defined manifold where the supports cross the patch boundaries.

Encyclopedia of Computational Mechanics. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

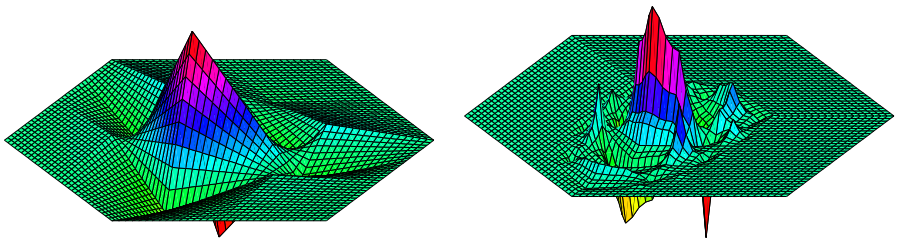


Figure 3. a) primal wavelet b) dual wavelet

Alternatively, hierarchies of uniform refinements of an arbitrary initial triangulation can be used to construct finite element based wavelets, see e.g. Dahmen and Stevenson, 1999; Stevenson, 2000.

All these constructions aim to realize (19). However, biorthogonolatity by itself is not quite sufficient in general to guarantee relations like (17). It is shown in Dahmen, 1996 that if in addition the multiresolution spaces S_j and $\tilde{S}_j$ each satisfy a (quite general form of a) direct (sometimes referred to as a *Jackson*) estimate which quantifies their approximation properties, and in addition satisfy an inverse (referred to as *Bernstein*) estimate, which quantifyies the regularity of these spaces, then an $L_2(\Omega)$- norm equivalences of the form (17) hold. One actually obtains norm equivalences for a whole range of smoothness spaces, (possibly weighted) Sobolev spaces, around L_2; a fact that is actually more important for the intended applications, Dahmen, 1996; Dahmen and Stevenson, 1999.

The above approach, in particular Dahmen and Stevenson, 1999, can be viewed as a special realization of the following general strategy. To describe this approach, it is now convenient to view a (countable) collection Θ of functions, such as a wavelet basis or a basis of scaling functions, as a column vector based on some fixed but unspecified ordering of its elements. Refinement relations of the form (14) take then the form $\Phi_j^T = \Phi_{j+1}^T \mathbf{M}_{j,0}$ where the columns of the matrix $\mathbf{M}_{j,0}$ consist of the mask coefficients in each two-scale relation for the elements of the scaling functions on level j. For instance, in the case of the Haar basis on $[0,1]$, (4) says that each column in $\mathbf{M}_{j,0}$ has at two successive positions the value $2^{-1/2}$ as the only nonzero entry. This format persists in much wider generality and can be used to represent two–scale relations for any hierarchy $\mathcal{S}$ of nested spaces spanned by scaling function bases $S_j = \operatorname{span} \Phi_j$. In the same way, a basis Ψ_j spanning some complement W_j of S_j in S_{j+1}, has the form $\Psi_j^T = \Phi_{j+1}^T \mathbf{M}_{1,j}$. It is easy to see that $\mathbf{M}_{j,1}$ *completes* $\mathbf{M}_{j,0}$ to an invertible operator $\mathbf{M}_j := (\mathbf{M}_{j,0}, \mathbf{M}_{j,1})$ if and only if $S_{j+1} = S_j \oplus W_j$ and that the complement bases are *uniformly scalewise stable* in the sense of (18) if and only if the condition numbers of the $\mathbf{M}_j$ with respect to the corresponding norms are uniformly bounded, Carnicer, Dahmen and Peña, 1996. Of course, in the case of orthonormal bases one has $\mathbf{G}_j = \mathbf{M}_j^T$.

One can now define multiscale transformations that change, for instance, the representation of an element in S_J with respect to Φ_J into the representation with respect to Φ_0 and the complement bases Ψ_j, $j < J$, in complete analogy to (8) and (9). In fact, the refinement relations imply that

$$\mathbf{c}^{j+1} = \left(\mathbf{M}_{j,0}\mathbf{c}^j + \mathbf{M}_{j,1}\boldsymbol{d}^j\right). \tag{20}$$

The refinement matrix $\mathbf{M}_{j,0}$ can be viewed as a *prediction operator*. When the detail coefficients

Encyclopedia of Computational Mechanics. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

are zero it provides an exact representation of the data on the next higher level of resolution. It is also sometimes called a *subdivision* operator, see e.g. Arandiga, Donat and Harten, 1998; Arandiga, Donat and Harten, 1999; Carnicer, Dahmen and Peña, 1996. It follows from (20) that the transformation $\mathbf{T}_J^{-1}$, taking the detail coefficients into the single scale coefficients $\mathbf{c}^J$, uses the matrices $\mathbf{M}_j$ in each cascadic step of (9).

Conversely, setting $\mathbf{G}_j := \mathbf{M}_j^{-1} = \binom{\mathbf{G}_{j,0}}{\mathbf{G}_{j,1}}$, one has

$$\mathbf{c}^j = \mathbf{G}_{j,0}\mathbf{c}^{j+1} \text{ and } \boldsymbol{d}^j = \mathbf{G}_{j,1}\mathbf{c}^{j+1}. \tag{21}$$

Hence the transformation $\mathbf{T}_J$ that decomposes $\mathbf{c}^J$ into details $\boldsymbol{d}^j$ and coarse scale coefficients $\mathbf{c}^0$ has the same cascadic structure as (8), now based on the filter matrices $\mathbf{G}_j$.

Furthermore, one can show that the transformations $\mathbf{T}_J$ have uniformly bounded spectral condition numbers independent of J if and only if the corresponding union Ψ of the complement bases Ψ_j and the coarse scale basis Φ_0 forms a Riesz basis for L_2, Dahmen, 1994; Carnicer, Dahmen and Peña, 1996.

While it is often difficult to directly construct a Riesz basis for the space of interest, in many cases, it is easy to find for each level j some *initial* complement bases $\check{\Psi}_j$. For instance, when working with a hierarchy of nodal finite element bases, complement bases are provided by the *hierarchical basis* consisting of those nodal basis functions at the nodes of the next higher level of resolution, see e.g. Yserentant, 1986. As a second step one can then generate from this initial multiscale decomposition another one that has certain desirable properties, for instance, a higher order of vanishing moments. The important point to be made in this regard is that all of this can be done completely on a discrete level. To this end, suppose that an *initial* completion $\check{\mathbf{M}}_{j,1}$ of the refinement matrix $\mathbf{M}_{j,0}$ (and $\check{\mathbf{G}}_j$) is known. Then *all* other *stable completions* have the form Carnicer, Dahmen and Peña, 1996

$$\mathbf{M}_{j,1} = \mathbf{M}_{j,0}\mathbf{L} + \check{\mathbf{M}}_{j,1}\mathbf{K} \tag{22}$$

with inverse blocks

$$\mathbf{G}_{j,0} = \check{\mathbf{G}}_{j,0} - \check{\mathbf{G}}_{j,1}(\mathbf{K}^T)^{-1}\mathbf{L}^T, \quad \mathbf{G}_{j,1} = \check{\mathbf{G}}_{j,1}(\mathbf{K}^T)^{-1}. \tag{23}$$

In fact, this follows from the identity

$$\mathbf{I} = \check{\mathbf{M}}_j\check{\mathbf{G}}_j = \check{\mathbf{M}}_j \begin{pmatrix} \mathbf{I} & \mathbf{L} \\ \mathbf{0} & \mathbf{K} \end{pmatrix} \begin{pmatrix} \mathbf{I} & -\mathbf{L}\mathbf{K}^{-1} \\ \mathbf{0} & \mathbf{K}^{-1} \end{pmatrix} \check{\mathbf{G}}_j =: \mathbf{M}_j\mathbf{G}_j.$$

The special case $\mathbf{K} = \mathbf{I}$ is often referred to as *Lifting Scheme*, Sweldens, 1996; Sweldens, 1998. The parameters in the matrices $\mathbf{L}, \mathbf{K}$ can be used to modify the complement bases. Such modifications of stable completions are used for instance in the construction of wavelets on an interval (see e.g. Dahmen, Kunoth and Urban, 1999) and hence in the above mentioned domain decomposition approach, Canuto, Tabacco and Urban, 1999; Canuto, Tabacco and Urban, 2000; Cohen and Masson, 1997; Dahmen and Schneider, 1999a, as well as in the construction of finite element based wavelets through coarse grid corrections, Carnicer, Dahmen and Peña, 1996; Dahmen, 1997; Dahmen and Kunoth, 1992; Dahmen and Stevenson, 1999; Stevenson, 2000; Vassilevski and Wang, 1997. A further important application concerns raising the order of *vanishing moments*: Choose $\mathbf{K} = \mathbf{I}$ and $\mathbf{L}$ such that

$$\int_\Omega \Psi_j^T P dx = \int_\Omega \Phi_{j+1}^T \mathbf{M}_{j,1} P dx = \int_\Omega \Phi_j^T \mathbf{L} P + \check{\Psi}_j^T P dx = 0, \quad P \in \mathbb{P}_{m^*}. \tag{24}$$

Encyclopedia of Computational Mechanics. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

The significance of the above formulations lies in the versatility of handling multiscale decompositions entirely on a discrete level. This allows one to circumvent (at least to some extent) the explicit construction of complicated basis functions, see Harten, 1996. However, statements about stability are often based on explicit knowledge of the underlying multiscale bases.

2.4. The Key Features

The ideas put forward in the previous section allow one to construct multiscale bases for a variety of domains and even closed surfaces. In this section we collect the main properties of these constructions that are valid on a domain Ω of spatial dimension d. We shall assume these properties in our subsequent applications. These key properties can be summarized as follows:

- *Locality* (L) • *Cancellation Properties* (CP) • *Norm Equivalences* (NE).

Locality: (L) means that the elements of Ψ all have compact support $S_\lambda := \operatorname{supp}\psi_\lambda$ that scales properly, i.e.

$$\operatorname{diam}(S_\lambda) \sim 2^{-|\lambda|}. \tag{25}$$

Locality is crucial for applications on bounded domains and for the efficiency of associated multiscale transforms.

Cancellation Properties: (CP) generalizes our earlier observation (12). It means that integrating a wavelet against a locally smooth function acts like differencing. Assume for example that the wavelets are normalized in L_2, i.e. $\|\psi_\lambda\|_{L_2} \sim 1$. Cancellation will then mean that

$$|\langle v, \psi_\lambda\rangle| \lesssim 2^{-|\lambda|\left(\tilde{m}+\frac{d}{2}-\frac{d}{p}\right)}|v|_{W_p^{\tilde{m}}(S_\lambda)}, \quad \lambda \in \mathcal{J}_\psi, \tag{26}$$

where $|v|_{W_p^n(G)}$ is the usual nth order seminorm of the corresponding Sobolev space on the domain G. Analogous relations can of course be formulated for the dual basis $\tilde{\Psi}$.

The integer $\tilde{m}$ signifies the strength of the cancellation properties because it says up to which order the local smoothness of the function is rewarded by the smallness of the coefficients (in this case of the dual expansion). Obviously, when Ω is a Euclidean domain, (26) implies that the wavelets have *vanishing polynomial moments* of order $\tilde{m}$, i.e.,

$$\langle P, \psi_\lambda\rangle_\Omega = 0, \quad P \in \mathbb{P}_{\tilde{m}},\ \lambda \in \mathcal{J}_\psi. \tag{27}$$

Conversely, as in (12), the vanishing moments imply that for $\frac{1}{p}+\frac{1}{p'}=1$

$$\begin{aligned} |\langle v, \psi_\lambda\rangle| &= \inf_{P\in\mathbb{P}_{\tilde{m}}} |\langle v-P, \psi_\lambda\rangle| \le \inf_{P\in\mathbb{P}_{\tilde{m}}} \|v-P\|_{L_p(S_\lambda)}\|\psi_\lambda\|_{L_{p'}}, \\ &\lesssim 2^{-|\lambda|\left(\frac{d}{2}-\frac{d}{p}\right)} \inf_{P\in\mathbb{P}_{\tilde{m}}} \|v-P\|_{L_p(S_\lambda)}, \end{aligned}$$

where we have used that

$$\|\psi_\lambda\|_{L_{p'}} \sim 2^{|\lambda|\left(\frac{d}{p'}-\frac{d}{2}\right)} \sim 2^{|\lambda|\left(\frac{d}{2}-\frac{d}{p}\right)} \quad \text{when} \quad \|\psi_\lambda\|_{L_2} \sim 1. \tag{28}$$

Now standard estimates on *local polynomial approximation* (see e.g. DeVore and Sharpley, 1984) tell us that

$$\inf_{P\in\mathbb{P}_k} \|v-P\|_{L_p(G)} \lesssim (\operatorname{diam} G)^k |v|_{W_p^k(G)}$$

Encyclopedia of Computational Mechanics. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

which yields (26). We refer to (26) as the cancellation property of order $\tilde{m}$ rather than to (27), since it makes sense for domains where ordinary polynomials are not defined.

Norm Equivalences: The cancellation properties tell us under what circumstances wavelet coefficients are small. One expects to have only relatively few significant coefficients when the expanded function is very smooth except for singularities on lower dimensional manifolds. This helps to recover a function with possibly few coefficients only if small perturbations in the coefficients give rise to perturbations of the function that are also small with respect to the relevant norm. Recall that for function spaces $\mathcal{X}$ with local norms it is usually easy to construct multiscale bases Ψ that are uniformly *scalewise* stable, i.e.

$$\|\sum_{|\lambda|=j} d_\lambda \psi_\lambda\|_{\mathcal{X}} \sim \|(\|d_\lambda \psi_\lambda\|_{\mathcal{X}})_{|\lambda|=j}\|, \tag{29}$$

uniformly in j, where $\|\cdot\|$ is some appropriate discrete norm.

In some cases this stability property can be extended to the whole array Ψ *over all* scales. In the particular case when $\mathcal{X} = \mathcal{H}$ is a Hilbert space, this is expressed by saying that, with the normalization $\|\psi_\lambda\|_{\mathcal{H}} \sim 1$, the family Ψ is a *Riesz basis* for the whole function space $\mathcal{H}$, i.e. every element $v \in \mathcal{H}$ possesses a unique expansion in terms of Ψ and there exist finite positive constants c_Ψ, C_Ψ such that

$$c_\Psi \|(v_\lambda)\|_{\ell_2} \leq \|\sum_\lambda v_\lambda \psi_\lambda\|_{\mathcal{H}} \leq C_\Psi \|(v_\lambda)\|_{\ell_2}, \quad \forall\, \mathbf{v} = (v_\lambda) \in \ell_2. \tag{30}$$

Thus, while relaxing the requirement of orthonormality, a Riesz basis still establishes a strong coupling between the continuous world, in which the the mathematical model is often formulated, and the discrete realm which is more apt to computational realizations. Therefore, it should not be a surprise that the availability of such bases for function spaces may be exploited for numerical methods.

We shall exploit norm equivalences for the problem class (2) where the relevant spaces are *Sobolev spaces* or tensor products of them. Recall that for $n \in \mathbb{N}$ the space $H^n(\Omega)$ consists of those elements of $L_2(\Omega)$ whose nth order weak derivatives are also in $L_2(\Omega)$. More generally, for $1 \leq p \leq \infty$ we have

$$W_p^n(\Omega) := \{f : \partial^\alpha f \in L_p(\Omega),\ |\alpha| \leq n\}, \tag{31}$$

and the corresponding (semi-)norms are given by $|f|_{W_p^n(\Omega)} := \left(\sum_{|\alpha|=n} \|\partial^\alpha f\|^p_{L_p(\Omega)}\right)^{1/p}$ and $\|v\|^p_{W_p^n(\Omega)} := \sum_{m=0}^n |f|^p_{W_p^m(\Omega)}$. Dealing with traces of functions on boundary manifolds, for instance, forces one to consider also non-integer smoothness orders $t \in \mathbb{R}$. For $t > 0$ these spaces can be defined either by interpolation between spaces of integer order (see e.g. Bergh and Löfström, 1976) or directly through intrinsic norms of the form

$$\|v\|_{W_p^t(\Omega)} = \left(\|v\|^p_{W_p^n(\Omega)} + \sum_{|\alpha|=n} \int_\Omega \int_\Omega \frac{|\partial^\alpha v(x) - \partial^\alpha v(y)|^p}{|x-y|^{d+(t-n)p}} dx dy\right)^{1/p}, \quad n := \lfloor t \rfloor.$$

Moreover, for Lipschitz domains Ω and $\Gamma \subset \partial\Omega$ we denote by $H^t_{0,\Gamma}(\Omega)$ the closure of those C^∞ functions on Ω with respect to the H^t-norm that vanish on Γ. We briefly write $H^1_0(\Omega)$ when $\Gamma = \partial\Omega$. We refer to Adams, 1978 for more details on Sobolev spaces.

Encyclopedia of Computational Mechanics. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

In the sequel for $t \geq 0$, H^t will denote some closed subspace of $H^t(\Omega)$ either of the form $H_0^t(\Omega) \subseteq H^t \subseteq H^t(\Omega)$ or with finite codimension in $H^t(\Omega)$. For $t < 0$ we define H^t as the *dual* space $H^t = (H^{-t})'$. Starting with a suitable wavelet Riesz basis Ψ for $\mathcal{H} = L_2(\Omega)$, a whole family of realizations of (30) can be formulated as follows. There exist positive constants $\gamma, \tilde{\gamma} > 0$ (depending on the regularity of the wavelet basis) with the following property: For $s \in (-\tilde{\gamma}, \gamma)$ there exist positive constants c_s, C_s, such that every $v \in H^s$ possesses a unique expansion $v = \sum_{\lambda \in \mathcal{J}} v_\lambda 2^{-s|\lambda|} \psi_\lambda$ such that

$$c_s \|(v_\lambda)_\lambda\|_{\ell_2} \leq \| \sum_{\lambda \in \mathcal{J}} v_\lambda 2^{-s|\lambda|} \psi_\lambda \|_{H^s} \leq C_s \|(v_\lambda)_\lambda\|_{\ell_2}. \tag{32}$$

Thus properly scaled versions of the wavelet basis Ψ for L_2 are Riesz bases for a *whole range* of smoothness spaces, including of course $s = 0$ as a special case. This range depends on the *regularity* of the wavelets. In many constructions one has $\gamma = 3/2$ corresponding to globally continuous wavelets, Canuto, Tabacco and Urban, 1999; Canuto, Tabacco and Urban, 2000; Cohen and Masson, 1997; Dahmen and Schneider, 1999a; Dahmen and Stevenson, 1999.

Establishing (32) works actually the other way around. It is usually easier to verify (32) for *positive s*. This can be derived from the validity of *Bernstein and Jackson* estimates for the *primal* multiresolution sequences only. If one can do this, however, for the primal *and* for the dual multiresolution sequences associated with a dual pair of multiscale bases $\Psi, \tilde{\Psi}$, (32) follows for the whole range of regularity indices s by an interpolation argument, see e.g. Dahmen, 1996; Dahmen and Stevenson, 1999. In particular, this says that the Riesz basis property for $L_2(\Omega)$ follows from that of scaled versions of $\Psi, \tilde{\Psi}$ for positive Sobolev regularity, see also Cohen, 2000; Cohen, 2003; Dahmen, 1997; Dahmen, 2003.

Remark 2.1. *We emphasize the case (32) because it implies further relations that will be important later for robustness. To describe these, recall our convention of viewing a collection* Θ *of basis functions sometimes as a vector whose entries are ordered in a fixed but unspecified way. Ordering the wavelet coefficient arrays in a natural way, we can write* $\sum_{\lambda \in \mathcal{J}} v_\lambda 2^{-s|\lambda|} \psi_\lambda =: \mathbf{v}^T \mathbf{D}^{-s} \Psi$ *where* $\left(\mathbf{D}^t := \left(2^{t|\lambda|} \delta_{\lambda,\nu}\right)_{\lambda,\nu} \right)$ *and* $\mathbf{v} := (v_\lambda)_{\lambda \in \mathcal{J}}$. *In the problem class (2) one often encounters Hilbert (energy) spaces endowed with a norm of the type* $\|v\|_{\mathcal{H}_\epsilon}^2 := \epsilon \langle \nabla v, \nabla v \rangle + \langle v, v \rangle$. *The performance of multilevel preconditioners for such problems often depends on* ϵ. *It will be seen that a remedy for this can be based on* robust *equivalences of the following form that can be derived from (32) for* $s = 0$ *and* $s = 1$, *Cohen, Dahmen and DeVore, 2001; Dahmen, 2001 : assume that* $\gamma > 1$ *and define the diagonal matrix*
$\mathbf{D}_\epsilon := \left((1 + \sqrt{\epsilon} 2^{|\lambda|}) \delta_{\lambda,\mu} \right)_{\lambda,\mu \in \mathcal{J}}$. *Then*

$$\left(2(c_0^{-2} + c_1^{-2})\right)^{-1/2} \|\mathbf{v}\|_{\ell_2} \leq \|\mathbf{v}^T \mathbf{D}_\epsilon^{-1} \Psi\|_{\mathcal{H}_\epsilon} \leq \left(C_0^2 + C_1^2\right)^{1/2} \|\mathbf{v}\|_{\ell_2}. \tag{33}$$

We wish to conclude this section with the following remarks concerning *duality*, see e.g. Dahmen, 2003 for more details. As indicated before, the known constructions of a wavelet basis Ψ, that satisfy norm equivalences of the form (32), involve to some extent the simultaneous construction of a dual basis $\tilde{\Psi}$. Conversely, the existence of such a dual basis is actually a consequence of the Riesz basis property in the following sense. It is not hard to show that the validity of (30) implies the existence of a collection $\tilde{\Psi} \subset \mathcal{H}'$ such that $\langle \psi_\lambda, \tilde{\psi_\nu} \rangle = \delta_{\lambda,\nu}$, where $\langle \cdot, \cdot \rangle$ is the duality pairing that identifies the representation of $\mathcal{H}'$. Moreover, $\tilde{\Psi}$ is a Riesz basis for $\mathcal{H}'$, i.e.

$$C_\Psi^{-1} \|\mathbf{w}\|_{\ell_2} \leq \|\mathbf{w}^T \tilde{\Psi}\|_{\mathcal{H}'} \leq c_\Psi^{-1} \|\mathbf{w}\|_{\ell_2}, \quad \mathbf{w} \in \ell_2. \tag{34}$$

Encyclopedia of Computational Mechanics. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

This will mainly be used in the equivalent form

$$C_\Psi^{-1}\|\langle\Psi, v\rangle\|_{\ell_2} \leq \|v\|_{\mathcal{H}'} \leq c_\Psi^{-1}\|\langle\Psi, v\rangle\|_{\ell_2}, \tag{35}$$

where we have abbreviated $\langle\Psi, v\rangle := (\langle\psi_\lambda, v\rangle : \lambda \in \mathcal{J})^T$.

In particular, if we want to construct a Riesz basis for L_2 then the dual basis (with respect to the L_2–inner product as a dual pairing) must also be a Riesz basis in L_2, in agreement with the above remarks concerning (32). This rules out the practically convenient so called *hierarchical bases* induced by interpolatory scaling functions since the dual basis is essentially comprised of Dirac distributions. An important further example is $\mathcal{H} = H_0^1(\Omega)$ for which, according to (32), a Riesz basis is obtained by renormalizing the functions ψ_λ with the weight $2^{-|\lambda|}$, which also amounts to redefining Ψ as $\mathbf{D}^{-1}\Psi$. In this case, (35) offers a convenient way of evaluating the H^{-1}-norm which is usefull for least squares formulations of second order elliptic problems or their mixed formulations, Dahmen, Kunoth and Schneider, 2002. Note that in this particular case, unlike the situation in Remark 2.1, Ψ need not be a Riesz basis for L_2.

2.5. Wavelets and Linear Operators

So far we have focussed on wavelet representations of functions. For the problem class (2), in particular, it will be important to deal with wavelet representations of *operators*. In this section we collect a few important facts concerning linear operators that follow from the above features. As a simple guiding example consider Poisson's equation on some bounded domain $\Omega \subset \mathbb{R}^d$

$$-\Delta u = f \quad \text{in} \;\; \Omega, \quad u = 0 \;\; \text{on} \;\; \Gamma := \partial\Omega. \tag{36}$$

It will be crucial to interpret this equation properly. Multiplying both sides of (36) by smooth test functions that vanish on Γ, and integrating by parts, shows that the solution u satisfies

$$\langle\nabla v, \nabla u\rangle = \langle v, f\rangle \quad \text{for all smooth} \;\; v, \tag{37}$$

where $\langle v, w\rangle := \int_\Omega vw dx$. However, this latter form makes sense even when u belongs only to the Sobolev space $H_0^1(\Omega)$ (recall Section 2.4) and when the test functions also just belong to $H_0^1(\Omega)$. Moreover, the right hand side makes sense whenever f is only a distribution in the *dual* $H^{-1}(\Omega)$ of $H_0^1(\Omega)$. Here $\langle\cdot,\cdot\rangle$ is then understood to be the dual form on $H_0^1(\Omega) \times H^{-1}(\Omega)$ induced by the standard L_2-inner product. Thus, defining the linear operator $\mathcal{A}$ by $\langle\nabla v, \nabla u\rangle = \langle v, \mathcal{A}u\rangle$ for all $v, u \in H_0^1(\Omega)$, the boundary value problem (36) is equivalent to the operator equation

$$\mathcal{A}u = f, \tag{38}$$

where, roughly speaking, $\mathcal{A}$ is in this case the Laplacian (with incorporated homogeneous boundary conditions), taking $H_0^1(\Omega)$ into its dual $H^{-1}(\Omega)$.

The Standard Wavelet Representation: Suppose now that as in the Laplace case described above, we have a linear operator $\mathcal{A} : \mathcal{H} \mapsto \mathcal{H}'$ and that Ψ is a Riesz-basis for $\mathcal{H}$ i.e. (30) holds. Then for any $v = \sum_\lambda v_\lambda\psi_\lambda \in \mathcal{H}$ one has

$$\mathcal{A}v \;\; = \;\; \sum_\lambda \langle\psi_\lambda, \mathcal{A}v\rangle\tilde{\psi}_\lambda = \sum_\lambda \langle\psi_\lambda, \mathcal{A}(\sum_\nu v_\nu\psi_\nu)\rangle\tilde{\psi}_\lambda = \sum_\lambda \left(\sum_\nu \langle\psi_\lambda, \mathcal{A}\psi_\nu\rangle v_\nu\right)\tilde{\psi}_\lambda$$

Encyclopedia of Computational Mechanics. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

Thus the coeffcient array $\mathbf{w}$ of $\mathcal{A}v \in \mathcal{H}'$ with respect to the dual basis $\tilde{\Psi}$ is given by

$$\mathbf{w} = \boldsymbol{A}\mathbf{v} \quad \text{where} \quad \boldsymbol{A} := (\langle \psi_\lambda, A\psi_\nu \rangle)_{\lambda,\nu}\,, \quad \mathbf{v} = (v_\nu)_\nu. \tag{39}$$

The above example (36) is a typical application where Ψ and $\tilde{\Psi}$ are biorthogonal Riesz bases in the Sobolev space $\mathcal{H} = H_0^1(\Omega)$ and its dual $\mathcal{H}' = H^{-1}(\Omega)$, respectively.

$\boldsymbol{A}$ is often referred to as the *standard* wavelet representation of $\mathcal{A}$. Note that in conventional discretizations such as Finite Elements and Finite Differences, the operators can usually only be *approximated*. A basis allows one to capture, at least conceptually, all of the full infinite dimensional operator, a fact that will later be seen to have important consequences.

Well–Posedness and an Equivalent ℓ_2–Problem: We say that an operator equation of the form (38) is *well-posed* if (either $\mathcal{A}$ maps $\mathcal{H}$ onto $\mathcal{H}'$ or $f \in \text{range}\,(\mathcal{A})$ and) there exist positive constants $c_\mathcal{A}, C_\mathcal{A}$ such that

$$c_\mathcal{A}\|v\|_\mathcal{H} \le \|\mathcal{A}v\|_{\mathcal{H}'} \le C_\mathcal{A}\|v\|_\mathcal{H} \quad \text{for all} \quad v \in \mathcal{H}. \tag{40}$$

Here $\mathcal{H}'$ is the dual of $\mathcal{H}$ endowed with the norm

$$\|w\|_{\mathcal{H}'} := \sup_{v\in\mathcal{H}} \frac{\langle v, w\rangle}{\|v\|_\mathcal{H}} \tag{41}$$

and $\langle\cdot,\cdot\rangle$ is a dual form on $\mathcal{H}\times\mathcal{H}'$ (which is induced as before by the standard inner product in some pivot L_2 space).

Clearly (40) means that for any data f in the dual $\mathcal{H}'$ – the range of $\mathcal{A}$ – there exists a unique solution u which depends continuously on the data f. Thus well–posedness refers to continuity with respect to a specific topology given by the *energy space* $\mathcal{H}$. It is not hard to show that in the case of Poisson's problem (36), (40) is a consequence of H^1–ellipticity

$$\langle \nabla v, \nabla v\rangle \ge c\|v\|^2_{H^1(\Omega)} := \|v\|^2_{L_2(\Omega)} + \|\nabla v\|^2_{L_2(\Omega)}, \quad |\langle \nabla v, \nabla w\rangle| \le C\|v\|_{H^1(\Omega)}\|w\|_{H^1(\Omega)}, \tag{42}$$

which in turn follows from Poincaré's inequality. While in this special case the right space $\mathcal{H} = H_0^1(\Omega)$ is easily recognized, the identification of a suitable $\mathcal{H}$ such that (40) holds is sometimes a nontrivial task, an issue that will be taken up again later.

An important observation is that, once the mapping property (40) has been established, the Riesz basis property (30) for the energy space allows one to transform the original problem into an equivalent one which is now *well–posed* in the Euclidean metric. This is of particular importance in parameter dependent cases such as the Hilbert space $\mathcal{H}_\epsilon$ considered in the previous section.

Theorem 2.2. *Suppose that $\mathcal{A} : \mathcal{H} \mapsto \mathcal{H}'$ satisfies (40) and that Ψ is a Riesz basis for $\mathcal{H}$, i.e. (30) holds. Let $\boldsymbol{A}$ denote the standard representation of $\mathcal{A}$ with respect to Ψ. Then (38) is equivalent to $\boldsymbol{A}\mathbf{u} = \mathbf{f}$, where $u = \sum_{\lambda\in\mathcal{J}} u_\lambda\psi_\lambda$, $\mathbf{f} = (\langle\psi_\lambda, f\rangle)_{\lambda\in\mathcal{J}}$. Moreover, $\boldsymbol{A}$ is boundedly invertible on ℓ_2, i.e.*

$$c_\Psi^2 c_\mathcal{A}\|\mathbf{v}\|_{\ell_2} \le \|\boldsymbol{A}\mathbf{v}\|_{\ell_2} \le C_\Psi^2 C_\mathcal{A}\|\mathbf{v}\|_{\ell_2}, \quad \textit{for all} \quad \mathbf{v} \in \ell_2. \tag{43}$$

Proof: By (30) one has for any $v = \sum_\lambda v_\lambda\psi_\lambda$

$$\|\mathbf{v}\|_{\ell_2} \le c_\Psi^{-1}\|v\|_\mathcal{H} \le c_\Psi^{-1}c_\mathcal{A}^{-1}\|\mathcal{A}v\|_{\mathcal{H}'} \le c_\Psi^{-2}c_\mathcal{A}^{-1}\|(\langle\psi_\lambda, \mathcal{A}v\rangle)_{\lambda\in\mathcal{J}}\|_{\ell_2} = c_\Psi^{-2}c_\mathcal{A}^{-1}\|\boldsymbol{A}\mathbf{v}\|_{\ell_2},$$

where we have used (34). The reverse estimate follows analogously. □

Encyclopedia of Computational Mechanics. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

3. Evolution problems – compression of flow fields

We shall now address the problem class (1) of evolution equations. In many relevant instances, the evolution of the quantity $u(x,t)$ expresses a *conservation law* in the sense that for any test volume V in the domain Ω of interest,

$$\partial_t \int_V u dx + \int_{\partial V} f(u) \cdot n ds = 0, \tag{44}$$

where $f(u)$ is the *flux function* and n denotes the outward normal on the boundary ∂V. When the solution is sufficiently smooth, (44) leads to a first order system of partial differential equations of the form

$$\partial_t u + \mathrm{div}_x f(u) = 0, \tag{45}$$

which is said to be *hyperbolic* when the Jacobian matrix $Df(x)$ has for all x real eigenvalues with a full basis of eigenvectors. Hyperbolic systems of conservation laws are used to model phenomenon as diverse as traffic, information and fluid flows. Perhaps the most well-known example is the system of Euler equations which model compressible fluid flow in terms of balance equations for mass, momentum and energy. Such system have to be complemented by suitable initial/boundary conditions. For simplicity we shall assume pure initial data $u_0(x) = u(x,0)$ with compact support.

Numerical methods for solving (44) are typically based on evolving cell averages $\bar{u}_C(t) := (\int_C u(x,t)dx)/|C|$, where C runs over a partition $\mathcal{P}$ of the domain Ω into disjoint cells. In fact, the balance relation (44) also reads

$$\bar{u}_C(t+\Delta t) = \bar{u}_C(t) + \frac{\Delta t}{|C|} B_C(t), \quad B_C(t) := \frac{1}{\Delta t} \int_t^{t+\Delta t} \int_{\partial C} f(u) \cdot n dx dt. \tag{46}$$

A *finite volume scheme* will mimic this time evolution by replacing the exact flux balance $B_C(t)$ by a numerical approximation computed from the current approximation of the exact cell-average. More precisely, given a time step Δt, the scheme computes approximate values $u_C^n \approx u_C(n\Delta t)$ according to

$$u_C^{n+1} = u_C(t) + \frac{\Delta t}{|C|} \sum_{C' \cap C \neq \emptyset} F_{C,C'}^n, \tag{47}$$

where $F_{C,C'}^n$ is the numerical flux across the common boundary of C and an adjacent cell C' for the time interval $[n\Delta t, (n+1)\Delta t]$. This numerical flux typically depends on the values u_C^n and $u_{C'}^n$, and possibly on other neighboring values. We shall always assume that the scheme is *conservative*, i.e. $F_{C,C'}^n = -F_{C',C}^n$. The initialization of the scheme uses the exact (or approximately computed) averages $u_C^0 := (\int_C u_0(x)dx)/|C|$ of the initial data u_0.

Denoting by $\mathbf{u}^n = (u_C^n)_{C \in \mathcal{P}}$ the array of cell averages, the finite volume scheme is thus summarized by a one step relation

$$\mathbf{u}^{n+1} = \mathbf{E}\mathbf{u}^n, \tag{48}$$

where $\mathbf{E}$ is a nonlinear discrete evolution operator. The computationally expensive part of this numerical method is the evaluation of the numerical fluxes $F_{C,C'}^n$, which is typically based

Encyclopedia of Computational Mechanics. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

on the (approximate) solution of local Riemann problems. An a–priori error analysis of these classical numerical methods is only available to a limited extent. It refers to scalar problems, not to systems, and is rigorously founded only for uniform meshes. The proven error estimates are of low approximation orders like $h^{1/2}$ where h is the mesh size, i.e. the maximal diameter of the cells.

It is well-known that the solutions of hyperbolic conservation laws exhibit a highly nonhomogeneous structure. Discontinuities in the solution can develop after finite time even for arbitraily smooth initial data. So the solution exhibits regions of high regularity separated by regions of discontinuities (shocks). To capture the singular effects in the solution by using classical discretizations based on uniform (or even quasi-uniform) partitions into cells would require a very fine resolution near the singularities and thus lead to enormous problem sizes. We see that the nature of the solution begs for the use of adaptive methods which would give finer resolution in the regions of shock discontinuities and maintain coarser resolution otherwise. The usual numerical approach is to generate such partitions adaptively. The difficulties in such an approach are to determine these regions and properly perform the time evolution on these inhomogeneous discretizations.

The analytic structure of solutions to (45) also points to possible advantages in using multiscale decompositions of the solution u in a numerical procedure. Because of the cancellation property (26), the coefficients of u would be small in those regions where the solution is smooth and would have significant size only near shocks. Thus, a multiscale decomposition would be excellent at identifying the regions of discontinuities by examining the size of the coefficients, and providing economical representations of the approximate solution at time $n\Delta t$ by an adapted set of wavelet coefficients $(d_\lambda^n)_{\lambda\in\Lambda^n}$. The approximate solution would therefore be given by

$$u_{\Lambda^n} = \sum_{\lambda\in\Lambda^n} d_\lambda^n \psi_\lambda, \tag{49}$$

where the set Λ^n is allowed to vary with n. The main difficulty in this approach is how to perform the evolution step strictly in terms of the wavelet coefficients. In other words, given Λ_n and the coefficients $(d_\lambda^n)_{\lambda\in\Lambda^n}$, how would we evolve on these data to obtain a good approximation at the next time step? This has led to the introduction of *dynamically adaptive schemes* in Maday, Perrier and Ravel, 1991, in which the derivation of $(\Lambda^{n+1}, u_{\Lambda^{n+1}})$ from $(\Lambda^n, u_{\Lambda^n})$ typically goes in three basic steps:

(i) **Refinement:** a larger set $\tilde\Lambda^{n+1}$ with $\Lambda^n \subset \tilde\Lambda^{n+1}$ is derived from an *a-posteriori* analysis of the computed coefficients d_λ^n, $\lambda \in \Lambda^n$.
(ii) **Evolution:** a first numerical solution $u_{\tilde\Lambda^{n+1}} = \sum_{\lambda\in\tilde\Lambda^{n+1}} d_\lambda^{n+1}\psi_\lambda$ is computed from u_n and the data of the problem.
(iii) **Coarsening:** the smallest coefficients of $\tilde u_{n+1}$ are thresholded, resulting in the numerical solution $u_{\Lambda^{n+1}} = \sum_{\lambda\in\Lambda^{n+1}} d_\lambda^{n+1}\psi_\lambda$ supported on the smaller set $\Lambda^{n+1} \subset \tilde\Lambda^{n+1}$.

A few words are in order concerning the initialization of the scheme: ideally, we can obtain an adaptive expansion u_{Λ^0} of the intial value data u_0 into a linear combination of wavelets by a thresholding procedure on its global expansion, i.e.

$$u_{\Lambda^0} = \sum_{\lambda\in\Lambda^0} d_\lambda^0\psi_\lambda, \quad \Lambda^0 := \{\lambda \text{ s.t. } \|d_\lambda^0\psi_\lambda\|_{\mathcal{X}} \geq \eta\}, \tag{50}$$

Encyclopedia of Computational Mechanics. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

where $\mathcal{X}$ is some prescribed norm in which we target to measure the error, η a prescribed threshold and $d^0_\lambda := \langle u_0, \tilde{\psi}_\lambda \rangle$ are the wavelet coefficients of u_0. In practice, we cannot compute all the values of these coefficients, and one thus needs a more reasonable access to a compressed representation. This is typically done through some *a-priori* analysis of the initial value u_0. In particular, if u_0 is provided by an analytic expression, or if we have some information on the local size of its derivatives, estimates on the decay of wavelet coefficients, such as (26), can be used to avoid the computation of most details which are below threshold. With such a strategy, we expect to obtain Λ^0 and $(u^0_\lambda)_{\lambda \in \Lambda^0}$ with a memory and computational cost which is proportional to $\#(\Lambda^0)$.

Then, assuming that at time $n\Delta t$ the approximate solution u_{Λ^n} has the form (49) for some set Λ^n of coefficients, the problem is thus both to select a correct set of indices Λ^{n+1} and to compute the new coefficients d^{n+1}_λ for $\lambda \in \Lambda^{n+1}$. As we already explained, this is done by (i) refining Λ^n into an intermediate set $\tilde{\Lambda}^{n+1}$ which is well fitted to describing the solution at time $(n+1)\Delta t$, (ii) computing $u_{\tilde{\Lambda}^{n+1}}$ supported by $\tilde{\Lambda}^{n+1}$ and (iii) deriving $(u_{\Lambda^{n+1}}, \Lambda^{n+1})$ from $u_{\tilde{\Lambda}^{n+1}}$ by a thresholding process. The selection of the intermediate set $\tilde{\Lambda}^{n+1}$ should thus take into account the effect of the evolution operator $\mathcal{E}$ on the sparse expansion (49), integrated between $n\Delta t$ and $(n+1)\Delta t$. Once a procedure for the refinement of Λ^n into $\tilde{\Lambda}^{n+1}$ has been prescribed, several strategies are available for computing $u_{\tilde{\Lambda}^{n+1}}$ from u_{Λ^n}, such as Petrov-Galerkin methods in Maday, Perrier and Ravel, 1991 or collocation methods in Bertoluzza, 1997. All these strategies are based on the computation of the inner products $\langle \mathcal{E}(u_{\Lambda^n}), \tilde{\psi}_\lambda \rangle$ for $\lambda \in \tilde{\Lambda}^{n+1}$ up to some precision. In the case where the evolution operator $\mathcal{E}$ is linear, this amounts to a matrix-vector product, and one can make use of the sparse multiplication algorithm which will be discussed in § 6. However, in many cases of interest, the evolution operator $\mathcal{E}$ is nonlinear, making this computation more difficult and costly. Generally speaking, the discretization of nonlinear operators is a less simple task in the wavelet coefficient domain than in the physical domain.

In the following, we shall present a systematic approach which allows us to solve this problem, by a suitable combination of the representations of the numerical solution by its wavelet coefficients and its physical values such as cell-averages. This approach was first advocated by Ami Harten, Harten, 1993; Harten, 1995. The idea of Harten is to use multiscale decompositions where they do well - namely in finding the discontinuities in the solution at a given time, and to use classical finite volume solvers, based on cell averages for the evolution step according to (47) and (51), since the properties of these solvers are well understood. To accomplish this, we need to build cell averages into our multiscale structure. This is easily accomplished (as is detailed below) by using multiscale bases that use characteristic functions of cells as the dual scaling functions. This means that at any given time step one can view the numerical solution through one of two microscopes. The one is the decomposition as a sum of characteristic functions of cells (on the finest level of decomposition); the other is the multiscale decomposition. The first is good for the evolution; the second is good for identifying shocks and regions of smoothness. As described in §2.3, there are fast methods for transforming between the two sets of coefficients (scaling coefficients and multiscale coefficients).

Let us first amplify on our claim that cell averages lend themselves naturally to multiresolution techniques based on multilevel bases as described in Section 2.3. In fact, given a hierarchy of nested meshes and corresponding partitions $(\mathcal{P}_j)_{j \geq 0}$ of the flow domain, the cell averages u_C correspond directly to inner products $\langle u, \chi_C \rangle / |C|$ which suggests that the

Encyclopedia of Computational Mechanics. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

L_1–normalized functions $\chi_C/|C|$ for $C \in \mathcal{P}_j$ play the role of the dual scaling functions $\tilde{\phi}_{j,k}$. In complete analogy to (4) the indicator functions $\chi_C/|C|$ satisfy two–scale relations. The prediction operators have then the form (20) based on the refinement matrices $\mathbf{M}_{j,0}$. Similarly to (4) one can construct Haar–like orthogonal bases ψ_λ. Here, as we have described in §2.3, one has a choice in the construction of the complement bases. We shall see that there is an advantage in having more vanishing moments in the dual basis then is provided by the classical Haar decomposition. Since Haar type bases have only first order vanishing moments, recall (12), one cannot expect a significant data compression. Therefore, one can use (22), (23) (with $\mathbf{K} = \mathbf{I}$) to raise the order of vanishing moments of the dual wavelets $\tilde{\psi}_\lambda$ as explained in (24) at the end of Section 2.3, see Dahmen, Gottschlich-Müller and Müller, 2001; Müller, 2003. This amounts to changing the primal multiresolution spaces and in turn the primal wavelets ψ_λ while the dual scaling functions remain defined as $\chi_C/|C|$.

Given any array $\mathbf{v}_J = (v_C)_{C\in\mathcal{P}_J}$ of cell averages on the partition $\mathcal{P}_J$, we can transform this vector into the multiscale format $\boldsymbol{d}^J := (\mathbf{v}_0, \boldsymbol{d}_0, \ldots, \boldsymbol{d}_{J-1})$, where the arrays $\boldsymbol{d}_j$ encode detail information needed to update the coarse cell averages in $\mathbf{v}_j$ to $\mathbf{v}_{j+1}$ on the next level of resolution. Thus the $\boldsymbol{d}$ vectors correspond to the multiscale coefficients. It is important to note that in generating the multiscale coefficients, we do not need explicit information on the multiscale basis functions. The transformation $\mathbf{T}_J$ that maps a cell average vector $\mathbf{v}_J$ on to its multiscale decomposition $\boldsymbol{d}^J$, can be executed in an entirely discrete way as in the cascade algorithm of §2.3 (see also Arandiga, Donat and Harten, 1998; Arandiga, Donat and Harten, 1999; Carnicer, Dahmen and Peña, 1996; Sweldens, 1996; Sweldens, 1998). To go the other way, from multiscale coefficients to the scaling coefficients, we again use the cascade structure. Recall that scaling coefficients $\mathbf{v}_{j+1}$, for a resolution level $j+1$, are obtained from the scaling coefficients $\mathbf{v}_j$ at the coarser level in a two step process. The first is to predict $\mathbf{v}_{j+1}$ by some rule (the lifting rule) and the second is to correct for the deviation in the prediction from the actual values. The deviation of the true coefficients $\mathbf{v}_{j+1}$ from the true coefficients is given by the detail $\boldsymbol{d}_{j+1}$.

Our adaptive numerical scheme will be designed as a combination of a *reference* finite volume scheme which operates at the finest resolution level J according to

$$\mathbf{u}_J^{n+1} = \mathbf{E}_J \mathbf{u}_J^n, \tag{51}$$

and of the transformation $\mathbf{T}_J$ and $\mathbf{T}_J^{-1}$ that relate the cell-average vector $\mathbf{u}_J^n$ and its multiscale coefficients $(d_\lambda^n)_{|\lambda|\le J-1}$. In an adaptive context, we want to encode only a small relevant portion of this vector corresponding to the adaptive set Λ^n. Ideally, this set would correspond to the indices λ such that

$$\|d_\lambda^n \psi_\lambda\|_{\mathcal{X}} > \eta, \tag{52}$$

where $\|\cdot\|_{\mathcal{X}}$ is the norm in which we plan to measure the error and η is some prescribed threshold. In practice, we precisely want to avoid the encoding of $\mathbf{u}_J^n$ and of the full multiscale vector, and therefore we cannot invoke a thresholding procedure applied to the reference numerical solution. Therefore, we shall develop an adaptive strategy that iteratively computes some η–significant sets Λ^n and multiscale coefficients $(d_\lambda^n)_{\lambda\in\Lambda^n}$ which might differ from those obtained by thresholding $\mathbf{u}_J^n$. One of our goals is to keep track of the error between $\mathbf{u}_J^n$ and the adaptive solution $\mathbf{v}_J^n$ which is defined as the reconstruction on the finest partition $\mathcal{P}_J$ from the details $(d_\lambda^n)_{\lambda\in\Lambda^n}$.

At this stage, a key observation is that a restricted multiscale vector $(d_\lambda)_{\lambda\in\Lambda}$ exactly encodes the cell averages on an *adaptive partition* $\mathcal{P}(\Lambda)$ which includes cells of different resolution levels

Encyclopedia of Computational Mechanics. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

$j = 0, \cdots, J$ as illustrated in Figure 4, provided that the set Λ has a *graded tree structure*. Such a tree structure ensures that all the detail coefficients which are necessary to reconstruct the exact average on a cell $C \in \mathcal{P}(\Lambda)$ are contained in Λ. Note that a graded tree structure is not guaranteed on an adaptive set Λ produced by a thresholding procedure, yet it can be ensured by a suitable enlargement of Λ. In addition, it can be seen that the *graded tree structure* induces a *grading property* on the partition $\mathcal{P}(\Lambda)$ which essentially means that two adjacent cells differ at most by one resolution level. The concepts of tree structure and grading will also serve in §6 in the context of nonlinear variational problems. Another key observation is that the cost of the transformation $\mathbf{T}_\Lambda$ which maps the cell-averages $(u_C)_{C\in\mathcal{P}(\Lambda)}$ onto the restricted multiscale vector $(d_\lambda)_{\lambda\in\Lambda}$ is of the order $\#(\Lambda)$ and similarly for the inverse transformation $\mathbf{T}_\Lambda^{-1}$. A detailed description of such techniques and the design of appropriate data structures can be found in Müller, 2003.

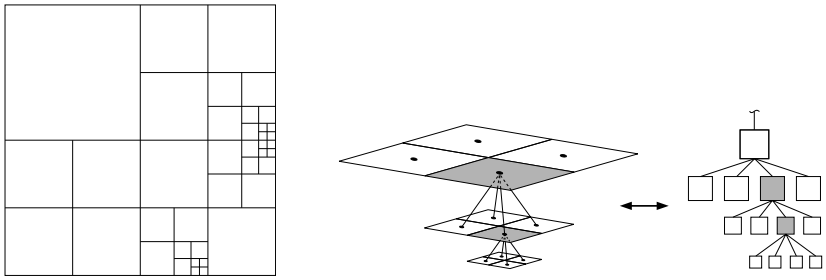


Figure 4. a) Adaptive mesh b) Tree

Based on this observation, we can propose the following adaptive scheme which follows the same principles as the dynamically adaptive scheme introduced in Maday, Perrier and Ravel, 1991:

(i) Initial values: Apply the multiscale transform $\mathbf{T}_J$ to the initial cell averages $\mathbf{u}_J^0$ to obtain the array of detail or wavelet coefficients $(d_\lambda^0)_{|\lambda|\le J}$ (including the cell averages on the coarsest level) for the time level $n = 0$. Choose a threshold parameter $\eta > 0$ and set Λ^0 to be the smallest graded tree containing those λ such that $\|d_\lambda^0\psi_\lambda\|_{\mathcal{X}} > \eta$.

(ii) Predicting the significant indices on the next time level: Given the η–significant tree Λ^n for the time level n and the details $(d_\lambda^n)_{\lambda\in\Lambda^n}$, *predict* a set $\tilde{\Lambda}^{n+1}$ that should contain the η–significant graded tree for time level $n+1$. We extend the detail vector by setting $d_\lambda^n = 0$ for $\lambda \in \tilde{\Lambda}^{n+1} \setminus \Lambda^n$ and we derive the cell-averages $(v_C^n)_{C\in\mathcal{P}(\tilde{\Lambda}^{n+1})}$ by applying the adaptive multiscale transform $\mathbf{T}_{\tilde{\Lambda}^{n+1}}^{-1}$.

(iii) Time evolution step: Compute the evolved cell-averages $(v_C^{n+1})_{C\in\mathcal{P}(\tilde{\Lambda}^{n+1})}$ at time $n+1$, by some discrete evolution operator to be specified later. Of course it is important that this evolution can be done at less cost than would be necessary to evolve the uncompressed data.

(iv) Reverse transform and thresholding: Apply the localized transform $\mathbf{T}_{\tilde{\Lambda}^{n+1}}$ to

$(v_C^{n+1})_{C\in\mathcal{P}(\tilde{\Lambda}^{n+1})}$ which yields an array of detail coefficients $(d_\lambda^{n+1})_{\lambda\in\tilde{\Lambda}^{n+1}}$. Set Λ^{n+1} to be the smallest graded tree containing those λ such that $\|d_\lambda^{n+1}\psi_\lambda\|_{\mathcal{X}} > \eta$. Set $n+1$ to n and go to (ii).

Any concrete realization of this scheme has to address the following issues.

(1) Choice of the norm $\|\cdot\|_{\mathcal{X}}$ and of the threshold parameter η;
(2) Strategy for predicting the set $\tilde{\Lambda}^{n+1}$;
(3) Specification of the evolution operator.

Regarding (1), since the norm $\|\cdot\|_{\mathcal{X}}$ will typically measure the deviation between the reference and adaptive solution $\mathbf{u}_J^n$ and $\mathbf{v}_J^n$, a relevant choice for this norm should be such that we already have at our disposal an error estimate between the reference solution $\mathbf{u}_J^n$ and the exact solution at time $n\Delta t$ in the same norm. As we shall see further, it will also be important that the reference scheme is stable with respect to such a norm. In the context of conservation laws, this limits us to the choice $\mathcal{X} = L_1$. For a discrete vector $\mathbf{u}_J$ indexed by the finest partition $\mathcal{P}_J$, we define $\|\mathbf{u}_J\|_{L^1}$ as the L_1 norm of the corresponding piecewise constant function on $\mathcal{P}_J$, i.e.

$$\|\mathbf{u}_J\|_{L_1} := \sum_{C\in\mathcal{P}_J} |C|\,|u_C|. \tag{53}$$

Assuming that the ψ_λ are normalized in L_1, it follows from the triangle inequality that the error e_η produced by discarding those multiscale coefficients of $\mathbf{u}^J$ satisfying $\|d_\lambda\psi_\lambda\|_{L_1} \le \eta$ is bounded by

$$e_\eta \le \sum_{\|d_\lambda\psi_\lambda\|_{L_1}\le\eta} \eta = \eta\,\#\{\lambda : \|d_\lambda\psi_\lambda\|_{L_1} \le \eta\}. \tag{54}$$

Since the above sum is limited to $|\lambda| \le J-1$, we can derive the estimate

$$e_\eta \le \#(\mathcal{P}_J)\eta \;\lesssim\; 2^{dJ}\eta, \tag{55}$$

where d is the spatial dimension of the problem. It follows that a prescribed thresholding error δ can be obtained by using a threshold of the order

$$\eta \sim 2^{-dJ}\delta. \tag{56}$$

Since the dual scaling functions and wavelets are normalized in L_1, the primal scaling functions and wavelets are normalized in L_∞ so that $\|\psi_\lambda\|_{L^1} \sim 2^{-d|\lambda|}$. Therefore, the above strategy corresponds to applying to the coefficients d_λ a level dependent threshold $\eta_{|\lambda|}$ with

$$\eta_j \sim 2^{d(j-J)}\delta. \tag{57}$$

Note however that the estimate (55) is somehow pessimistic since some thresholded coefficients d_λ could actually be much smaller than $\eta_{|\lambda|}$.

Concerning (2), an ideal prediction should take into account the action of the reference scheme $\mathbf{E}_J$ on the adaptive solution in the sense that the detail coefficients of $\mathbf{E}_J\mathbf{v}_J^n$ which are not contained in $\tilde{\Lambda}^{n+1}$ are guaranteed to be below the threshold. A strategy for constructing $\tilde{\Lambda}^{n+1}$ was proposed by Harten, based on a heuristic argument concerning the finite propagation speed of information in hyperbolic problems. Basically, $\tilde{\Lambda}^{n+1}$ is formed as the union of certain fixed neighborhoods (on the same or at most one higher scale) of the elements in Λ^n. Recently, at least for scalar problems a rigorous analysis has been presented in Cohen, Kaber, Müller and

Encyclopedia of Computational Mechanics. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

Postel, 2002 which gives rise to sets $\tilde{\Lambda}^{n+1}$ that are *guaranteed* to fullfill the above prescription. In this case the neighborhoods are allowed to depend in a more precise way on the size of the elements in Λ^n. In practical experiments Harten's simpler choice seems to have worked so far well enough though.

Turning to (3), several strategies are available for evolving the cell averages $(v_C^n)_{C\in\mathcal{P}(\tilde{\Lambda}^{n+1})}$ into $(v_C^{n+1})_{C\in\mathcal{P}(\tilde{\Lambda}^{n+1})}$. The first one consists in computing the effect on these averages of the exact application of the reference scheme $\mathbf{E}_J$ to the adaptive solution $\mathbf{v}_J^n$ reconstructed on the fine grid. A key observation is that since we are only interested in the averages of $\mathbf{E}_J\mathbf{v}_J^n$ on the adaptive partition $\mathcal{P}(\tilde{\Lambda}^{n+1})$, the numerical fluxes which need to be computed are only those between the adjacent fine cells such that their interface lies on the edge of a cell of the adaptive partition. In the original concept proposed by Harten, this idea was exploited in order to obtain CPU savings on the number of flux evaluation, with the solution encoded in its non-adaptive form $\mathbf{v}_J^n$. In Cohen, Kaber, Müller and Postel, 2002, it was shown that the computation of the needed fluxes can be performed from the adaptive data $(v_C^{n+1})_{C\in\mathcal{P}(\tilde{\Lambda}^{n+1})}$ without the need of performing the reconstruction of the entire $\mathbf{v}_J^n$. This information can indeed be acquired by local reconstruction. However, in several space dimensions the resulting computational complexity, although still lower than that for the fully refined partitions, is suboptimal. A second more economical strategy is to employ the finite volume stencil of the uniform partition but for the currently local level of resolution. This makes use of the local quasi–uniformity of the mesh which can be made locally uniform by subdividing neighboring cells of lower generation. The gradedness of the partitions ensures that the subdivisions need only have depth one. In numerical experiments this strategy turns out to work well when using higher order finite volume schemes in connection with corresponding higher order multiscale decompositions, here corresponding to the higher order vanishing moments, see e.g. Müller, 2003.

One of the nice features of the adaptive approach that we have described is the possibility to monitor the error between the reference and adaptive numerical solution by a proper tuning of the threshold parameter. Here, we consider the evolution strategy that amounts in computing exactly the averages of $\mathbf{E}_J\mathbf{v}_J^n$ on the adaptive partition $\mathcal{P}(\tilde{\Lambda}^{n+1})$. It follows that we can write

$$\|\mathbf{u}_J^{n+1}-\mathbf{v}_J^{n+1}\|_{L_1}=\|\mathbf{E}_J\mathbf{u}_J^n-\mathbf{E}_J\mathbf{v}_J^n\|_{L_1}+p_n+t_n, \tag{58}$$

where

$$p_n:=\sum_{\lambda\notin\tilde{\Lambda}^{n+1}}\|d_\lambda(\mathbf{E}_J\mathbf{v}_J^n)\psi_\lambda\|_{L_1} \tag{59}$$

and

$$t_n:=\sum_{\lambda\in\tilde{\Lambda}^{n+1}\setminus\Lambda^{n+1}}\|d_\lambda(\mathbf{E}_J\mathbf{v}_J^n)\psi_\lambda\|_{L_1} \tag{60}$$

respectively denote the errors resulting from the restriction to the predicted set $\tilde{\Lambda}^{n+1}$ and to the set Λ^{n+1} obtained by thresholding. According to our previous remarks, these errors can be controlled by some prescribed δ provided that we use the level dependent threshold $\eta_j\sim 2^{d(j-J)}\delta$. Assuming in addition that the reference scheme is L_1-stable in the sense that for all $\mathbf{u}_J$ and $\mathbf{v}_J$,

$$\|\mathbf{E}_J\mathbf{u}_J-\mathbf{E}_J\mathbf{v}_J\|_{L_1}\le(1+C\Delta t)\|\mathbf{u}_J-\mathbf{v}_J\|_{L_1}, \tag{61}$$

we thus obtain

$$\|\mathbf{u}_J^{n+1}-\mathbf{v}_J^{n+1}\|_{L_1}\le\|\mathbf{E}_J\mathbf{u}_J^n-\mathbf{E}_J\mathbf{v}_J^n\|_{L_1}+2\delta, \tag{62}$$

Encyclopedia of Computational Mechanics. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

which yields the estimate at time $T = n\Delta t$

$$\|\mathbf{u}_J^n - \mathbf{v}_J^n\|_{L_1} \leq C(T) n\delta. \tag{63}$$

Therefore, if the reference numerical scheme is known to provide accuracy $\epsilon = \epsilon_J$ on level J, it is natural to choose δ such that $n\delta \sim \epsilon$. In many practical instances, however, this estimate turns out to be too pessimistic in the sense that thresholding and refinement errors do not really accumulate with time, so that δ and the threshold η can be chosen larger than the value prescribed by this crude analysis. A sharper analysis of the error between the adaptive and reference solution is still not available.

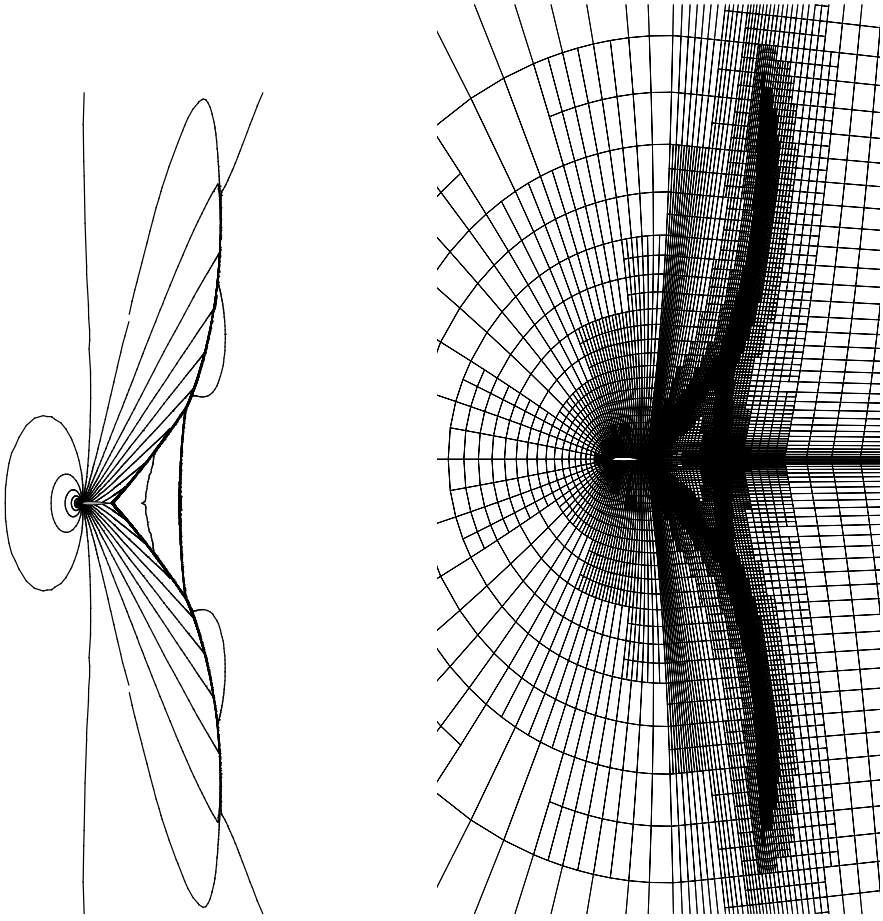


Figure 5. a) Pressure distribution b) Adaptive mesh

Encyclopedia of Computational Mechanics. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

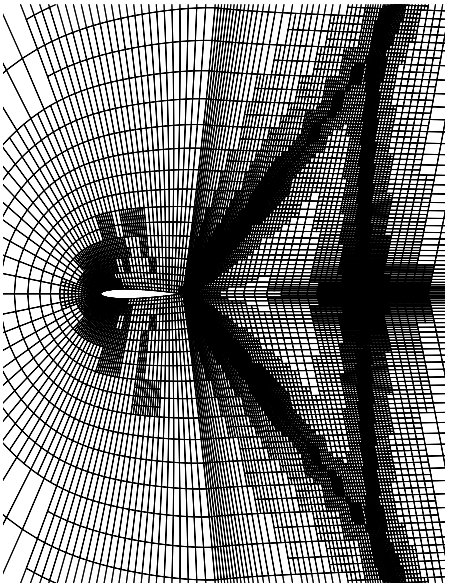


Figure 6. Adaptive mesh – close up

An adaptive solver based on the above concepts has been developed and implemented by S. Müller. It also incorporates implicit time discretizations. A detailed account can be found in Müller, 2003. The recently developed new flow solver QUADFLOW for hyperbolic conservation laws and for the Navier Stokes equations for compressible flows is based on these adaptive multiresolution techniques, on a finite volume discretization that can cope with hanging nodes and on a mesh generator based on block partitions. Each block corresponds to a B-spline based parametric mapping that allows a flexible mesh refinement through point evaluations of the B-spline representation. An outline of the scheme and extensive numerical tests can be found in Bramkamp, Gottschlich-Müller, Hesse, Lamby, Müller, Ballmann, Brakhage and Dahmen, 2001; Ballmann, Bramkamp and Müller, 2000. The numerical examples provided by S. Müller and F. Bramkamp should give an impression of the performance of such techniques. The first example in Figure 5 shows the results for an Euler computation concerning a flow at Mach 0.95 at an angle of attack $\alpha = 0$ around a bench mark NACA00012 profile. Here the main objective is to test the resolution of shock interactions even at a large distance from the airfoil. The mesh has appr. 5×10^4 cells as opposed to an estimated number of 7×10^7 cells needed by a uniformly refined mesh for a comparable target accuracy.

Figure 7 shows a series of adaptive refinements again for an Euler computation for a flow around a BAC 3-11/RES/30/21-Profile at $M = 0.85$ and an angle of attack $\alpha = 0°$. This test illustrates the reliable detection even of small shocks here in the lower region of the nose. Further detailed numerical studies for instationary problems such as moving wings, shock-bubble interactions, analogous studies for viscous flows and boundary layer resolution can be found in Bramkamp, Gottschlich-Müller, Hesse, Lamby, Müller, Ballmann, Brakhage and Dahmen, 2001; Ballmann, Bramkamp and Müller, 2000; Müller, 2003.

Encyclopedia of Computational Mechanics. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

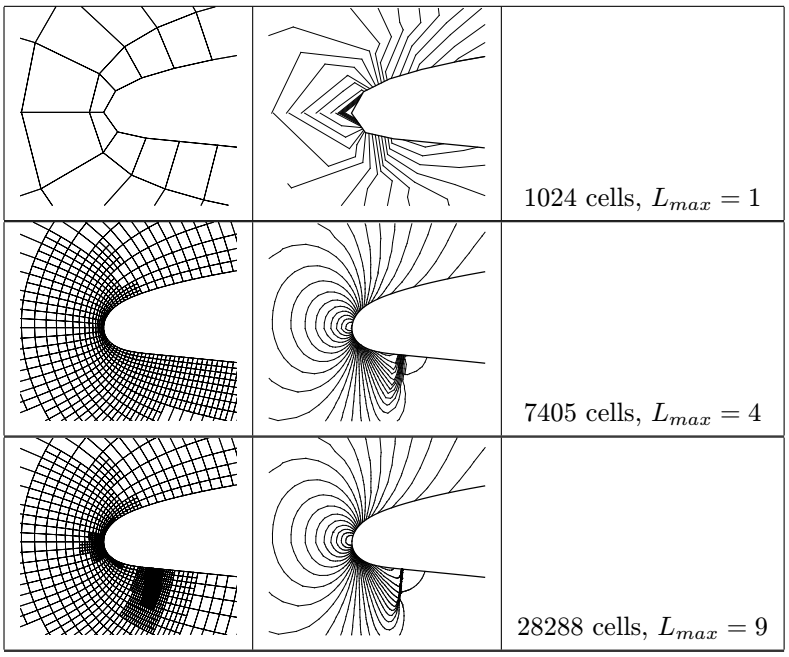


Figure 7. Adaptive mesh refinement

3.1. Concluding Remarks

The above framework is an example where multiscale bases for realistic geometries are conveniently realized. In spite of the promising numerical results it should be stressed though that many principal questions remain open, due to a number of obstructions. First, the current understanding of error analysis for hyperbolic problems is much less developed than for elliptic problems, partly due to the nature of the relevant function spaces. On one hand, there are only poor a-priori estimates that could serve as a bench mark. The central point of view is a *perturbation analysis*. The overall attainable accuracy is fixed by the a–priori choice of a highest level J of spatial resolution. All subsequent attempts aim at preserving the accuracy offered by a uniformly refined discretization with that resolution at possibly low cost. Thus whatever information is missed by the reference scheme cannot be recovered by the above adaptive solver.

On the other hand, the multiscale techniques did not unfold their full potential: one makes use of the cancellation properties of wavelet bases but not of the norm equivalences between wavelet coefficients and functions. Thus the primary focus here is on the compression of the conserved variables, i.e. on the *sparse approximation* of functions based on cancellation properties. This does so far not provide any estimates that relate the achieved accuracy ϵ to the size of the η–significant trees. This question will be addressed later again in a different context where stronger basis properties allow one to exploit not only the sparse approximation of functions but also of the involved operators.

Encyclopedia of Computational Mechanics. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

4. Boundary Integral Equations – Matrix Compression

A variety of problems in elasticity, fluid flow or electro-magnetism lead to a formulation in terms of *boundary integral equations* falling into the category (2). In principle, this is a feasible approach when the Green's function of the underlying (linear) partial differential equation is known explicitly. This is a particularly tempting option when the original formulation via a PDE refers to an exterior and hence unbounded domain since the corresponding boundary integral formulation lives on a compact manifold of lower spatial dimension. Wavelet concepts have had a significant impact on this problem area, see Dahmen, Harbrecht and Schneider, 2002; Dahmen, Prößdorf and Schneider, 1994; Harbrecht, 2001; Lage and Schwab, 1998; Lage, 1996; von Petersdorff and Schwab, 1996; von Petersdorff and Schwab, 1997; Schneider, 1998, primarily in finding sparse and efficient approximations of potential operators. We shall describe this in the following simple setting.

4.1. Classical Boundary Integral Equations

Let Ω^- be again a bounded domain in $\mathbb{R}^d$ ($d \in \{2,3\}$) and consider Laplace's equation

$$-\Delta w = 0, \text{ on } \Omega, \quad (\Omega = \Omega^- \text{ or } \Omega^+ := \mathbb{R}^3 \setminus \Omega^-), \tag{64}$$

subject to the boundary conditions

$$w = f \quad \text{on } \Gamma := \partial\Omega^- \quad (w(x) \to 0,\ |x| \to \infty \text{ when } \Omega = \Omega^+). \tag{65}$$

Of course, the unbounded domain Ω^+ poses an additional difficulty in the case of such an *exterior* boundary value problem. A well-known strategy is to transform (64), (65) into a *boundary integral equation* that lives only on the manifold $\Gamma = \partial\Omega$. There are several ways to do that. They all involve the fundamental solution $\mathcal{E}(x,y) = 1/4\pi|x-y|$ of the Laplace operator which gives rise to the *single layer potential* operator

$$(\mathcal{A}u)(x) = (\mathcal{V}u)(x) := \int_\Gamma \mathcal{E}(x,y)u(y)d\Gamma_y, \quad x \in \Gamma. \tag{66}$$

One can then show that the solution u of the first kind integral equation

$$\mathcal{V}u = f \quad \text{on } \Gamma \tag{67}$$

provides the solution w of (64) through the representation formula

$$w(x) = \int_\Gamma \mathcal{E}(x,y)u(y)d\Gamma_y, \quad x \in \Omega. \tag{68}$$

An alternative way uses the *double layer potential*

$$(\mathcal{K}v)(x) := \int_\Gamma \frac{\partial}{\partial n_y}\mathcal{E}(x,y)v(y)d\Gamma_y = \int_\Gamma \frac{1}{4\pi}\frac{n_y^T(x-y)}{|x-y|^3}v(y)\,d\Gamma_y, \quad x \in \Gamma, \tag{69}$$

where n_y is the outward normal to Γ at $y \in \Gamma$. Now the solution of the second kind integral equation

$$\mathcal{A}u := (\frac{1}{2} \pm \mathcal{K})u = f \quad (\Omega = \Omega^\pm) \tag{70}$$

Encyclopedia of Computational Mechanics. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

gives the solution to (64) through

$$w(x) = \int_\Gamma \mathcal{K}(x,y)u(y)d\Gamma_y. \tag{71}$$

One way of reformulating (64) with Neuman boundary conditions $\partial w/\partial n = g$ on Γ, where $\int_\Gamma g(x)d\Gamma_x = 0$, is offered by the so called *hypersingular operator* $(\mathcal{W}v)(x) := -\frac{\partial}{\partial n_x}\int_\Gamma \frac{\partial}{\partial n_y}\mathcal{E}(x,y)d\Gamma_y$. Now the solution of

$$\mathcal{A}u = \mathcal{W}u = g \quad \text{on} \quad \Gamma \tag{72}$$

leads to the solution of the Neuman problem for Laplace's equation through the representation formula (71), where the constraint $\int_\Gamma u(x)d\Gamma_x = 0$ is imposed in the case of an interior problem to ensure uniqueness.

4.2. Quasi–Sparsity of Wavelet Representations

We have encountered so far two classes of operators. The first case such as (36) concerns differential operators which are local in the sense that $\langle \psi_\lambda, \mathcal{A}\psi_\nu\rangle = 0$ whenever $S_\lambda \cap S_\nu = \emptyset$. Note that the wavelet representation $\boldsymbol{A}$ is not sparse in the classical sense since basis functions from different levels may overlap. In fact, it is easy to see that the number of entries in a principal section of $\boldsymbol{A}$ of size N contains $O(N\log N)$ entries. However, we shall see that many of these entries are so small in modulus that they can be neglected in a matrix vector multiplication without perturbing the result too much. The point of focus in this section is that this even holds true for the second class of *global* operators, which are roughly speaking inverses of differential operators such as the above boundary integral operators. They all share the property that they (or at least their global part) is of the form

$$(\mathcal{A}u)(x) = \int_\Gamma K(x,y)u(y)\,d\Gamma_y, \tag{73}$$

where for a given domain or manifold Γ the kernel $K(x,y)$ is smooth except on the diagonal $x = y$ and satisfies the decay conditions.

$$\left|\partial_x^\alpha \partial_y^\beta K(x,y)\right| \lesssim \operatorname{dist}(x,y)^{-(d+2t+|\alpha|+|\beta|)} \tag{74}$$

By (39), the entries of $\boldsymbol{A}$ are in this case given by

$$\boldsymbol{A}_{\lambda,\nu} = \langle K, \psi_\lambda \otimes \psi_\nu\rangle_{\Gamma\times\Gamma} = \int_\Gamma\int_\Gamma K(x,y)\psi_\lambda(x)\psi_\nu(y)d\Gamma_x d\Gamma_y. \tag{75}$$

Although none of the entries $\boldsymbol{A}_{\lambda,\nu}$ will generally be zero many of them are very small in modulus as specified by the following classical estimate, see e.g. Dahmen, Prößdorf and Schneider, 1994; von Petersdorff and Schwab, 1996; von Petersdorff and Schwab, 1997.

Theorem 4.1. *Suppose that the kernel K is of the above form and that $\mathbf{D}^{-s}\Psi$ is a Riesz-basis for H^s for $-\tilde{\gamma} < s < \gamma$ (see (32)) and has cancellation properties (see (26)) of order $\tilde{m}$. Moreover, assume that $\mathcal{A}$ given by (73) has order $2t$ and satisfies for some $r > 0$*

$$\|\mathcal{A}v\|_{H^{-t+a}} \lesssim \|v\|_{H^{t+a}}, \quad v \in H^{t+a}, 0 \le |a| \le r. \tag{76}$$

Encyclopedia of Computational Mechanics. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

Then, for any $\sigma > 0$ *such that* $0 < \sigma \leq \min\{r, d/2 + \tilde{m} + t\}$, $t + \sigma < \gamma$, *and* $t - \sigma > -\tilde{\gamma}$, *one has*

$$2^{-(|\nu|+|\lambda|)t}|\langle \psi_\lambda, \mathcal{A}\psi_\nu\rangle| \lesssim \frac{2^{-||\lambda|-|\nu||\sigma}}{(1+2^{\min(|\lambda|,|\nu|)}\operatorname{dist}(S_\lambda, S_\nu))^{d+2\tilde{m}+2t}}. \tag{77}$$

Thus the entries of the wavelet representation of operators of the above type exhibit a polynomial spatial decay, depending on the order of cancellation properties, and an exponential scalewise decay, depending on the regularity of the wavelets.

For a proof of such estimates one distinguishes two cases. When $\operatorname{dist}(S_\lambda, S_\nu) \lesssim 2^{-\min(|\lambda|,|\nu|)}$ one can use the continuity properties (76) in combination with the norm equivalences (32) to show that

$$|\langle \psi_\lambda, \mathcal{A}\psi_\nu\rangle| \leq 2^{t(|\lambda|+|\nu|)}2^{\sigma(|\nu|-|\lambda|)}, \tag{78}$$

see Dahlke, Dahmen, Hochmuth and Schneider, 1997 for more details.

On the other hand, when $\operatorname{dist}(S_\lambda, S_\nu) \gtrsim 2^{-\min(|\lambda|,|\nu|)}$, the wavelets are integrated against smooth parts of the kernel K. One can then exploit the cancellation properties for both wavelets to obtain the bound

$$|\langle \psi_\lambda, \mathcal{A}\psi_\nu\rangle| \lesssim \frac{2^{-(|\lambda|+|\nu|)(d/2+\tilde{m})}}{(\operatorname{dist}(S_\lambda, S_\nu))^{d+2\tilde{m}+2t}}, \tag{79}$$

see e.g. Dahmen, Prößdorf and Schneider, 1994; Dahmen and Stevenson, 1999; von Petersdorff and Schwab, 1996; von Petersdorff and Schwab, 1997 for more details. The decay estimate (77) follows then from (78) and (79).

However, the above argument for the case of overlapping supports is rather crude. Instead one can use the so called *second compression* due to Schneider, Schneider, 1998. In fact, when $|\lambda| \gg |\nu|$ and when S_λ does not intersect the singular support of ψ_ν then $\mathcal{A}\psi_\nu$ is smooth on the support of ψ_λ and one can again use the cancellation property of ψ_λ. Denoting by S'_ν the singular support of (the lower level wavelet) ψ_ν, this leads to

$$|\langle \psi_\lambda, \mathcal{A}\psi_\nu\rangle| \lesssim \frac{2^{|\lambda|/2}2^{-|\nu|(\tilde{m}+d/2)}}{\operatorname{dist}(S'_\nu, S_\lambda)^{2t+\tilde{m}}}. \tag{80}$$

Estimates of the type (77), (78) and (80) provide the basis of *matrix compression* strategies that aim at replacing the wavelet representation of a operator by a sparsified perturbation which can be used to expedite the numerical solution of corresponding linear systems.

4.3. Weak Formulations and Galerkin Schemes

As in the case of Poisson's equation (36) we are dealing again with an operator equation

$$\mathcal{A}u = f \tag{81}$$

this time of the type (66), (70) or (72). A classical approach to solving such an equation numerically is to return again to a proper weak formulation on which to base a Galerkin discretization. As before, the key is to identify first a suitable (Hilbert) space $\mathcal{H}$ such that the variational formulation

$$a(v,u) := \langle v, \mathcal{A}u\rangle = \langle v, f\rangle \quad \text{for all} \quad v \in \mathcal{H}, \tag{82}$$

is well–posed in the sense of (40). In terms of the operator $\mathcal{A}$ this can be rephrased by saying that $\mathcal{A}$ is boundedly invertible as a mapping from $\mathcal{H}$ onto $\mathcal{H}'$ which will often be referred to as *mapping property*.

All the above examples can be shown to fall into this framework, see e.g. Kress, 1989. The single layer potential is symmetric positive definite on the Sobolev space $\mathcal{H} := H^{-1/2}(\Gamma)$ whose dual is $\mathcal{H}' = H^{1/2}(\Gamma)$, i.e.,

$$a(v,v) = \langle v, \mathcal{V}v\rangle \gtrsim \|v\|_{H^{-1/2}(\Gamma)} \quad \text{for all} \quad v \in H^{-1/2}(\Gamma), \tag{83}$$

which is easily seen to imply (40).

The double layer potential is known to be compact when Γ is a C^2 manifold in which case the kernel is weakly singular. In general, the approriate *energy space* is $\mathcal{H} = L_2(\Gamma) = \mathcal{H}'$. Despite the lack of symmetry one can show that the bilinear form $a(\cdot,\cdot)$ is *coercive* and that (40) holds with $\mathcal{H} = L_2(\Gamma)$.

The hypersingular operator $\mathcal{W}$, in turn, is strongly singular with energy space $\mathcal{H} = H^{1/2}(\Gamma)$ (i.e. $\mathcal{H}' = H^{-1/2}(\Gamma)$), respectively $\mathcal{H} = H^{1/2}(\Gamma)/\mathbb{R}$ in the case of an interior problem. Again since it is then symmetric positive definite and (40) follows.

According to the shifts caused by these operators in the Sobolev scale, $\mathcal{A} : H^t(\Gamma) \to H^{-t}(\Gamma)$, the single layer potential, double layer potential and hypersingular operator have *order* $2t = -1, 0, 1$, respectively.

The (conforming) Galerkin method (for any operator equation (64)) consists now in choosing a finite dimensional space $S \subset \mathcal{H}$ and determining $u_S \in S$ such that

$$a(v, u_S) = \langle v, f\rangle \quad \text{for all} \quad v \in S. \tag{84}$$

Such a scheme is called ($\mathcal{H}$–)stable (for a family $\mathcal{S}$ of increasing spaces $S \in \mathcal{S}$) if (40) holds on the discrete level, uniformly in $S \in \mathcal{S}$. In other words, denoting by P_S any $\mathcal{H}$-bounded projector onto S, we need to ensure that (40) holds with $\mathcal{A}$ replaced by $P_S'\mathcal{A}P_S$, uniformly in $S \in \mathcal{S}$, where P_S' is the adjoint of P_S. This is trivially the case for any subsapce $S \subset \mathcal{H}$ when $\mathcal{A}$ is symmetric positive definite. In the coercive case, one can show that Galerkin discretizations are stable for families $\mathcal{S}$ of trial spaces that satisfy certain approximation and regularity properties formulated in terms of direct and inverse estimates, whenever the level of resolution is fine enough, see e.g. Dahmen, Prößdorf and Schneider, 1994.

Once this homework has been done, it remains to choose a basis for S by which (84) is turned into a linear system of equations. The unknowns are the coefficients of u_S with respect to the chosen basis.

The obvious advantage of the boundary integral approach is the reduction of the spatial dimension and that one has to discretize in all cases only on bounded domains. On the other hand, one faces several obstructions:

(i) Whenever the order of the operator is different from zero (e.g. for $\mathcal{A} = \mathcal{V}$), the problem of *growing condition numbers* arises because the operator treats high frequency components differently from slowly varying ones. In general, if an operator has order $2t$, the spectral condition numbers of the stiffness matrices grow like $h^{-2|t|}$, where h reflects the spatial resolution (e.g. the mesh size) of the underlying discretization.

(ii) Discretizations lead in general to *densely populated matrices*. This severly limits the number of degrees of freedom when using direct solvers. But iterative techniques are also problematic, due to the fact that the cost of each matrix/vector multiplication increases with the square of the problem size.

One possible strategy to overcome these obstructions will be outlined next.

Encyclopedia of Computational Mechanics. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

4.4. Wavelet–Galerkin Methods

We adhere to the above setting and consider the operator equation (81), where $\mathcal{A}$ has the form (73), (74) and satisfies (40) for $\mathcal{H} = H^t(\Gamma)$. Suppose now that we have a wavelet basis Ψ which is a Riesz basis for $H^t(\Gamma)$, constructed along the lines from Section 2.3, with the corresponding multiresolution sequence of spaces S_j spanned by all wavelets ψ_λ, $|\lambda| < j$, of level less than j.

We point out next how the use of the spaces S_j as trial spaces in Galerkin discretizations can help to cope with the above obstructions. To this end, let $\boldsymbol{A}_j := (\langle \psi_\lambda, \mathcal{A}\psi_\nu\rangle)_{|\lambda|,|\nu|<j}$ denote the stiffness matrix of $\mathcal{A}$ with respect to the (finite) wavelet basis of the trial space S_j. Thus (84) takes the form

$$\boldsymbol{A}_j \mathbf{u}_j = \mathbf{f}_j, \quad \mathbf{f}_j := (\langle \psi_\lambda, f\rangle)_{|\lambda|<j}. \tag{85}$$

The first observation concerns obstruction (i) above, see e.g. Dahmen and Kunoth, 1992; Dahmen, Prößdorf and Schneider, 1994.

Remark 4.2. *If the Galerkin discretizations are H^t–stable for $\mathcal{S} = \{S_j\}_{j\in\mathbb{N}_0}$, then the spectral condition numbers of $\boldsymbol{A}_j$ are uniformly bounded.*

In fact, when the bilinear form $a(\cdot,\cdot)$, defined in (82), is symmetric and H^t–elliptic, so that $\boldsymbol{A}$ is symmetric positive definite, the spectrum of $\boldsymbol{A}_j$ is contained in the convex hull of the spectrum of $\boldsymbol{A}$, so that the assertion follows immediately from Theorem 2.2. Since under this assumption Galerkin discretizations are always stable for any choice of subspaces this is a special case of the above claim. In the general case the argument is similar to that in the proof of Theorem 2.2. In fact, Galerkin stability means that $\|\boldsymbol{A}_j^{-1}\|_{\ell_2\to\ell_2} \lesssim 1$ which ensures the existence of a constant $\bar{c}$ such that $\bar{c}\|\mathbf{v}_j\|_{\ell_2} \le \|\boldsymbol{A}_j\mathbf{v}_j\|_{\ell_2}$ for any $v_j \in S_j$ with coefficient vector $\mathbf{v}_j$. Moreover, by (43),

$$\|\boldsymbol{A}_j\mathbf{v}_j\|_{\ell_2} = \|(\langle \psi_\lambda, \mathcal{A}v_j\rangle)_{|\lambda|<j}\|_{\ell_2} \le \|\boldsymbol{A}\mathbf{v}_j\| \le C_\Psi^2 C_{\mathcal{A}}\|\mathbf{v}_j\|,$$

which confirms the claim.

This observation applies to all our above examples of boundary integral operators. In fact, $\mathcal{V}$ and $\mathcal{W}$ are elliptic and the coercivity in the case (70) of the double layer potential ensures that (for $j \ge j_0$ large enough) the Galerkin discretizations are also stable in this case, see e.g. Dahmen, Prößdorf and Schneider, 1994.

Thus, a proper choice of wavelet bases for the respective energy space deals with obstruction (i) not only for the second kind integral equations with zero order operators but essentially for all classical potential operators. This is, for instance, important in the context of transmission problems.

Of course, the preconditioning can be exploited in the context of iterative solvers for (85) only if the cost of a matrix/vector multiplication can be significantly reduced below the square of the problem size. First, note that, due to the presence of discretization errors, it is not necessary to compute a matrix/vector multiplication *exactly* but it would suffice to *approximate* it within an accuracy tolerance that depends on the current discretization error provided by S_j. Thus, one faces the following central

Task: Replace as many entries of $\boldsymbol{A}_j$ as possible by zero so as to otain a perturbed matrix $\tilde{\boldsymbol{A}}_j$ with the following properties for all the above operator types:

(i) The $\tilde{\boldsymbol{A}}_j$ have still uniformly bounded condition numbers when the level j of resolution grows.
(ii) The solutions $\tilde{\mathbf{u}}_j$ of the perturbed systems $\tilde{\boldsymbol{A}}_j\tilde{\mathbf{u}}_j = \mathbf{f}_j$ have still the same order of accuracy as the solutions $\mathbf{u}_j$ of the unperturbed systems (85), *uniformly* in j.
(iii) Find efficient ways of computing the nonzero entries of $\tilde{\boldsymbol{A}}_j$.

These issues have been addressed in a number of investigations, see e.g. Dahmen, Prößdorf and Schneider, 1994; Dahmen, Harbrecht and Schneider, 2002; von Petersdorff and Schwab, 1996; von Petersdorff and Schwab, 1997; von Petersdorff, Schneider and Schwab, 1997; Harbrecht, 2001; Schneider, 1998. We shall briefly outline the current state of the art as reflected by Harbrecht, 2001; Dahmen, Harbrecht and Schneider, 2002. The key is a suitable level–dependent thresholding strategy based on the a–priori estimates (79) and (80). It requires a sufficiently high order of cancellation properties, namely $\tilde{m} > m - 2t$ where m is the approximation order provided by the multiresolution spaces S_j. Thus, whenever $\mathcal{A}$ has nonpositive order (such as the single and double layer potential operator), one must have $\tilde{m} > m$, ruling out orthonormal wavelets (in L_2). Given that Ψ meets this requirement, and considering a *fixed highest level* J of resolution, fix parameters $a, a' > 1$ and $m' \in (m, \tilde{m} + 2t)$ and define the cut-off parameters (see Dahmen, Harbrecht and Schneider, 2002; Harbrecht, 2001; Schneider, 1998)

$$\begin{aligned} c_{l,j} &:= a \max\left\{2^{-\min\{l,j\}}, 2^{\frac{2J(m'-t)-(j+l)(m'+\tilde{m})}{2(\tilde{m}+t)}}\right\} \\ c'_{l,j} &:= a' \max\left\{2^{-\max\{l,j\}}, 2^{\frac{2J(m'-t)-(j+l)m'-\max\{j,l\}\tilde{m}}{\tilde{m}+2t}}\right\}. \end{aligned} \tag{86}$$

Then the a–priori compression of $\boldsymbol{A}_J$ is given by

$$(\tilde{\boldsymbol{A}}_J)_{\lambda,\nu} := \begin{cases} 0, & \operatorname{dist}(S_\lambda, S_\nu) > c_{|\lambda|,|\nu|} \text{ and } |\lambda|, |\nu| \geq j_0, \\ 0, & \operatorname{dist}(S_\lambda, S_\nu) \lesssim 2^{-\min\{|\lambda|,|\nu|\}} \text{ and} \\ & \operatorname{dist}(S'_\lambda, S_\nu) > c'_{|\lambda|,|\nu|} \text{ if } |\nu| > |\lambda| \geq j_0 - 1, \\ & \operatorname{dist}(S_\lambda, S'_\nu) > c'_{|\lambda|,|\nu|} \text{ if } |\lambda| > |\nu| \geq j_0 - 1, \\ (\boldsymbol{A}_J)_{\lambda,\nu}, & \textit{otherwise.} \end{cases} \tag{87}$$

The first line is the classical "first compression" based on (79) when the wavelets have disjoint supports with a distance at least the diameter of the larger support. The number of nonzero entries that remain after this compression is of the order $N_J \log N_J$ where $N_J := \dim S_J \sim 2^{(d-1)J}$ when $d-1$ is the dimension of Γ. The second line reflects the "second compression" due to Schneider which discards entries for wavelets with overlapping support, Schneider, 1998. More importantly this affects also those entries involving the scaling functions on the coarsest level j_0. It has a significant effect when, due to a complicated geometry, the coarsest level already involves a relatively large number of basis functions. Asymptotically, it removes the log factor in the count of the nonzero entries of $\tilde{\boldsymbol{A}}_J$.

A sophisticated perturbation analysis, whose main ingredients are the a–priori estimates (79), (80) based on the cancellation properties of Ψ, the norm equivalences (30), and suitable versions of the Schur lemma, yields the following result, Dahmen, Harbrecht and Schneider, 2002; Harbrecht, 2001.

Theorem 4.3. *The compressed matrices $\tilde{\boldsymbol{A}}_J$, given by (87), have uniformly bounded condition numbers. The number of nonzero entries in $\tilde{\boldsymbol{A}}_J$ is of the order N_J, uniformly in J. Moreover, the solution $\tilde{\mathbf{u}}_J$ exhibits optimal discretization error estimates in the energy norm*

$$\|u - \tilde{u}_J\|_{H^t(\Gamma)} \lesssim 2^{J(t-m)} \|u\|_{H^m(\Gamma)}, \quad J \to \infty. \tag{88}$$

This result says that the above compression strategy is asymptotically optimal. In comparison with earlier versions the removal of log–factors even offers a strictly linear complexity. Note that for operators of negative order the relatively high computational efforts for Galerkin discretizations, due to the double integrals, pay off through the high order $m+|t|$.

The remaining crucial question concerns the complexity of computing the compressed matrices $\tilde{\boldsymbol{A}}_J$. A detailed analysis of this issue can be found in Harbrecht, 2001, see also Dahmen, Harbrecht and Schneider, 2002. The main facts can be summarized as follows.

Of course, the entries of $\tilde{\boldsymbol{A}}_J$ cannot be computed exactly but one has to resort to quadrature. The following nice observation from Harbrecht, 2001 tells us how much computational effort can be spent on the entry $(\boldsymbol{A}_J)_{\lambda,\nu}$ so as to keep the overall complexity of computing an approximation to $\tilde{\boldsymbol{A}}_J$ proportional to the system size N_J.

Theorem 4.4. *The complexity of approximately computing the nonzero entries of $\tilde{\boldsymbol{A}}_J$ is $O(N_J)$, provided that for some $\alpha > 0$ at most $O\big((J - (|\lambda| + |\nu|)/2)^\alpha\big)$ operations are spent on the computation of a nonzero coefficient $(\tilde{\boldsymbol{A}}_J)_{\lambda,\nu}$.*

Next, in analogy to the compression estimates one can ask which further perturbation is allowed for the approximate calculation of the entries of $\tilde{\boldsymbol{A}}_J$ so as to retain the above optimal convergence rates, Harbrecht, 2001; Dahmen, Harbrecht and Schneider, 2002.

Theorem 4.5. *If the quadrature error for $(\tilde{\boldsymbol{A}}_J)_{\lambda,\nu}$ is bounded by*

$$\delta \min\Big\{2^{-\frac{(d-1)||\lambda|-|\nu||}{2}}, 2^{-(d-1)(J-\frac{|\lambda|+|\nu|}{2})\frac{m'-t}{m+t}}\Big\} 2^{2Jt} 2^{-2m'\left(J-\frac{|\lambda|+|\nu|}{2}\right)}$$

for some fixed $\delta < 1$, then the (perturbed) Galerkin scheme is stable and converges with optimal order (88)

The main result now is that the accuracy bounds in Theorem 4.5 can be met by a sophisticated adapted quadrature strategy whose computational complexity remains also in the operations budget given by Theorem 4.4. Thus, in summary one obtains a fully discrete scheme that exhibits asymptotically optimal computational complexity that remains proportional to the problem size. This is illustrated below by some numerical examples.

Remark 4.6. *Another approach to matrix compression, first pointed out in Beylkin, Coifman, and Rokhlin, 1991, uses the so called* nonstandard form *of the operator $\mathcal{A}_S := P_S' \mathcal{A} P_S$, which involves a telescoping expansion for $\mathcal{A}_S$ but is not a* representation *of $\mathcal{A}_S$ in the strict sense. It consists of blocks whose entries involve only basis functions (wavelets and scaling functions) of the same level, which may simplify their computation in comparison to the standard form. On the other hand, the nonstandard form does not support preconditioning when dealing with operators of order different from zero and is therefore restricted to problems of the type (70). Due to the presence of scaling function coefficients it also does not allow us to combine matrix compression together with function compression. We shall point out later that this is indeed supported by the standard form.*

Encyclopedia of Computational Mechanics. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

4.5. Numerical Tests

The following example has been provided by H. Harbrecht. An interior Dirichlet problem for the Laplacian is solved by the indirect approach. We use both, the Fredholm integral equation of the first kind based on the single layer operator and the Fredholm integral equation of the second kind based on the double layer potential operator. Both approaches yield a density u, from which one derives the solution in the domain via potential evaluation, i.e. by applying the single layer operator and double layer operator, respctively, to the density, see (68), (71).

The domain Ω under consideration is the gearwheel shown in Figure 8. It has 15 teeth and is represented using 180 patches. As Dirichlet data we choose the restriction of the harmonic function

$$U(x) = \frac{(a,x)}{\|x\|^3}, \quad a = (1,2,4),$$

to Γ. Then, U is the unique solution of the Dirichlet problem. We discretize the given boundary integral equation by piecewise constant wavelets with three vanishing moments.

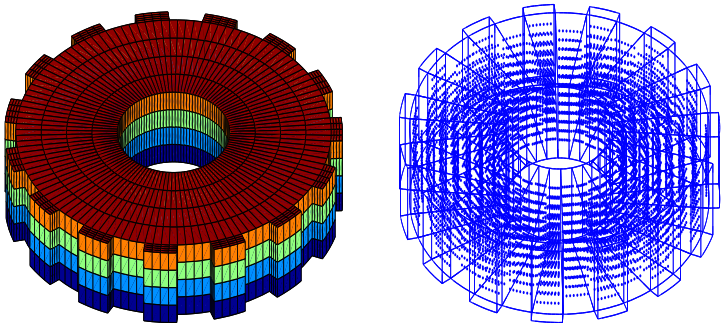


Figure 8. The surface mesh and the evaluation points x_i of the potential.

In order to measure the error produced by the method, we calculate the approximate solution $U_J = \mathcal{A}u_J$ at several points x_i inside the domain, plotted in Figure 8. The last column in the tables below reflects the effect of an a–posteriori compression applied to the computed entries of the stiffness matrix. The discrete potentials are denoted by

$$\mathbf{U} := [U(x_i)], \qquad \mathbf{U}_J := [(\mathcal{A}u_J)(x_i)],$$

where $\mathcal{A}$ stands for the single or double layer operator.

We list in Tables I and II the results produced by the wavelet Galerkin scheme. For the double layer approach, the optimal order of convergence of the discrete potential is quadratic with respect to l^∞-norm over all points x_i. For the single layer approach, this order is cubic. But one should mention that one cannot expect the full orders of convergence, due to the reentrant edges resulting from the teeth of the gearwheel.

Encyclopedia of Computational Mechanics. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

J	N_J	$\|\mathbf{U}-\mathbf{U}_J\|_\infty$	cpu-time	a-priori (%)	a-posteriori (%)
1	720	4.8e-1	1	27	7.9
2	2880	2.7e-1 (1.8)	10	8.7	2.3
3	11520	7.6e-2 (3.6)	107	3.4	0.6
4	46080	2.4e-2 (3.1)	839	1.0	0.2
5	184320	6.0e-3 (4.1)	4490	0.2	0.0

Table I. Numerical results with respect to the double layer operator.

J	N_J	$\|\mathbf{U}-\mathbf{U}_J\|_\infty$	cpu-time	a-priori (%)	a-posteriori (%)
1	720	4.9e-1	1	28	21
2	2880	5.7e-2 (8.6)	12	10	7.4
3	11520	1.2e-2 (4.5)	116	4.2	2.0
4	46080	2.8e-3 (4.5)	1067	1.3	0.5
5	184320	1.0e-3 (2.9)	6414	0.4	0.1

Table II. Numerical results with respect to the single layer operator.

4.6. Concluding Remarks

The above results show how to realize, for any (a–priori fixed) level J of resolution, a numerical scheme that solves a boundary integral equation with discretization error accuracy in linear time. As for the quantitative performance, the above examples indicate that accuracy is not degraded at all by the compression and quadrature errors. Moreover, the robustness with respect to the underlying geometry is surprisingly high. The experiences gained in Harbrecht, 2001 show that the number of basis functions on the coarsest level may go up to the square root of the overall problem size without spoiling the complexity significantly. Due to the built in preconditioning, the actual iterative solution of the linear systems is still by far dominated by the efforts for computing $\tilde{\boldsymbol{A}}_J$.

The concept is strictly based on a perturbation of the operator but makes no use of adaptivity with respect to the discretization. First tests in this direction have been made in Harbrecht, 2002. However, this is somewhat incompatible with the basic structure where all computations are tuned to an a-priori fixed highest level J of spatial resolution. Finite subsets of the wavelet basis serve to formulate a Galerkin discretization in the same way as classical settings so that no direct use is made of the full wavelet transformed representation of the boundary integral equation.

Thus, incorporating adaptivity may require an alternative to the entrywise computation of $\tilde{\boldsymbol{A}}_J$ which we shall comment on later.

Encyclopedia of Computational Mechanics. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

There are other ways of accelerating the calculation of matrix/vector products in the above context such as panel clustering (Hackbusch and Nowak, 1989), multipole expansions (Greengard and Rokhlin, 1987) or hierarchical matrices (Hackbusch, 1999). These concepts offer an even better robustness with respect to the geometry since they exploit the smoothness of the integral kernel in $\mathbb{R}^d$ and not of its trace on the manifold. However, these approaches do not allow one to build in preconditioning in such a straightforward way as above and adaptivity is even harder to incorporate. A combination of the different concepts has recently been proposed in Schmidlin, Lage and Schwab, 2002, combining the advantages of clustering and wavelet techniques.

5. A New Adaptive Paradigm

So far we have sketched essentially two different directions where the key features of wavelets listed in Section 2.4 played an essential role. In the context of boundary integral equations it was utilized that corresponding operators possess well–conditioned sparse wavelet representations. When dealing with hyperbolic conservation laws the typical piecewise smooth nature of solutions permits the compression of the flow field based on suitable thresholding strategies applied to multiscale representations of approximate solutions. In both cases an arbitrary but *fixed* level of resolution was considered and wavelet concepts were used to precondition or accelerate the numerical processing of the resulting fixed finite dimensional problem.

We shall now turn to recent developments that deviate from this line and aim at combining in a certain sense both effects, namely the sparse representation of functions and the sparse representation of (linear and nonlinear) operators. The subsequent developments are based on the results in Cohen, Dahmen and DeVore, 2001; Cohen, Dahmen and DeVore, 2002a; Cohen, Dahmen and DeVore, 2002b; Cohen, Dahmen and DeVore, 2002c; Dahlke, Dahmen and Urban, 2002

5.1. Road Map

Recall that the classical approach is to utilize a variational formulation of a differential or integral equation mainly as a starting point for the formulation of (Petrov-) Galerkin scheme which gives rise to a finite dimensional system of linear or nonlinear equations. The finite dimensional problem has then to be solved in an efficient way. As we have seen before, one then faces several obstructions such as ill–conditioning or the instability of the disretizations, for instance, due to the wrong choice of trial spaces. For instance, in the case of the double layer potential operator, stability of the Galerkin scheme is only guaranteed for sufficiently good spatial resolution, i.e. sufficient closeness to the infinite dimensional problem. As we shall see below, more severe stability problems are encountered when dealing with non–coercive problems such as saddle point problems. In this case, the trial spaces for the different solution components have to satisfy certain compatibility conditions known as Ladyšhenskaja-Babuška-Brezzi (LBB) condition. In brief, although the underlying infinite dimensional problem may be well–posed in the sense of (40), the corresponding finite dimensional problem may not always share this property.

In contrast we propose here a somewhat different paradigm that tries to exploit the well–posedness of the underlying continuous problem to the best possible extent along the following

Encyclopedia of Computational Mechanics. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

line:

(I) Establish well–posedness of the underlying variational problem;
(II) transform this problem into an *equivalent infinite dimensional* one which is now well–posed in ℓ_2;
(III) devise a *convergent iteration* for the infinite dimensional ℓ_2-problem;
(IV) only at that stage realize this iteration approximately with the aid of an *daptive application* of the involved (linear or nonlinear) operators.

5.2. The Scope of Problems – (I)

We describe first the scope of problems we have in mind. We begin with a general format which will then be exemplified by several examples.

For a given (possibly nonlinear) operator F the equation

$$\mathcal{F}(u) = f \tag{89}$$

is always understood in a weak sense, namely to find u in some normed space $\mathcal{H}$ such that for given data f

$$\langle v, \mathcal{F}(u)\rangle = \langle v, f\rangle, \quad \forall\ v \in \mathcal{H}. \tag{90}$$

This makes sense for any $f \in \mathcal{H}'$, the dual of $\mathcal{H}$ (recall (41)) and when $\mathcal{F}$ takes $\mathcal{H}$ onto its dual $\mathcal{H}'$. In principle, the conservation laws fit into this framework as well. They will be, however, excluded from major parts of the following discussion, since we will assume from now on that $\mathcal{H}$ is a Hilbert space.

The operator $\mathcal{F}$ is often given in a strong form so that the first task is to identify the right space $\mathcal{H}$ for which (90) is well–posed. We have already seen what this means when $\mathcal{F}$ is linear, see (40). The classical Dirichlet problem with $\mathcal{H} = H_0^1(\Omega)$ and the single layer potential equation (67) with $\mathcal{H} = H^{-1/2}(\Gamma)$ are examples. When dealing with nonlinear problems one may have to be content with locally unique solutions and it is natural to require that corresponding local linearizations are used to define well–posedness. Thus we assume that the *Frechét–derivative* $D\mathcal{F}(v)$ of $\mathcal{F}$ at v, defined by

$$\langle z, D\mathcal{F}(v)w\rangle = \lim_{t\to 0}\frac{1}{t}\langle z, \mathcal{F}(v+tw) - \mathcal{F}(v)\rangle, \quad \forall\ \ z \in \mathcal{H}, \tag{91}$$

exists for every v in a neighborhood $\mathcal{U}$ of a solution u of (90) as a mapping from $\mathcal{H}$ onto $\mathcal{H}'$. In analogy to (40), well–posedness now means that there exist for every $v \in \mathcal{U}$ positive finite constants $c_{\mathcal{F},v}, C_{\mathcal{F},v}$ such that

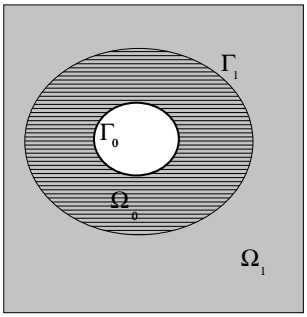
$$c_{\mathcal{F},v}\|w\|_{\mathcal{H}} \le \|D\mathcal{F}(v)w\|_{\mathcal{H}'} \le C_{\mathcal{F},v}\|w\|_{\mathcal{H}}, \quad \forall\ \ w \in \mathcal{H}. \tag{92}$$

We have already seen several examples in Sections 2.5 and 4.1 which, however, are all coercive. We shall therefore briefly review several further examples as models for different problem types. They also indicate that, as an additional practical obstruction, the topology for which the problem is well–posed involves norms that are usually difficult to deal with computationally. Most of the examples are actually linear but they may as well play the role of a (local) linearization.

Encyclopedia of Computational Mechanics. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

5.2.1. Transmission Problem The following example is interesting because it involves both local and global operators, see Costabel and Stephan, 1990

$$\begin{aligned} -\nabla\cdot(\mathbf{a}\nabla u) &= f \qquad \text{in } \Omega_0, \\ -\Delta u &= 0 \qquad \text{in } \Omega_1, \\ u|_{\Gamma_0} &= 0 \\ \mathcal{H} &:= H^1_{0,\Gamma_D}(\Omega_0)\times H^{-1/2}(\Gamma_1). \end{aligned}$$



Both boundary value problems are coupled by the *interface conditions:*

$$u^- = u^+, \quad (\partial_{\mathbf{n}})u^- = (\partial_{\mathbf{n}})u^+.$$

A well-posed weak formulation of this problem with respect to the above $\mathcal{H}$ is

$$\begin{aligned} \langle \mathbf{a}\nabla u, \nabla v\rangle_{\Omega_0} + \langle \mathcal{W}u - (\tfrac{1}{2}\mathcal{I} - \mathcal{K}')\sigma, v\rangle_{\Gamma_1} &= \langle f, v\rangle_{\Omega_0}, \qquad v \in H^1_{0,\Gamma_D}(\Omega_0), \\ \langle (\tfrac{1}{2}\mathcal{I} - \mathcal{K})u, \delta\rangle_{\Gamma_1} + \langle \mathcal{V}\sigma, \delta\rangle_{\Gamma_1} &= 0, \qquad \delta \in H^{-1/2}(\Gamma_1), \end{aligned}$$

where $\mathcal{K}, \mathcal{V}, \mathcal{W}$ are the double, single layer potential and hypersingular operator, see Dahmen, Kunoth and Schneider, 2002.

Note that, as an additional obstruction, the occurrence and evaluation of *difficult norms* like $\|\cdot\|_{H^{1/2}(\Gamma)}$, $\|\cdot\|_{H^{-1/2}(\Gamma)}$, $\|\cdot\|_{H^{-1}(\Omega)}$ arises.

5.2.2. Saddle Point Problems All the above examples involve coercive bilinear forms. An important class of problems which are no longer coercive are *saddle point problems.* A classical example is

The Stokes System The simplest model for viscous incompressible fluid flow is the Stokes system

$$\begin{aligned} -\nu\Delta u + \nabla p &= f \qquad \text{in } \Omega, \\ \text{div } u &= 0 \qquad \text{in } \Omega, \\ u|_\Gamma &= 0, \end{aligned} \tag{93}$$

where u and p are the velocity, respectively pressure, see Brezzi and Fortin, 1991; Girault and Raviart, 1986. The relevant function spaces are

$$X := (H^1_0(\Omega))^d, \quad M = L_{2,0}(\Omega) := \{q \in L_2(\Omega) : \int_\Omega q = 0\}. \tag{94}$$

In fact, one can show that the range of the divergence operator is $L_{2,0}(\Omega)$. The weak formulation of (93) is

$$\begin{aligned} \nu\langle \nabla v, \nabla u\rangle_{\mathbf{L}_2(\Omega)} + \langle \text{ div } v, p\rangle_{L_2(\Omega)} &= \langle f, v\rangle, \quad v \in (H^1_0(\Omega))^d \\ \langle \text{ div } u, q\rangle_{L_2(\Omega)} &= 0, \qquad q \in L_{2,0}(\Omega), \end{aligned} \tag{95}$$

Encyclopedia of Computational Mechanics. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

i.e., one seeks a solution (u,p) in the energy space $\mathcal{H} = X \times M = (H_0^1(\Omega))^d \times L_{2,0}(\Omega)$, for which the mapping property (92) can be shown to hold.

First Order Systems One is often more interested in derivatives of the solution of an elliptic boundary value problem which leads to *mixed formulations*. Introducing the *fluxes* $\theta := -a\nabla u$, (36) can be written as a system of first order equations whose weak formulation reads

$$\begin{array}{rcrcll} \langle \theta, \eta\rangle & + & \langle \eta, a\nabla u\rangle & = & 0, & \forall\ \eta \in (L_2(\Omega))^d, \\ -\langle \theta, \nabla v\rangle & & & = & \langle f, v\rangle, & \forall v \in H^1_{0,\Gamma_D}(\Omega). \end{array} \tag{96}$$

One now looks for a solutiom $(\theta, u) \in \mathcal{H} := (L_2(\Omega))^d \times H^1_{0,\Gamma_D}(\Omega)$. For a detailed discussion in the finite element context, see e.g. Bramble, Lazarov, and Pasciak, 1997. It turns out that in this case the Galerkin discretization inherits the stability from the original second order problem.

The General Format The above examples are special cases of the following general problem class. A detailed treatment can be found in Brezzi and Fortin, 1991; Girault and Raviart, 1986. Suppose X, M are Hilbert spaces and that $a(\cdot,\cdot)$, $b(\cdot,\cdot)$ are bilinear forms on $X\times X$, respectively $X \times M$ which are continuous

$$|a(v,w)| \lesssim \|v\|_X \|w\|_X, \quad |b(q,v)| \lesssim \|v\|_X \|q\|_M. \tag{97}$$

Given $f_1 \in X'$, $f_2 \in M'$, find $(u,p) \in X \times M =: \mathcal{H}$ such that one has for all $(v,q) \in \mathcal{H}$

$$\langle (v,q), \mathcal{F}(u,p)\rangle := \left\{ \begin{array}{rcrcl} a(u,v) & + & b(p,v) & = & \langle v, f_1\rangle, \\ b(q,u) & & & = & \langle q, f_2\rangle. \end{array} \right. \tag{98}$$

Note that when $a(\cdot,\cdot)$ is positive definite symmetric, the solution component u minimizes the quadratic functional $J(w) := \frac{1}{2}a(w,w) - \langle f_1, w\rangle$ subject to the constraint $b(u,q) = \langle q, f_2\rangle$, for all $q \in M$, which corresponds to

$$\inf_{v\in X} \sup_{q\in M} \left(\frac{1}{2} a(v,v) + b(v,q) - \langle f_1, v\rangle - \langle q, f_2\rangle \right).$$

This accounts for the term saddle point problem (even under more general assumptions on $a(\cdot,\cdot)$).

In order to write (98) as an operator equation, define the operators A, B by

$$a(v,w) =: \langle v, Aw\rangle, \quad v \in X, \quad b(v,p) =: \langle Bv, q\rangle, \quad q \in M,$$

so that (98) becomes

$$\mathcal{F}(u,p) := \begin{pmatrix} A & B' \\ B & 0 \end{pmatrix} \begin{pmatrix} u \\ p \end{pmatrix} = \begin{pmatrix} f_1 \\ f_2 \end{pmatrix} =: f. \tag{99}$$

As for the mapping property (92), a simple (sufficient) condition reads as follows, see Brezzi and Fortin, 1991; Girault and Raviart, 1986. If $a(\cdot,\cdot)$ is *elliptic* on

$$\ker B := \{v \in X : b(v,q) = 0,\ \forall\ q \in M\},$$

i.e.,

$$a(v,v) \sim \|v\|_X^2, \quad v \in \ker B, \tag{100}$$

and if $b(\cdot,\cdot)$ satisfies the *inf-sup condition*

$$\inf_{q\in M} \sup_{v\in X} \frac{b(v,q)}{\|v\|_X \|q\|_M} > \beta \tag{101}$$

for some positive β, then (97) is well-posed in the sense of (92). Condition (101) means that B is surjective (and thus has closed range). Condition (100) is actually too strong. It can be replaced by requiring bijectivity of A on $\ker B$, see Brezzi and Fortin, 1991.

Aside from leading to large ill conditioned systems the additional obstructions are the *indefiniteness* of this type of problem and that the well–posedness of the infinite dimensional problem is not automatically inherited by Galerkin discretizations, say. In fact, the trial spaces in X and M have to be compatible in the sense that they satisfy the inf-sup condition (101) *uniformly* with respect to the resolution of the chosen discrtetizations. This is called the *Ladyšhenskaya-Babuška-Brezzi-condition* (LBB) and may, depending on the problem, be a delicate task.

5.2.3. A Nonlinear Model Problem A wide range of phenomena involve the interaction of a (linear) diffusion with a nonlinear reaction or advection part. We therefore close the list of examples with the simple class of semilinear elliptic boundary value problems. On one hand, it permits a rather complete analysis. On the other hand, it still exhibits essential features that are relevant for a wider scope of nonlinear problems. In this section, we follow Cohen, Dahmen and DeVore, 2002b and suppose 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

$$c\|v\|_{\mathcal{H}}^2 \le a(v,v), \quad a(v,w) \le C\|v\|_{\mathcal{H}}\|w\|_{\mathcal{H}}, \quad \forall\ v,w \in \mathcal{H}. \tag{102}$$

The simplest example is

$$a(v,u) := \langle \nabla v, \nabla u\rangle + \kappa\langle v,u\rangle, \quad \kappa \ge 0, \quad \langle v,w\rangle = \int_\Omega vw, \tag{103}$$

and $\mathcal{H} = H_0^1(\Omega)$ endowed with the norm $\|v\|_{\mathcal{H}}^2 := \|\nabla v\|_{L_2(\Omega)}^2 + \kappa\|v\|_{L_2(\Omega)}^2$.

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

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

$$\|\mathcal{G}(u) - \mathcal{G}(v)\|_{\mathcal{H}'} \le C(\max\{\|u\|_{\mathcal{H}}, \|v\|_{\mathcal{H}}\})\|u-v\|_{\mathcal{H}}, \quad u,v \in \mathcal{H}, \tag{104}$$

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

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

$$\langle v, \mathcal{F}(u)\rangle := a(v,u) + \langle v, \mathcal{G}(u)\rangle = \langle v, f\rangle, \quad \forall\ \ v \in \mathcal{H}, \tag{105}$$

is of the form (90) with $\mathcal{F}(u) = \mathcal{A}u + \mathcal{G}(u)$. Note that with the bilinear form from (103), (105) may also arise through an implicit time discretization of a nonlinear parabolic equation.

Encyclopedia of Computational Mechanics. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

The unique solvability of (105) is easily ensured when $\mathcal{G}$ is in addition assumed to be *monotone*, i.e., $(x-y)(\mathcal{G}(x)-\mathcal{G}(y)) \geq 0$ for $x, y \in \mathbb{R}$. In this case $\mathcal{F}(u) = f$ is the Euler equation of a convex minimization problem, Cohen, Dahmen and DeVore, 2002b.

A simple example is

$$\langle v, \mathcal{F}(u)\rangle = \int_\Omega \nabla v^T \nabla u + v u^3 dx, \quad \mathcal{H} = H_0^1(\Omega),\ \mathcal{H}' = H^{-1}(\Omega). \tag{106}$$

That (at least for $d \leq 3$) $\mathcal{H} = H_0^1(\Omega)$ is indeed the right choice can be seen as follows. The fact that $H^1(\Omega)$ is continuously embedded in $L_4(\Omega)$ for $d = 1, 2, 3$, readily implies that $\mathcal{G}(v) \in H^{-1}(\Omega)$ for $v \in H_0^1(\Omega)$. Moreover, $\langle z, \mathcal{F}(v+tw) - \mathcal{F}(v)\rangle = t\langle \nabla z, \nabla w\rangle + \langle z, t3v^2 w + t^2 3vw^2 + t^3 w^3\rangle$ so that $\langle z, D\mathcal{F}(v)w\rangle = \langle \nabla z, \nabla w\rangle + 3\langle z, v^2 w\rangle$ and hence

$$D\mathcal{F}(v)w = -\Delta w + 3v^2 w. \tag{107}$$

Therefore, again by the embedding $H^s \hookrightarrow L_p$ if $\frac{1}{2} < \frac{s}{d} + \frac{1}{p}$, i.e. $\|v\|_{L_4} \lesssim \|v\|_{\mathcal{H}}$ for $d < 4$ with $\mathcal{H} = H_0^1(\Omega)$, we see that

$$\|D\mathcal{F}(v)w\|_{\mathcal{H}'} = \sup_{z\in\mathcal{H}} \frac{\langle \nabla z, \nabla w\rangle + 3\langle z, v^2 w\rangle}{\|z\|_{\mathcal{H}}} \lesssim \|w\|_{\mathcal{H}} + \|v\|_{\mathcal{H}}^2 \|w\|_{\mathcal{H}}.$$

On the other hand,

$$\|D\mathcal{F}(v)w\|_{\mathcal{H}'} \geq \frac{\langle \nabla w, \nabla w\rangle + 3\langle w, v^2 w\rangle}{\|w\|_{\mathcal{H}}} \geq \frac{\|\nabla w\|_{L_2}^2}{\|w\|_{\mathcal{H}}} \geq (1 + c(\Omega)^2)^{-1} \|w\|_{\mathcal{H}},$$

where we have used Poincaré's inequality in the last step. Hence we obtain

$$(1 + c(\Omega)^2)^{-1} \|w\|_{\mathcal{H}} \leq \|D\mathcal{F}(v)w\|_{\mathcal{H}'} \lesssim (1 + \|v\|_{\mathcal{H}}^2)\|w\|_{\mathcal{H}}, \quad w \in \mathcal{H}, \tag{108}$$

which is (92).

The scope of of problems is actually not limited at all to the variational formulation of integral or partial differential equations but covers, for instance, also optimal control problems with PFEs as constraints, see Dahmen and Kunoth, 2002; Kunoth, 2001.

5.3. Transformation into a Well–Posed ℓ_2–Problem – (II)

Suppose that (90) is well–posed with respect to the energy space $\mathcal{H}$. In all previous examples $\mathcal{H}$ was a (closed subspace of a) Sobolev or a product of such spaces. As indicated in Section 2 it is known how to construct wavelet bases for such spaces. In the following, we will therefore assume that Ψ is a Riesz basis for $\mathcal{H}$.

The transformation of (90) into wavelet coordinates is analogous to the linear case (see Theorem 2.2 in Section 2.5). In fact, testing in (90) with $v = \psi_\lambda$ for all $\lambda \in \mathcal{J}$, defines through $\mathbf{F}(\mathbf{v}) := (\langle \psi_\lambda, \mathcal{F}(v)\rangle)_{\lambda\in\mathcal{J}}$ a sequence valued nonlinear mapping $\mathbf{F}$ which depends on the array $\mathbf{v} \in \ell_2$ via the wavelet expansion $v = \sum_{\lambda\in\mathcal{J}} v_\lambda \psi_\lambda$. Similarly, the Jacobian of $\mathbf{F}$ at $\mathbf{v}$ acting on $\mathbf{w} \in \ell_2$ is defined by $D\mathbf{F}(\mathbf{v})\mathbf{w} = (\langle \psi_\lambda, D\mathcal{F}(v)w\rangle)_{\lambda\in\mathcal{J}}$. Finally, setting as before $\mathbf{f} := (\langle \psi_\lambda, f\rangle)_{\lambda\in\mathcal{J}}$ the following fact can be derived by the same arguments as in Theorem 2.2.

Theorem 5.1. *Assume that (92) and (30) hold. Then the variational problem (90) is equivalent to* $\mathbf{F}(\mathbf{u}) = \mathbf{f}$ *where* $u = \sum_{\lambda\in\mathcal{J}} u_\lambda \psi_\lambda$. *Moreover, when the latter problem is well-posed in* ℓ_2, *i.e. for* $v = \sum_{\lambda\in\mathcal{J}} v_\lambda \psi_\lambda$ *in some neighborhood* $\mathcal{U}$ *of the locally unique solution* u *of (90)*

$$c_\Psi^{-2} c_{\mathcal{F},v}^{-1} \|\mathbf{w}\|_{\ell_2} \leq \|D\mathbf{F}(\mathbf{v})\mathbf{w}\|_{\ell_2} \leq C_\Psi^2 C_{\mathcal{F},v} \|\mathbf{w}\|_{\ell_2}, \quad \mathbf{w} \in \ell_2. \tag{109}$$

As for the special case (105), note that the monotonicity of $\mathcal{G}$ implies the monotonicity of $\mathbf{G}(\cdot)$, defined by $\mathbf{G}(\mathbf{v}) := (\langle \psi_\lambda, \mathcal{G}(v)\rangle)_{\lambda\in\mathcal{J}}$ and hence the positive semidefiniteness of $D\mathbf{G}(\mathbf{v})$, see Cohen, Dahmen and DeVore, 2002b for details.

5.4. An Iteration for the Infinite dimensional Problem – (III)

Once the problem attains a well–conditioned form in ℓ_2, it makes sense to devise an iterative scheme for the *full* infinite dimensional problem $\mathbf{F}(\mathbf{u}) = \mathbf{f}$ that converges with a guaranteed error reduction rate. These iterations will take the form

$$\mathbf{u}^{n+1} = \mathbf{u}^n - \mathbf{C}_n(\mathbf{F}(\mathbf{u}^n) - \mathbf{f}), \quad n = 0, 1, 2, \ldots, \tag{110}$$

where the (infinite) matrix $\mathbf{C}_n$ is possibly stage dependent. It can be viewed a a fixed point iteration based on the trivial identity $\mathbf{u} = \mathbf{u} - \mathbf{C}(\mathbf{F}(\mathbf{u}) - \mathbf{f})$. We shall indicate several ways of choosing $\mathbf{C}_n$, depending on the nature of the underlying variational problem (90).

Gradient Iterations: In the above case of elliptic semilinear problems, the transformed problem still identifies the unique minimum of a convex functional. Thus, it makes sense to consider *gradient iterations*, i.e. $\mathbf{C}_n = \alpha\mathbf{I}$

$$\mathbf{u}^{n+1} = \mathbf{u}^n - \alpha(\mathbf{F}(\mathbf{u}^n) - \mathbf{f}), \quad n = 0, 1, 2, \ldots, \tag{111}$$

In fact, one can estimate a suitable positive damping parameter $\alpha > 0$ from the constants in (30), (92) and (104), see Cohen, Dahmen and DeVore, 2002b for details, so that (111) converges for $\mathbf{u}^0$ say with a fixed reduction rate $\rho < 1$,

$$\|\mathbf{u}^{n+1} - \mathbf{u}\|_{\ell_2} \le \rho\|\mathbf{u}^n - \mathbf{u}\|_{\ell_2}, \quad n \in \mathbb{N}_0. \tag{112}$$

For instance, in the linear case $\mathcal{G} \equiv 0$, one can take any $\alpha < 2/(C_\Psi^2 C_{\mathcal{A}})$ (see (43)) and verify that $\rho = \max\{1 - \alpha c_\Psi^2 c_{\mathcal{A}}, |1 - C_\Psi^2 C_{\mathcal{A}}|\}$ works.

Least Squares Iteration: Of course, when dealing with indefinite Jacobians the above scheme will in general no longer work. However, the well–posedness in ℓ_2 offers, in principle, always a remedy which we explain first in the linear case $\mathcal{A}u = f$, where this may stand for any of the well-posed problems discussed in Section 5.2. Since then (109) reduces to (43) one can again find a positive α so that $\mathbf{C}_n = \alpha\boldsymbol{A}^T$ leads to an iteration

$$\mathbf{u}^{n+1} = \mathbf{u}^n - \alpha\boldsymbol{A}^T(\boldsymbol{A}\mathbf{u}^n - \mathbf{f}), \quad n = 0, 1, 2, \ldots, \tag{113}$$

that satisfies (112) with some fixed $\rho < 1$. Clearly this is simply a gradient iteration for the *least squares formulation* $\boldsymbol{A}^T\boldsymbol{A} = \boldsymbol{A}^T\mathbf{f}$ in the wavelet coordinate domain, see also Dahmen, Kunoth and Schneider, 2002 for connections with least squares finite element methods.

This is interesting because it suggests an analog also in the *general nonlinear case* even when $D\mathbf{F}(\mathbf{v})$ is indefinite but (92) (or equivalently (109)) holds. In fact, the role of $\boldsymbol{A}^T$ is played by $D\mathbf{F}(\mathbf{u}^n)^T$. Setting

$$\boldsymbol{R}(\mathbf{v}) := \mathbf{F}(\mathbf{v}) - \mathbf{f}, \tag{114}$$

and noting that

$$\boldsymbol{R}(\mathbf{v}) = \mathbf{F}(\mathbf{v}) - \mathbf{F}(\mathbf{u}) = \Big(\int_0^1 D\mathbf{F}(\mathbf{u} + s(\mathbf{v} - \mathbf{u}))ds\Big)(\mathbf{v} - \mathbf{u}), \tag{115}$$

one can derive from (109) that the iteration

$$\mathbf{u}^{n+1} = \mathbf{u}^n - \alpha D\mathbf{F}(\mathbf{u}^n)^T \boldsymbol{R}(\mathbf{u}^n), \tag{116}$$

can be made to satisfy (112) for a suitable positive damping factor α, depending on the constants in (109) and a sufficiently good initial guess $\mathbf{u}^0$ (which is always needed in the case of only locally unique solutions). Moreover, when being sufficiently close to the solution $\mathbf{C}_n = \alpha D\mathbf{F}(\mathbf{u}^n)^T$ can be frozen to $\mathbf{C} = D\mathbf{F}(\mathbf{u}^0)$ so as to still realize a strict error reduction (112), see Cohen, Dahmen and DeVore, 2002b for details.

Uzawa Iteration: Of course, in the above least squares iterations (113) and (116) the damping factor α may have to be chosen rather small which entails a poor error reduction rate ρ. Whenever $\mathcal{A}$ or the linearization $D\mathbf{F}(\mathbf{v})$ corresponds to a saddle point operator (98) or (99), the squaring of the condition number caused by the least squares formulation can be avoided with the aid of an *Uzawa iteration*, see Dahlke, Dahmen and Urban, 2002; Dahmen, Urban and Vorloeper, 2002. We indicate this only for the linear case. Instead of working directly with the wavelet representation $\mathbf{F}(\mathbf{u}, \mathbf{p}) = \mathbf{f}$, one can first eliminate the solution component $\mathbf{u}$, the velocity in the case of the Stokes problem, on the infinite dimensional operator level. Recall from (99) that $\mathbf{F}$ takes the form

$$\mathbf{F}(\mathbf{u}, \mathbf{p}) = \begin{pmatrix} \boldsymbol{A} & \mathbf{B}^T \\ \mathbf{B} & \mathbf{0} \end{pmatrix} = \mathbf{f} := \begin{pmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \end{pmatrix}, \quad \mathbf{f}_1 := (\langle \psi_{X,\lambda}, f_1 \rangle)_{\lambda \in \mathcal{J}_X}, \ \mathbf{f}_2 := (\langle \psi_{M,\lambda}, f_2 \rangle)_{\lambda \in \mathcal{J}_M}, \tag{117}$$

where for given Riesz bases Ψ_X, Ψ_M of the component spaces of $\mathcal{H} = X \times M$, $\boldsymbol{A} := a(\Psi_X, \Psi_X)$, $\mathbf{B} := b(\Psi_M, \Psi_X)$ are the wavelet representations of the operators A, B in (99), respectively. Recall also that $\boldsymbol{A}$ need only be invertible on the kernel of $\mathbf{B}$. To eliminate $\mathbf{u}$ one may first have to modify the system as explained in Dahlke, Dahmen and Urban, 2002 so as to make the modification of $\boldsymbol{A}$ an automorphism on all of $\ell_2(\mathcal{J}_X)$. Without loss of generality we may therefore assume that this is already the case and that $\boldsymbol{A}$ satisfies (43). Then from the first system $\mathbf{A}\mathbf{u} + \mathbf{B}^T\mathbf{p} = \mathbf{f}_1$ we conclude that $\mathbf{u} = \mathbf{A}^{-1}\mathbf{f}_1 - \mathbf{A}^{-1}\mathbf{B}^T\mathbf{p}$ which, by the second equation in (117), gives $\mathbf{B}\mathbf{u} = \mathbf{B}\mathbf{A}^{-1}\mathbf{f}_1 - \mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T\mathbf{p} = \mathbf{f}_2$. Hence (117) is equivalent to the *Schur complement* problem

$$\mathbf{S}\mathbf{p} = \mathbf{B}\boldsymbol{A}^{-1}\mathbf{f}_1 - \mathbf{f}_2, \quad \mathbf{S} := \mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T. \tag{118}$$

Once (118) has been solved the eliminated component can be recovered from (the elliptic problem)

$$\boldsymbol{A}\mathbf{u} = \mathbf{f}_1 - \mathbf{B}^T\mathbf{p}. \tag{119}$$

From the well–posedness of (117) on ℓ_2 one easily derives that $\mathbf{S}$ is also boundedly invertible on $\ell_2(\mathcal{J}_M)$. So one can find a positive α such that the gradient iteration $\mathbf{p}^{n+1} = \mathbf{p}^n - \alpha(\mathbf{S}\mathbf{p}^n - (\mathbf{B}\boldsymbol{A}^{-1}\mathbf{f}_1 - \mathbf{f}_2))$ satisfies (112) for some $\rho < 1$. (Actually, the residual may have to be modified again when, as in the case of the Stokes problem, the Lagrange multiplier space M is a subspace of finite codimension in a larger space for which a wavelet Riesz basis is given. For simplicity, we suppress this issue here and refer to Dahlke, Dahmen and Urban, 2002 for details). The problem is that the Schur complement is generally not easily accessible, due to the presence of the factor $\boldsymbol{A}^{-1}$. However, note that $\mathbf{S}\mathbf{p}^n - (\mathbf{B}\boldsymbol{A}^{-1}\mathbf{f}_1 - \mathbf{f}_2) = \mathbf{f}_2 - \mathbf{B}\mathbf{u}^n$ whenever $\mathbf{u}^n$ is

the solution of the elliptic problem $\boldsymbol{A}\mathbf{u}^n = \mathbf{f}_1 - \mathbf{B}^T\mathbf{p}^n$. The gradient iteration for the Schur complement problem then takes the form

$$\mathbf{p}^{n+1} = \mathbf{p}^n - \alpha(\mathbf{f}_2 - \mathbf{B}\mathbf{u}^n), \quad \text{where } \boldsymbol{A}\mathbf{u}^n = \mathbf{f}_1 - \mathbf{B}^T\mathbf{p}^n, \quad n = 0, 1, 2, \ldots. \tag{120}$$

Thus, the iteration is again of the form we want but each step requires as an input the solution of a subproblem.

Newton Iteration: Finally, the choice $\mathbf{C}_n := (D\mathbf{F}(\mathbf{u}^n))^{-1}$ in (110) gives rise to the *Newton scheme*

$$\mathbf{u}^{n+1} = \mathbf{u}^n + \mathbf{w}^n, \quad D\mathbf{F}(\mathbf{u}^n)\mathbf{w}^n = \mathbf{f} - \mathbf{F}(\mathbf{u}^n) = -\boldsymbol{R}(\mathbf{u}^n), \tag{121}$$

where each step requires the solution of a linear subproblem. While all previous examples have convergence order one, the Newton scheme, in principle, offers even better convergence behavior. We shall address this issue later in more detail.

5.5. Perturbed Iteration Schemes – (IV)

We have indicated several ways of forming an (idealized) iteration scheme on the full infinite dimensional transformed system $\mathbf{F}(\mathbf{u}) = \mathbf{f}$. We have made essential use of the fact that the transformed system is well posed in ℓ_2 in the sense of (109). Recall that this hinges on the mapping property (92) induced by the original continuous problem and the availability of a Riesz basis Ψ for the corresponding energy space $\mathcal{H}$, (30). The final step is to realize the idealized iteration numerically. We shall do so *not* by choosing some fixed finite dimensional space on which the problem is projected, as was done in previous sections, but rather by approximating at each step the true residual $\mathbf{r}^n := \mathbf{C}_n(\mathbf{F}(\mathbf{u}^n) - \mathbf{f}) = \mathbf{C}_n\boldsymbol{R}(\mathbf{u}^n)$ within a suitable dynamically updated accuracy tolerance. Again we wish to avoid choosing for this approximation any a-priori fixed finite dimensional space but try to realize the required accuracy at the expense of possibly few degrees of freedom. The whole task can then be split into two steps:

(a) Assuming that, in each case at hand, a computational scheme is available that allows us to approximate the residuals within each desired target accuracy, determine first for which dynamic tolerances the iteration will converge in the sense that for any given target accuracy ϵ, the perturbed iteration outputs a finitely supported vector $\mathbf{u}(\epsilon)$ such that $\|\mathbf{u} - \mathbf{u}(\epsilon)\|_{\ell_2} \leq \epsilon$. Thus on the numerical side all approximations will take place in the Euclidean metric. Note however,that due to (30) this implies that

$$\|u - \sum_{\operatorname{supp}\mathbf{u}(\epsilon)} (\mathbf{u}(\epsilon))_\lambda \psi_\lambda\|_{\mathcal{H}} \leq C_\Psi \epsilon. \tag{122}$$

(b) Once this has been clarified, one has to come up with concrete realizations of the residual approximations appearing in the idealized iteration. It is clear from the above examples that the crucial task is to approximate with possibly few terms the sequence $\mathbf{F}(\mathbf{u}^n)$ in ℓ_2. Moreover, we will have to make precise what we mean by "possibly few terms", i.e. we have to analyze the computational complexity of the numerical schemes.

We shall address frst (a) under the assumption that we are given a numerical scheme

Encyclopedia of Computational Mechanics. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

RES$[\eta, \mathbf{C}, \mathbf{F}, \mathbf{f}, \mathbf{v}] \to \mathbf{r}_\eta$: WHICH FOR ANY POSITIVE TOLERANCE η AND ANY FINITELY SUPPORTED INPUT $\mathbf{v}$ OUTPUTS A FINITELY SUPPORTED VECTOR $\mathbf{r}_\eta$ SATISFYING

$$\|\mathbf{C}\boldsymbol{R}(\mathbf{v}) - \mathbf{r}_\eta\|_{\ell_2} \le \eta. \tag{123}$$

The second ingredient is a routine

COARSE$[\eta, \mathbf{w}] \to \mathbf{w}_\eta$: WHICH FOR ANY POSITIVE TOLERANCE η AND ANY FINITELY SUPPORTED INPUT VECTOR $\mathbf{w}$ PRODUCES A FINITELY SUPPORTED OUTPUT $\mathbf{w}_\eta$ WITH POSSIBLY FEW ENTRIES (SUBJECT TO CONSTRAINTS THAT WILL BE SPECIFIED LATER) SUCH THAT

$$\|\mathbf{w} - \mathbf{w}_\eta\|_{\ell_2} \le \eta. \tag{124}$$

This routine will be essential later to control the complexity of the overall perturbed iteration scheme.

Next we need some initialization, i.e., an initial guess $\mathbf{u}^0$ and an error bound

$$\|\mathbf{u} - \mathbf{u}^0\|_{\ell_2} \le \epsilon_0. \tag{125}$$

In general, such an estimate depends on the problem at hand. In the case of semi–linear elliptic problems, such a bound is, for instance, $\epsilon_0 = c_\Psi^{-2} c_{\mathcal{A}}^{-1}(\|\mathbf{G}(\mathbf{0})\|_{\ell_2} + \|\mathbf{f}\|_{\ell_2})$ for $\mathbf{u}^0 = \mathbf{0}$, see Cohen, Dahmen and DeVore, 2002b for details.

As a final prerequisite, one can find, again as a consequence of (109), for each of the above iteration schemes (110) a constant β such that

$$\|\mathbf{u} - \mathbf{v}\|_{\ell_2} \le \beta \|\mathbf{C}\boldsymbol{R}(\mathbf{v})\|_{\ell_2} \tag{126}$$

holds in a neigborhood of the solution $\mathbf{u}$. The perturbed iteration scheme may now be formulated as follows.

SOLVE$[\epsilon, \mathbf{C}, \boldsymbol{R}, \mathbf{u}^0] \to \bar{\mathbf{u}}(\epsilon)$

(I) CHOOSE SOME $\bar\rho \in (0,1)$. SET $\bar{\mathbf{u}}^0 = \mathbf{u}^0$, THE CORRESPONDING INITIAL BOUND ϵ_0 ACCORDING TO THE ABOVE INITIALIZATION, AND $j = 0$;

(II) IF $\epsilon_j \le \epsilon$ STOP AND OUTPUT $\bar{\mathbf{u}}(\epsilon) := \bar{\mathbf{u}}^j$; ELSE SET $\mathbf{v}^0 := \bar{\mathbf{u}}^j$ AND $k = 0$

(II.1) SET $\eta_k := \bar\rho^k \epsilon_j$ AND COMPUTE

$$\mathbf{r}^k = \text{RES}[\eta_k, \mathbf{C}, \mathbf{F}, \mathbf{f}, \mathbf{v}^k], \qquad \mathbf{v}^{k+1} = \mathbf{v}^k - \mathbf{r}^k.$$

(II.2) IF

$$\beta\big(\eta_k + \|\mathbf{r}^k\|_{\ell_2}\big) \le \epsilon_j/(2(1+2C^*)), \tag{127}$$

SET $\tilde{\mathbf{v}} := \mathbf{v}^k$ AND GO TO (III). ELSE SET $k+1 \to k$ AND GO TO (II.1).

(III) COARSE$[\frac{2C^*\epsilon_j}{2(1+2C^*)}, \tilde{\mathbf{v}}] \to \bar{\mathbf{u}}^{j+1}$, $\epsilon_{j+1} = \epsilon_j/2$, $j+1 \to j$, GO TO (II).

The constant C^* depends on the particular realization of the routine COARSE and will be specified later.

Note that step (II) is just an approximation of the updates in (110). This applies until the stopping criterion (127) is met. This is a–posteriori information based on the numerical

residual. The fact that there is actually a uniform bound K for the number of updates in step (II), independent of the data and the target accuracy, until (127) is satisfied and a coarsening step (III) is carried out, relies on (126) and the underlying well–posedness (109). The parameter $\bar{\rho}$ is actually allowed here to be smaller than the true error reduction rate ρ in (112) for which only a poor or a possibly too pessimistic estimate may be available.

One can show by fairly straightforward perturbation arguments that the choice of accuracy tolerances in SOLVE implies convergence, Cohen, Dahmen and DeVore, 2002b.

Proposition 5.2. *The iterates* $\bar{\mathbf{u}}^j$ *produced by the scheme* SOLVE *satisfy*

$$\|\mathbf{u} - \bar{\mathbf{u}}^j\|_{\ell_2} \leq \epsilon_j, \tag{128}$$

so that in particular $\|\mathbf{u} - \bar{\mathbf{u}}(\epsilon)\|_{\ell_2} \leq \epsilon$. *By (30), this means*

$$\|u - \sum_{\lambda \in \Lambda(\epsilon)} \bar{u}(\epsilon)_\lambda \psi_\lambda\|_{\mathcal{H}} \leq C_\Psi \epsilon, \tag{129}$$

where C_Ψ *is the constant from (30) and* $\Lambda(\epsilon) := \operatorname{supp} \mathbf{u}(\epsilon)$.

6. Construction of Residual Approximations and Complexity Analysis

It remains to construct concrete realizations of the routines RES and COARSE. It turns out that the development of such routines is closely intertwined with their complexity analysis. Since the conceptual tools are probably unfamiliar in the context of numerical simulation we highlight some of them in the next section.

6.1. Best N–Term Approximation

A *lower bound* for the computational complexity of SOLVE is, of course, the growth of the supports of outputs in step (II.1), which determines how the overall number of degrees of freedom grows until the target accuracy is reached. Therefore a lower bound for the computational complexity of SOLVE is given by the number of terms needed to recover the true solution $\mathbf{u}$ in ℓ_2 within accuracy ϵ. This is the issue of *best N–term approximation* in ℓ_2. Thus the question arises whether, or under which circumstances, SOLVE can actually attain this lower bound, at least asymptotically. Since best N–term approximation limits what can be achieved at best we briefly review some relevant features concerning best N–term approximation.

There are two intimately connected ways of looking at an error analysis for N-term approximation. In the first, we can specify the target accuracy ϵ and ask what is the smallest number $N(\epsilon)$ of terms needed to recover a given object? The second view is to assume we are given a budget N of terms and ask what accuracy $\epsilon(N)$ can be achieved with the best selection of N terms? The process of selecting such N terms is obviously nonlinear. This can be formalized by defining the following *error of approximation*:

$$\sigma_{N,\ell_2}(\mathbf{v}) := \inf_{\#\operatorname{supp} \mathbf{w} \leq N} \|\mathbf{v} - \mathbf{w}\|_{\ell_2}. \tag{130}$$

Obviously, $\sigma_{N,\ell_2}(\mathbf{v})$ is attained by $\mathbf{w} = \mathbf{v}_N$ comprised of the N largest terms of $\mathbf{v}$ in modulus. Note that this is not necessarily unique since several terms may have equal modulus.

Encyclopedia of Computational Mechanics. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

Analogously one can define $\sigma_{N,\mathcal{H}}(v)$ by looking for the best approximation of v in $\mathcal{H}$ by a linear combination of at most N wavelets. One easily infers from (30) that

$$c_\Psi \sigma_{N,\ell_2}(\mathbf{v}) \le \sigma_{N,\mathcal{H}}(v) \le C_\Psi \sigma_{N,\ell_2}(\mathbf{v}). \tag{131}$$

Best N-term approximations in the Euclidean metric yield therefore *near best* N-term approximations in $\mathcal{H}$. Hence an element in $\mathcal{H}$ can be approximated well with relatively few terms if and only if this is true for its coefficient array in ℓ_2.

We shall proceed with identifying classes of sequences in ℓ_2 for which $\epsilon(N)$ decays like N^{-s} since these are the rates that can be expected from approximations based on spatial refinements (h-methods). To this end, consider the classes

$$\mathcal{A}^s(\mathcal{H}) := \{v \in \mathcal{H} : \sigma_{N,\mathcal{H}}(v) \lesssim N^{-s}\}, \quad \mathcal{A}^s(\ell_2) := \{\mathbf{v} \in \ell_2 : \sigma_{N,\ell_2}(\mathbf{v}) \lesssim N^{-s}\}. \tag{132}$$

These are normed linear spaces endowed with the norms

$$\|v\|_{\mathcal{A}^s(\mathcal{H})} := \sup_{N\in\mathbb{N}} N^s \sigma_{N,\mathcal{H}}(v), \quad \|\mathbf{v}\|_{\mathcal{A}^s(\ell_2)} := \sup_{N\in\mathbb{N}} N^s \sigma_{N,\ell_2}(\mathbf{v}).$$

Thus to achieve a target accuracy ϵ, the order of $N(\epsilon) \sim \epsilon^{-1/s}$ terms are needed for $v \in \mathcal{A}^s(\mathcal{H})$ or $\mathbf{v} \in \mathcal{A}^s(\ell_2)$. Hence the larger $s > 0$ the less terms suffice.

Which property makes a function v or its coefficient sequence $\mathbf{v}$ *sparse* in the above sense is best explained when $\mathcal{H}$ is a Sobolev space $\mathcal{H} = H^s$ over some domain in $\mathbb{R}^d$. One can then show that for any positive δ (cf. DeVore, 1998; Bergh and Löfström, 1976; Cohen, 2000)

$$\ell_q \subset \mathcal{A}^s(\ell_2) \subset \ell_{q+\delta}, \quad B_q^{\alpha+sd}(L_q) \subset \mathcal{A}^s(H^\alpha) \subset B_q^{\alpha+sd-\delta}(L_q), \quad \frac{1}{q} = s + \frac{1}{2}. \tag{133}$$

Here ℓ_q consists of the q-summable sequences while $B_q^\alpha(L_q)$ denotes a *Besov space* consisting roughly of functions with smoothness α measured in L_q, see DeVore, 1998; Bergh and Löfström, 1976; Cohen, 2003 for precise definitions. For a certain smoothness range, depending on the regularity of the wavelet basis these spaces can also be characterized through wavelet bases. In fact for $0 < q \le \infty$ one has

$$\|v\|^q_{B_q^\alpha(L_q)} \sim \|v\|^q_{L_q} + \sum_{\lambda\in\mathcal{J}} 2^{\alpha q|\lambda|} \|\langle v, \tilde{\psi}_\lambda\rangle \psi_\lambda\|^q_{L_q}, \tag{134}$$

which, due to the equivalence $H^s = B_2^s(L_2)$, covers (32) as a special case, see e.g. Cohen, 2000; Cohen, 2003; DeVore, 1998. It is important to note here that the smaller q the weaker is the smoothness measure. By the Sobolev embedding theorem, the value of q given by (133) gives the weakest possible measure so that smoothness of order $sd + \alpha$ in L_q guarantees Sobolev regularity of order α corresponding to the anchor space $\mathcal{H} = H^\alpha$ (a Sobolev space of order α or a closed subspace defined e.g. by homogeneous boundary conditions). This is illustrated in Figure 9 below. Each point in the $(1/q, s)$–plane corresponds to a smoothness space (actually to a class of smoothness spaces) describing smoothness s measured in L_q. In our case we have $X = H^\alpha$ and $p = 2$. The spaces located left of the line with slope d emanating from X are embedded in X. The spaces of smoothness $\alpha + sd$ on the vertical line above X are essentially those whose elements can be approximated with accuracy $O(N^{-s})$ by approximants from *quasi–uniform meshes*, i.e. with equidistributed degrees of freedom. In the present terms this means just keeping all wavelets up to some scale J say (or equivalently working with uniform meshes), so

Encyclopedia of Computational Mechanics. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

that $N \sim 2^{Jd}$ would require the function v to belong to $B_\infty^{t+sd}(L_2)$ which is very close to the Sobolev space H^{t+sd}. The spaces on the critical embedding line, however, are characterized by *nonlinear approximation* like best N–term approximation. Thus, while the spaces obtained when moving to the right away from the vertical line on the same smoothness level grow and admit increasingly stronger singularities, this loss of regularity can be compensated by judiciously placing the degrees of freedom so as to retain the same convergence rates in terms of degrees of freedom N. Since H^{t+sd} is a much smaller space than $B_q^{t+sd}(L_q)$ this indicates the possible gain offered by nonlinear approximation schemes like best N-term approximation over simpler schemes based on a-priori fixed discretizations.

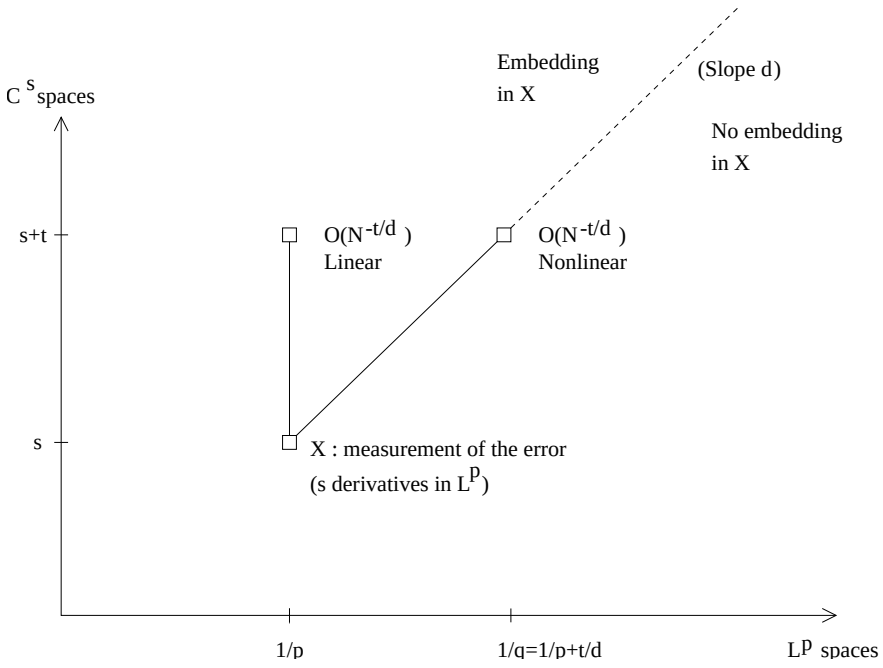


Figure 9. Topography of smoothness spaces

Of course, it remains to see whether this potential can be exploited by adaptive schemes.

Tree Structures: The above notion of best N–term approximation puts no constraints on the distribution of significant coefficients. In the context of conservation laws it was important that the significant coefficients were arranged in tree like structures which corresponds to local mesh refinements. Thus interrelating the selection of wavelets with locally refined meshes is one reason for imposing some sort of constraint on the distribution of wavelet coefficients. Another reason arises when approximating the quatities $\mathbf{F}(\mathbf{v})$ when $\mathbf{v}$ is some finitely supported vector. Intuitively, one might expect that the nonlinearity in $\mathbf{F}$ makes the effect of a term v_λ with large $|\lambda|$ cascade down to lower levels in a neighborhood of the support S_λ, which also gives rise to tree like structures.

Encyclopedia of Computational Mechanics. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

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

$$\mu \prec \lambda \Rightarrow S_\mu \subset S_\lambda, \tag{135}$$

where we 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 $\lambda \prec \mu$.

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

$$\mathcal{L} := \{\lambda \in \mathcal{J} : \lambda \notin \mathcal{T},\ \lambda \prec \mu \implies \mu \in \mathcal{T}\}. \tag{136}$$

The set $\mathcal{L}(\mathcal{T})$ plays the role of a (locally refined) mesh. In fact, one readily confirms that

$$\|\mathbf{v} - \mathbf{v}|_{\mathcal{T}}\|_{\ell_2}^2 = \sum_{\lambda \notin \mathcal{T}} |v_\lambda|^2 = \sum_{\lambda \in \mathcal{L}(\mathcal{T})} \sum_{\nu : \nu \preceq \lambda} |v_\lambda|^2, \tag{137}$$

which suggests considering the quantities

$$\tilde{v}_\lambda := \sum_{\nu : S_\nu \subset S_\lambda} |v_\nu|^2. \tag{138}$$

These quantities measure in some sense a local error associated with the spatial location of ψ_λ. To see this, suppose that the wavelets have the form $\psi_\lambda = \omega_\lambda \theta_\lambda$ where ω_λ are some positive weights (see (33)) and Θ is a Riesz basis for L_2 (which is the case for all constructions considered in Section 2). Then, by (30)

$$\|\sum_{\lambda \in \mathcal{J}} v_\lambda \psi_\lambda\|_{\mathcal{H}} \sim \|\mathbf{v}\|_{\ell_2} \sim \|\sum_{\lambda \in \mathcal{J}} v_\lambda \theta_\lambda\|_{L_2},$$

so that

$$\|v - \sum_{\lambda \in \mathcal{T}} v_\lambda \psi_\lambda\|_{\mathcal{H}} \sim \|\sum_{\lambda \notin \mathcal{T}} v_\lambda \theta_\lambda\|_{L_2}. \tag{139}$$

Note that the right hand side can be localized. In fact, for $\mu \in \mathcal{L}(\mathcal{T})$

$$\begin{aligned} \|\sum_{\lambda \notin \mathcal{T}} v_\lambda \theta_\lambda\|_{L_2(S_\mu)}^2 &= \|\sum_{|\nu| \geq |\mu| : S_\nu \cap S_\mu \neq \emptyset} v_\nu \theta_\nu\|_{L_2(S_\mu)}^2 \lesssim \sum_{|\nu| \geq |\mu| : S_\nu \cap S_\mu \neq \emptyset} v_\nu^2 \\ &\lesssim \sum_{\lambda \in \mathcal{L}(\mathcal{T}), S_\lambda \cap S_\mu \neq \emptyset} \tilde{v}_\lambda^2. \end{aligned} \tag{140}$$

It has been shown in Cohen, Dahmen and DeVore, 2002c that any tree $\mathcal{T}$ can be expanded to a tree $\tilde{\mathcal{T}}$ such that $\#(\tilde{\mathcal{T}}) \lesssim \#(\mathcal{T})$ but for any $\mu \in \mathcal{L}(\tilde{\mathcal{T}})$ only for a uniformly bounded finite number of $\lambda \in \mathcal{L}(\tilde{\mathcal{T}})$ one has $S_\lambda \cap S_\mu \neq \emptyset$. Hence a finite number of the terms $\tilde{v}_\lambda$ bound the local error on S_μ.

Encyclopedia of Computational Mechanics. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

A natural idea for constructing "good" meshes – or equivalently "good trees" identifying spans of wavelets – is to *equilibrate* these local errors. However, it turns out that this will not necessarily minimize the error $\|\mathbf{v} - \mathbf{v}|_{\mathcal{T}}\|_{\ell_2}$ over *all* trees of a fixed cardinality $N = \#(\mathcal{T})$, see Cohen, Dahmen and DeVore, 2002c. To formalize this, we define an error for N-term tree approximation which is the exact *tree analog* of the best (unconstrained) N–term approximation defined in (130).

$$\sigma^{\rm tree}_{N,\ell_2}(\mathbf{v}) := \min\{\|\mathbf{v} - \mathbf{w}\|_{\ell_2} : \ \mathcal{T} := \operatorname{supp}\mathbf{w} \ \text{ is a tree and } \ \#\mathcal{T} \le N\}. \tag{141}$$

Any minimizing tree will be denoted by $\mathcal{T}_N(\mathbf{v})$. We define now in analogy to (131) the sequence space

$$\mathcal{A}^s_{\rm tree}(\ell_2) := \{\mathbf{v} \in \ell_2 : \sigma^{\rm tree}_{N,\ell_2}(\mathbf{v}) \ \lesssim \ N^{-s}\}, \tag{142}$$

endowed with the quasi–norm

$$\|\mathbf{v}\|_{\mathcal{A}^s_{\rm tree}(\ell_2)} := \sup_{n\in\mathbb{N}} N^s \sigma^{\rm tree}_{N,\ell_2}(\mathbf{v}). \tag{143}$$

Analogously, we can define the counterpart $\mathcal{A}^s_{\rm tree}(\mathcal{H})$ in $\mathcal{H}$. As in (131) the error of tree approximation of $\mathbf{v} \in \mathcal{A}^s_{\rm tree}(\ell_2)$ decays like the error of the corresponding tree approximations to v in $\mathcal{H}$.

In spite of the conceptual similarity there is an important difference between best tree and best unconstrained N–term approximation. At least for any finitely supported $\mathbf{v}$ the latter one is easily determined by (quasi–) sorting by size – thresholding. Determining the best tree however, is much harder. However, since one obtains a near best approximation in the energy norm anyway we can be content with *near best* tree approximation in ℓ_2 as well. More precisely, given a fixed constant $C^* \ge 1$, a tree $\mathcal{T}(\eta, \mathbf{v})$ is called an (η, C^*)–near best tree for $\mathbf{v}$ if $\|\mathbf{v} - \mathbf{v}|_{\mathcal{T}(\eta,\mathbf{v})}\|_{\ell_2} \le \eta$ and whenever any other tree $\mathcal{T}$ satisfies $\|\mathbf{v} - \mathbf{v}|_{\mathcal{T}}\|_{\ell_2} \le \eta/C^*$ one has $\#(\mathcal{T}(\eta, \mathbf{v})) \le C^*\#(\mathcal{T})$. It is remarkable that, according to Binev and DeVore, 2002, such near best trees can be constructed in *linear time*. This can be achieved with the aid of modified thresholding strategies working in the present setting with the quantities $\tilde v^2_\lambda$ (138) as local error functionals. We shall invoke this method to construct near best trees.

Since the selections of terms are constrained by the imposed tree structure one always has

$$\sigma_{\#\mathcal{T}_\eta(\mathbf{v}),\ell_2}(\mathbf{v}) \le \|\mathbf{v} - \mathbf{v}|_{\mathcal{T}_\eta(\mathbf{v})}\|_{\ell_2}. \tag{144}$$

However, for a wide class of functions in $\mathcal{H}$ one actually does not loose too much with regard to an optimal work/accuracy rate. To explain this we consider again the above scenario $\mathcal{H} = H^t$. The following fact has been shown in Cohen, Dahmen and DeVore, 2002c.

Remark 6.1. *For $\mathcal{H} = H^t$ one has $B^{t+sd}_q(L_q) \hookrightarrow \mathcal{A}^s_{\rm tree}(\mathcal{H})$ whenever $q^{-1} < s + 1/2$.*

Thus as soon as the smoothness space is strictly left of the Sobolev embedding line its elements have errors of tree approximations that decay like $(\#\mathcal{T}_N(\mathbf{v}))^{-s}$, see Figure 9. Moreover, this rate is known to be sharp, i.e.

$$\sup_{\|v\|_{B^{t+sd}_q(L_q)}=1} \ \inf_N N^s \sigma_{N,H^t}(v) \gtrsim 1, \tag{145}$$

which means that on the class $B^{t+sd}_q(L_q)$, under the above restriction of q, tree approximations give the same asymptotic error decay as best N–term approximations. The smaller the

Encyclopedia of Computational Mechanics. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

discrepancy $\delta := s + \frac{1}{2} - \frac{1}{q} > 0$, the larger the space $B_q^{t+sd}(L_q)$ admitting stronger singularities. In fact, when $\sup\{s : u \in H^{t+sd}\}$ is strictly smaller than $\sup\{s : u \in B_q^{t+sd}(L_q)\}$ the asymptotic work/accuracy rate achieved by meshes corresponding to the trees $\mathcal{T}_N(\cdot)$ is strictly better than that for uniform mesh refinements. This is known to be the case, for instance, for solutions u of elliptic boundary value problems on Lipschitz domains when δ is sufficiently small, see Dahlke and DeVore, 1997.

Thus the question that guides the subsequent discussion can be formulated as follows: *Can one devise the routines* RES *and* COARSE *in such a way that the computational work and storage needed to produce the output* $\mathbf{u}(\epsilon)$ *of* SOLVE*, stays proportional to* $\epsilon^{-1/s}$*, uniformly in* ϵ*, whenever the unknown solution* u *belongs to* $\mathcal{A}^s_{\rm tree}(\mathcal{H})$*, or even to* $\mathcal{A}^s(\mathcal{H})$*?*

6.2. Realization of Residual Approximations

We shall always assume that we have full access to the given data $\mathbf{f}$. Depending on some target accuracy one should therefore think of $\mathbf{f}$ as a finite array that approximates some "ideal" data accurately enough. Moreover, these data are (quasi-)ordered by their modulus. Such a quasi-ordering, based on binary binning can be realized in linear time, see e.g Barinka, 2003. In particular, this allows us to obtain coarser approximations $\mathbf{f}_\eta$, satisfying $\|\mathbf{f} - \mathbf{f}_\eta\|_{\ell_2} \le \eta$ with the aid of the simplest version of the routine COARSE, realized by adding $|\mathbf{f}_\lambda|^2$ in direction of increasing size until the sum exceeds η^2, see Cohen, Dahmen and DeVore, 2001 for details. As a central task one further has to approximate the sequence $\mathbf{F}(\mathbf{v})$ for any given finitely supported input $\mathbf{v}$ which we shall now describe.

Linear Operators: It will be instructive to consider first the linear case $\mathbf{F}(\mathbf{v}) = \boldsymbol{A}\mathbf{v}$ when $\boldsymbol{A}$ is the wavelet representation of the underlying operator. We shall describe an algorithm for the fast computation of $\boldsymbol{A}\mathbf{v}$. So far, the only property of $\boldsymbol{A}$ that we have used is the norm equivalence (30). Now the cancellation properties (26) come into play. We have seen in Section 4.2 that they imply the quasi-sparsity of a wide class of linear operators. The relevant notion can be formulated as follows (Cohen, Dahmen and DeVore, 2001). *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 matrix* $\mathbf{C}_j$ *with the following properties: For some summable sequence* $(\alpha_j)_{j=1}^\infty$ $(\sum_j \alpha_j < \infty)$ $\mathbf{C}_j$ *is obtained by replacing all but the order of* $\alpha_j 2^j$ *entries per row and column in* $\mathbf{C}$ *by zero and satisfies*

$$\|\mathbf{C} - \mathbf{C}_j\| \le C\alpha_j 2^{-js}, \quad j \in \mathbb{N}. \tag{146}$$

Specifically, wavelet representations of differential and also the singular integral operators from Sections 4.2 and 4.1 fall into this category for values of s^*, that depend, in particular, on the regularity of the wavelets, see Cohen, Dahmen and DeVore, 2001; Dahlke, Dahmen and Urban, 2002; Stevenson, 2003.

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 $\mathbf{v}_{[j]} := \mathbf{v}_{2^j}$ $(\mathbf{v}_{[-1]} \equiv \mathbf{0})$ and define

$$\mathbf{w}_j := \boldsymbol{A}_j\mathbf{v}_{[0]} + \boldsymbol{A}_{j-1}(\mathbf{v}_{[1]} - \mathbf{v}_{[0]}) + \cdots + \boldsymbol{A}_0(\mathbf{v}_{[j]} - \mathbf{v}_{[j-1]}), \tag{147}$$

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

compression estimates yield for any fixed $s < s^*$

$$\|\boldsymbol{A}\mathbf{v} - \mathbf{w}_j\|_{\ell_2} \quad \leq \quad c \left(\underbrace{\|\mathbf{v} - \mathbf{v}_{[j]}\|_{\ell_2}}_{\sigma_{2^j,\ell_2}(\mathbf{v})} + \sum_{l=0}^{j} \alpha_l 2^{-ls} \underbrace{\|\mathbf{v}_{[j-l]} - \mathbf{v}_{[j-l-1]}\|_{\ell_2}}_{\lesssim \ \sigma_{2^{j-l-1},\ell_2}(\mathbf{v})} \right). \tag{148}$$

One can now exploit the *a-posteriori* information offered by the quantities $\sigma_{2^{j-l-1},\ell_2}(\mathbf{v})$ to choose the smallest j for which the right hand side of (148) 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 Cohen, Dahmen and DeVore, 2001; Barinka, Barsch, Charton, Cohen, Dahlke, Dahmen and Urban, 2001 for a detailed description, analysis and implementation) which we summarize as follows:

APPLY $[\eta, \boldsymbol{A}, \mathbf{v}] \to \mathbf{w}_\eta$: DETERMINES FOR ANY FINITELY SUPPORTED INPUT $\mathbf{v}$ A FINITELY SUPPORTED OUTPUT $\mathbf{w}_\eta$ SUCH THAT

$$\|\boldsymbol{A}\mathbf{v} - \mathbf{w}_\eta\|_{\ell_2} \leq \eta. \tag{149}$$

Depending on the compressibility range s^* this scheme can be shown to exhibit the same work/accuracy rate as the best (unconstrained) N–term approximation in ℓ_2 as stated by the folloing result, Cohen, Dahmen and DeVore, 2001.

Theorem 6.2. *Suppose that $\boldsymbol{A} \in \mathcal{C}_{s^*}$ and that for some $0 < s < s^*$, $\mathbf{v} \in \mathcal{A}^s(\ell_2)$. Then, $\boldsymbol{A}\mathbf{v}$ is also in $\mathcal{A}^s(\ell_2)$. Moreover, for any finitely supported $\mathbf{v}$ the output $\mathbf{w}_\eta = \text{APPLY}\,[\eta, \mathbf{C}, \mathbf{v}]$ satisfies:*

(i) $\|\mathbf{w}_\eta\|_{\mathcal{A}^s(\ell_2)} \ \lesssim \ \|\mathbf{v}\|_{\mathcal{A}^s(\ell_2)}$;

(ii) $\#\text{supp}\,\mathbf{w}_\eta \ \lesssim \ \|\mathbf{v}\|^{1/s}_{\mathcal{A}^s(\ell_2)} \eta^{-1/s}$, $\#\text{flops} \ \lesssim \ \#\text{supp}\,\mathbf{v} + \|\mathbf{v}\|^{1/s}_{\mathcal{A}^s(\ell_2)} \eta^{-1/s}$,

where the constants in these estimates depend only on s when s is small.

The above work count is based on the tacit assumption that the entries of $\boldsymbol{A}$ can be computed with sufficient accuracy on average at unit cost. This can be verified for constant coefficient differential operators and spline wavelets. In general, the justification of such an assumption is less clear. We shall return to this point later.

The Nonlinear Case – Prediction of Significant Coefficients: In this case the point of view changes somewhat. The question to be addressed first is the following:

Given any $\eta > 0$ and an (η, C^*)–near best tree $\mathcal{T}(\eta, \mathbf{v})$ of $\mathbf{v}$, find a possibly small tree $\mathcal{T}_\eta$ such that for some constant C

$$\mathcal{T}^*(C\eta, \mathbf{F}(\mathbf{v})) \subseteq \mathcal{T}_\eta, \tag{150}$$

where $\mathcal{T}^*(C\eta, \mathbf{F}(\mathbf{v}))$ is a smallest *tree realizing accuracy $C\eta$.*

Thus we are asking for quantitative estimates concerning the effect of a nonlinearity on contributions with different length scales, a question of central importance in several areas

Encyclopedia of Computational Mechanics. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

of applications such as turbulence analysis. Using trees now already anticipates the need for taking the (quasi-local) effect of higher frequency on lower ones into account.

Of course, tight estimates of that sort must incorporate some knowledge about the character of the nonlinearity. Nonlinearities of at most *power growth* have been studied recently in Cohen, Dahmen and DeVore, 2002c and we briefly review some of the main findings. For instance, when the operator involves a local composition operator G as in (105) "power growth" means that for some $p > 0$ $|\mathcal{G}^{(n)}(x)| \lesssim (1 + |x|)^{(p-n)_+}$. In fact, one can show that for $\mathcal{H} = H^t$ (on some domain of spatial dimension d) one has $\mathcal{G} : \mathcal{H} \to \mathcal{H}'$ provided that

$$p < p^* := \frac{d + 2t}{d - 2t} \quad \text{when} \quad t < d/2, \qquad p > 0 \quad \text{when} \quad t \geq d/2, \tag{151}$$

see Cohen, Dahmen and DeVore, 2002c. The analysis in Cohen, Dahmen and DeVore, 2002c covers a much wider class of nonlinear operators including those that depend on several components involving also derivatives of the arguments $G(D^{\alpha_1} v_1, \ldots, D^{\alpha_l} v_l)$. For instance, the convective term in the Navier Stokes equations is covered. In order to convey the main ideas while keeping the exposition as simple as possible we confine the subsequent discussion to the above special situation. Using the locality of the nonlinearity, the cancellation properties of the wavelets as well as certain norm equivalences for Besov spaces in terms of weighted sequence norms for the wavelet coefficients, one can derive estimates of the form

$$|\mathbf{F}(\mathbf{v})_\lambda| \lesssim \sup_{S_\mu \cap S_\lambda \neq \emptyset} |v_\mu| 2^{-\gamma(|\lambda| - |\mu|)}, \tag{152}$$

where for $\mathcal{H} = H^t$ a typical value for γ is $\gamma = t + \tilde{m} + d/2$. It measures in some sense the compressibility of the nonlinear map.

How to predict good trees for $\mathbf{F}(\mathbf{v})$ from those for $\mathbf{v}$ for the above mentioned type of nonlinearities can be sketched as follows, cf. Cohen, Dahmen and DeVore, 2002c. For a given target accuracy ϵ, $j = 0, 1, \cdots$ and a given $\mathbf{v}$, we consider the near best trees

$$\mathcal{T}_j := \mathcal{T}(\frac{2^j \epsilon}{1 + j}, \mathbf{v}), \tag{153}$$

and the corresponding expanded trees $\tilde{\mathcal{T}}_j$ mentioned before. By construction, these trees are nested in the sense that $\tilde{\mathcal{T}}_j \subset \tilde{\mathcal{T}}_{j-1}$. We shall use the difference sets

$$\Delta_j := \tilde{\mathcal{T}}_j \setminus \tilde{\mathcal{T}}_{j+1}. \tag{154}$$

in order to build a tree which will be adapted to $\mathbf{w} = \mathbf{F}(\mathbf{v})$. They represent the "energy" in $\mathbf{v}$ reflecting the next higher level of accuracy. Now we introduce the parameter

$$\alpha := \frac{2}{2\gamma - d} > 0, \tag{155}$$

where γ is the constant in (152) and for each $\mu \in \Delta_j$, we define the *influence set*

$$\Lambda_{\epsilon,\mu} := \{\lambda \ : \ S_\lambda \cap S_\mu \neq \emptyset \text{ and } |\lambda| \leq |\mu| + \alpha j\}. \tag{156}$$

Thus the amount αj by which the level $|\mu|$ is exceeded in $\Lambda_{\epsilon,\mu}$ depends on the "strength" of v_μ expressed by the fact that $\mu \in \Delta_j$. We then define $\mathcal{T}_\epsilon$ as the union of these influence sets

$$\mathcal{T}_\epsilon := \mathcal{J}_\phi \cup \left(\cup_{\mu \in \tilde{\mathcal{T}}_0} \Lambda_{\epsilon,\mu} \right). \tag{157}$$

The main result can then be stated as follows, Cohen, Dahmen and DeVore, 2002c.

Theorem 6.3. *Given any* $\mathbf{v}$ *and* $\mathcal{T}_\epsilon$ *defined by (157), we have the error estimate*

$$\|\mathbf{F}(\mathbf{v}) - \mathbf{F}(\mathbf{v})|_{\mathcal{T}_\epsilon}\|_{\ell_2} \lesssim \epsilon. \tag{158}$$

Moreover, if $\mathbf{v} \in \mathcal{A}^s_{\rm tree}(\ell_2)$ *with* $0 < s < \frac{2\gamma - d}{2d}$, *then we have the estimate*

$$\#(\mathcal{T}_\epsilon) \lesssim \|\mathbf{v}\|^{1/s}_{\mathcal{A}^s_{\rm tree}(\ell_2)} \epsilon^{-1/s} + \#(\mathcal{J}_\phi). \tag{159}$$

We therefore have $\mathbf{F}(\mathbf{v}) \in \mathcal{A}^s_{\rm tree}(\ell_2)$ *and*

$$\|\mathbf{F}(\mathbf{v})\|_{\mathcal{A}^s_{\rm tree}(\ell_2)} \lesssim 1 + \|\mathbf{v}\|_{\mathcal{A}^s_{\rm tree}(\ell_2)}. \tag{160}$$

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

This result provides the basis for the following evaluation scheme.

EVAL $[\epsilon, \mathbf{F}, \mathbf{v}] \to \mathbf{w}(\epsilon)$ PRODUCES FOR ANY FINITELY SUPPORTED VECTOR $\mathbf{v}$ A FINITELY SUPPORTED VECTOR $\mathbf{w}(\epsilon)$ SUCH THAT $\|\mathbf{w}(\epsilon) - \mathbf{F}(\mathbf{v})\|_{\ell_2} \le \epsilon$ USING THE FOLLOWING STEPS:

(1) INVOKE THE ALGORITHM IN BINEV AND DEVORE, 2002 TO COMPUTE THE TREES

$$\mathcal{T}_j := \mathcal{T}(\frac{2^j \epsilon}{C_0(j+1)}, \mathbf{v}), \tag{161}$$

WHERE $C_0 = C_0(\|\mathbf{v}\|)$ IS THE CONSTANT INVOLVED IN (158), FOR $j = 0, \ldots, J$, AND STOP FOR THE SMALLEST J SUCH THAT $\mathcal{T}_J$ IS EMPTY (WE ALWAYS HAVE $J \lesssim \log_2(\|\mathbf{v}\|/\epsilon)$).

(2) DERIVE THE EXPANDED TREES $\tilde{\mathcal{T}}_j$, THE LAYERS Δ_j AND THE OUTCOME TREE $\mathcal{T}_\epsilon$ ACCORDING TO (157).

(3) COMPUTE $\mathbf{F}(\mathbf{v})|_{\mathcal{T}_\epsilon}$ (APPROXIMATELY WITHIN ACCURACY ϵ).

Clearly any finitely supported $\mathbf{v}$ belongs to $\mathcal{A}^s_{\rm tree}(\ell_2)$ for every $s > 0$. Moreover, the trees $\mathcal{T}_j$ will be empty for $j \ge J$ and some $J \in \mathbb{N}$. Thus the scheme terminates after finitely many steps. We postpone some comments on step (3) to § 6.4. The following theorem summarizes the properties of Algorithm EVAL.

Theorem 6.4. *Given the inputs* $\epsilon > 0$, *a nonlinear function* $\mathcal{F}$ *such that* $\mathbf{F}$ *(satisfying assumptions of the type mentioned before), and a finitely supported vector* $\mathbf{v}$, *then the output tree* $\mathcal{T}_\epsilon$ *has the following properties: One has* $\|\mathbf{F}(\mathbf{v}) - \mathbf{F}(\mathbf{v})|_{\mathcal{T}_\epsilon}\|_{\ell_2} \le \epsilon$. *Furthermore, for any* $0 < s < \frac{2\gamma - d}{2d}$ *(see Theorem 6.3), one has*

$$\#(\mathcal{T}_\epsilon) \le C\|\mathbf{v}\|^{1/s}_{\mathcal{A}^s} \epsilon^{-1/s} + \#(\mathcal{J}_\phi) =: N_\epsilon \tag{162}$$

with C *a constant depending only on the constants appearing in Theorem 6.3. Moreover, the number of computations needed to find* $\mathcal{T}_\epsilon$ *is bounded by* $C(N_\epsilon + \#\mathcal{T}(\mathbf{v}))$, *where* N_ϵ *is the right hand side of (162) and* $\mathcal{T}(\mathbf{v})$ *is the smalles tree containing* $\operatorname{supp} \mathbf{v}$.

Finally, we need a praticular coarsening strategy that respects tree structures.

Encyclopedia of Computational Mechanics. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

COARSE $[\eta, \mathbf{w}] \to \bar{\mathbf{w}}_\eta$ 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 COARSE is based on the results in Binev and DeVore, 2002, which ensure linear complexity. This version of COARSE can also be used in the linear case. As such it can be used to show that Theorem 6.2 remains valid for compressible matrices when the spaces $\mathcal{A}^s(\ell_2)$ are replaced by $\mathcal{A}^s_{\rm tree}(\ell_2)$ (see Cohen, Dahmen and DeVore, 2002b).

The above results allow one to show that the scheme RES, in all the above examples satisfies the following:

Whenever the exact solution u of (90) belongs to $\mathcal{A}^s_{\rm tree}(\mathcal{H})$ for some $s < s^*$, then one has for any finitely supported input $\mathbf{v}$ and any tolerance $\eta > 0$ that the output $\mathbf{r}_\eta := \text{RES}\,[\eta, \mathbf{C}, \mathbf{F}, \mathbf{f}, \mathbf{v}]$ satisfies

$$\begin{aligned} \#\text{supp}\,\mathbf{r}_\eta &\leq C\eta^{-1/s}\left(\|\mathbf{v}\|^{1/s}_{\mathcal{A}^s_{\rm tree}(\ell_2)} + \|\mathbf{u}\|^{1/s}_{\mathcal{A}^s_{\rm tree}(\ell_2)} + 1\right), \\ \|\mathbf{r}_\eta\|_{\mathcal{A}^s_{\rm tree}(\ell_2)} &\leq C\left(\|\mathbf{v}\|_{\mathcal{A}^s_{\rm tree}(\ell_2)} + \|\mathbf{u}\|_{\mathcal{A}^s_{\rm tree}(\ell_2)} + 1\right), \end{aligned} \tag{163}$$

where (in addition to the dependence given in the previous theorems) C depends only on s when $s \to s^*$. Moreover, the number of operations needed to compute $\mathbf{w}_\eta$ stays proportional to $\#\text{supp}\,\mathbf{r}_\eta$.

One can show that the number of perturbed updates in step (II) of SOLVE executed before branching off into a coarsening step (III), remains uniformly bounded independent of the data and of the target accuracy ϵ. Therefore the s^*-sparsity of the rountine RES ensures that the $\mathcal{A}^s_{\rm tree}(\ell_2)$-norms of the approximations $\mathbf{v}^k$ remain bounded in each update block (II). The coarsening step is applied exactly in order to prevent the constants in these estimates from building up over several subsequent update blocks (II). This is the consquence of the following Coarsening Lemma, Cohen, Dahmen and DeVore, 2002b.

Proposition 6.5. *If $\mathbf{v} \in \mathcal{A}^s_{\rm tree}(\ell_2)$ and $\|\mathbf{v} - \mathbf{w}\|_{\ell_2} \leq \eta$ with $\#\text{supp}\,\mathbf{w} < \infty$. Then $\bar{\mathbf{w}}_\eta :=$ COARSE $[\mathbf{w}, C^*\eta]$ satisfies*

$$\#\text{supp}\,\bar{\mathbf{w}}_\eta \lesssim \|\mathbf{v}\|^{1/s}_{\mathcal{A}^s_{\rm tree}(\ell_2)}\eta^{-1/s}, \quad \|\mathbf{v} - \bar{\mathbf{w}}_\eta\|_{\ell_2} \leq (1 + C^*)\eta,$$

and

$$\|\bar{\mathbf{w}}_\eta\|_{\mathcal{A}^s_{\rm tree}(\ell_2)} \lesssim \|\mathbf{v}\|_{\mathcal{A}^s_{\rm tree}(\ell_2)},$$

where C^* is the constant from the near best tree construction scheme in Binev and DeVore, 2002.

One can then use the above building blocks to show that SOLVE is optimal in the following sense, Cohen, Dahmen and DeVore, 2002b.

Theorem 6.6. *If the exact solution $u = \sum_{\lambda\in\mathcal{J}} u_\lambda\psi_\lambda$ belongs to $\mathcal{A}^s_{\rm tree}(\mathcal{H})$, for any $s < s^*(F, \Psi)$, then, for any target accuracy $\epsilon > 0$, the approximations $\mathbf{u}(\epsilon)$ produced by* SOLVE *satisfy*

$$\|u - \sum_\lambda \bar{u}(\epsilon)_\lambda\psi_\lambda\|_H \leq C_\Psi\epsilon,$$

Encyclopedia of Computational Mechanics. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

and

$$\#\mathrm{supp}\,\bar{\mathbf{u}}(\epsilon),\ \text{comp. work} \lesssim \epsilon^{-1/s}.$$

The above results cover the examples from Section 5.2, in particular the mixed formulations. Note that there is no restriction on the choice of wavelet bases for the different solution components such as velocity and pressure in the Stokes problem. In contrast, in the classical approach a compatible choice verifying the LBB condition is essential. In the adaptive context such constraints become void. In addition to these qualitative asymptotic results, the experiments in Dahlke, Dahmen and Urban, 2002; Dahmen, Urban and Vorloeper, 2002 show that also quantitatively the performence of the adaptive scheme stays the same even when wavelet bases for velocity and pressure are used that, in connection with an a–priori choice of finite dimensional trial spaces, would violate the LBB condition.

6.3. The Newton Scheme

The scheme SOLVE is based on an ideal iteration of order one. The choice $\mathbf{C}_n := (D\mathbf{F}(\mathbf{u}^n))^{-1}$ offers in principle a better convergence behavior of the outer iteration. In fact, for problems of the type (105) one can show that the Newton iteration (121) converges quadratically for a sufficiently good initial guess $\mathbf{u}^0$. On the other hand, it is not clear what the cost of each linear subproblem

$$D\mathbf{F}(\mathbf{u}^n)\mathbf{w}^n = -(\mathbf{F}(\mathbf{u}^n) - \mathbf{f}) \tag{164}$$

will amount to. A detailed analysis is given in Cohen, Dahmen and DeVore, 2002b where it is shown that the perturbed Newton iteration still retains a quadratic convergence while preserving an overall asymptoticaly optimal complexity in the sense of Theorem 6.6. It is perhaps worth stressing the following two points. The well–posedness (109) ensures that the problem (164) is well conditioned which suggests employing SOLVE for its approximate solution. Nevertheless, this raises two questions. First, the scheme APPLY needed to realize the residual approximation in SOLVE would require assembling in some sense the matrix $D\mathbf{F}(\mathbf{u}^n)$ in each update step with sufficient accuracy, which could be prohibitively expensive. However, the result of the application of $D\mathbf{F}(\mathbf{u}^n)$ to a vector $\mathbf{w}$ can be interpreted as the array of dual wavelet coefficients of a nonlinear composition with *two* wavelet expansions, since

$$D\mathbf{F}(\mathbf{u}^n)\mathbf{w} = (\langle \psi_\lambda, DF(u^n)w\rangle)_{\lambda\in\mathcal{J}} =: (\langle \psi_\lambda, Q(u^n, w)\rangle)_{\lambda\in\mathcal{J}} =: \boldsymbol{Q}(\mathbf{u}^n, \mathbf{w}).$$

The approximation of $\boldsymbol{Q}(\mathbf{u}^n, \mathbf{w})$ can be realized again with the aid of the scheme EVAL without assembling the Jacobian which in this sense leads to a *matrix free* Newton scheme. In fact, in Cohen, Dahmen and DeVore, 2002b EVAL is described for functions of several components. Further remarks on the related computational issues will be given in the next Section.

Secondly, the complexity analysis is not completely straightforward because one cannot directly infer from the sparseness of $\mathbf{u}$ to the sparseness of the Newton systems (164). However, one can show that these systems may be viewed as perturbations of another system whose solution is sparse whenever $\mathbf{u}$ is sparse. This, however, limits the target accuracy for (164). Nevertheless, one can show that this still suffices to ensure second order convergence of the outer iteration (121), see Cohen, Dahmen and DeVore, 2002b.

6.4. Computational Issues

The above complexity analysis works under the assumption that in the linear case the entries of $\boldsymbol{A}$ are computatble at unit cost in order to invoke APPLY. Likewise in the nonlinear case, the

Encyclopedia of Computational Mechanics. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

entries of $\mathbf{F}(\mathbf{v})$ in the predicetd sets $\mathcal{T}_\epsilon$ have to be computed. Under fairly general assumptions both tasks can be handled by the following strategy. By definition, one has

$$\mathcal{F}(v) = \sum_{\lambda\in\mathcal{J}} \langle \psi_\lambda, \mathcal{F}(v)\rangle \tilde{\psi}_\lambda = \sum_{\lambda\in\mathcal{J}} (\mathbf{F}(\mathbf{v}))_\lambda \tilde{\psi}_\lambda,$$

where $\tilde{\Psi}$ is the dual basis for Ψ and hence a Riesz basis for $\mathcal{H}'$. The idea is now to use an efficient *recovery scheme*, as desribed in Dahmen, Schneider and Xu, 2000; Barinka, Dahmen, Schneider and Xu, 2003, that produces for a given target accuracy ϵ an approximation $g = \sum_{\lambda\in\Lambda_\epsilon} g_\lambda \tilde{\psi}_\lambda \in \mathcal{H}'$ to $\mathcal{F}(v)$ such that $\|\mathcal{F}(v) - g\|_{\mathcal{H}'} \le \epsilon$ at a computational cost that stays proportional to the size of the prediction set $\mathcal{T}_\epsilon$ from EVAL. The norm equivalence now guarantees that the corresponding coefficient arrays exhibit essentially the same accuracy $\|\mathbf{F}(\mathbf{v}) - \mathbf{g}\|_{\ell_2} \le c_\Psi^{-1}\epsilon$. The important point is that individual coefficients are never computed but, solely based on the knowledge of the prediction set $\mathcal{T}_\epsilon$, quadrature is always used to generate on the function side an approximation to the whole object $\mathcal{F}(v)$ by local quasi-interpolant techniques to be able to keep the computational work proportional to the number of current degrees of freedom. This strategy justifies the above assumptions in a wide range of cases, see Barinka, 2003; Dahmen, Schneider and Xu, 2000; Barinka, Dahmen, Schneider and Xu, 2003 for details.

6.5. Concluding Remarks

The primary goal of this section was to bring out the essential mechanisms and the potential of wavelet based multiscale techniques and to understand under which circumstances optimal complexity can be achieved. A crucial role was played by the mapping properties of the underlying variational problem in conjunction with the availability of a wavelet Riesz basis for the corresponding energy space. This also revealed where principal difficulties may arise. Depending on the nature of $\mathcal{H}$ or when dealing with complex geometries the construction of a good basis Ψ may be very difficult or even impossible. Poor constants in (30) would spoil the quantitative performance of SOLVE significantly. Likewise poor constants in (92) would have the same effect. In fact, these constants may be parameter dependent and further work in this direction is under progress. But at least, the analysis reveals what one should be looking for in each concrete case which might certainly require much more additional work. More information on first computational studies for elliptic but also indefinite problems can be found in Barinka, Barsch, Charton, Cohen, Dahlke, Dahmen and Urban, 2001; Dahlke, Dahmen and Urban, 2002; Dahmen, Urban and Vorloeper, 2002.

The quantitative improvement of evaluation schemes like EVAL in conjunction with the strategies in Barinka, 2003; Dahmen, Schneider and Xu, 2000; Barinka, Dahmen, Schneider and Xu, 2003 certainly plays an important role. But already the pure prediction result in Theorem 6.3 at least gives rigorous bounds on the effect of certain nonlinearities concerning the interaction of fine and coarse scales – a problem that is at the heart of many multiscale phenomena in technology and science.

Finally, the difficulties of finding stable pairs of trial functions in many mixed formulations may help one to appreciate the principal merits of techniques that inherit the stability properties of the original problem. In this context the above multiscale techniques incorporate a natural stabilization effect.

Encyclopedia of Computational Mechanics. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

ACKNOWLEDGEMENT

We are very indebted to Frank Bramkamp, Helmut Harbrecht, Siegfried Müller and Reinhold Schneider for providing us with illustrations and numerical results concerning the material discussed in this chapter.

REFERENCES

R.A. Adams, *Sobolev Spaces*, Academic Press, 1978.

B. Alpert, A class of bases in L_2 for sparse representation of integral operators, SIAM J. Math. Anal., 24 (1993), 246–262.

F. Arandiga, R. Donat, A. Harten, Multiresolution based on weighted averages of the hat function I: Linear reconstruction techniques, SIAM J. Numer. Anal., 36 (1998), 160–203.

F. Arandiga, R. Donat, A. Harten, Multiresolution based on weighted averages of the hat function II: Nonlinear reconstruction techniques, SIAM Sci. Comput., 20 (1999), 1053–1093.

J. Ballmann, F. Bramkamp, S. Müller, Development of a flow solver employing local adaptation based on multiscale analysis based on B–spline grids, in: *Proceedings of the 8th Annual Conference of the CFD Society of Canada, June 11 to 13, 2000, Montreal, 2000.*

F.Bramkamp, B. Gottschlich-Müller, M. Hesse, Ph. Lamby, S. Müller, J. Ballmann, K.-H. Brakhage, W. Dahmen, H–adaptive multiscale schemes for the compressible Navier Stokes equations – Polyhedral discretization, data compression and mesh generation, Notes on Numerical Fluid Mechanics, Flow Modulation and Fluid-Structure-Interaction at Airplane Wings (ed. by J. Ballmann), Springer-Verlag, 125–204.

A. Barinka, *Fast Evaluation Tools for Adaptive Wavelet Schemes*, PhD Thesis, RWTH Aachen, in preparation.

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. Sci. Comp., 23, No. 3 (2001), 910–939.

A. Barinka, S. Dahlke, W. Dahmen, Adaptive Application of Operators in Standard Representation, in preparation.

A. Barinka, W. Dahmen, R. Schneider, Y. Xu, An algorithm for the evaluation of nonlinear functionals of wavelet expansions, in preparation

M. Berger, J. Oliger, Adaptive mesh refinement for hyperbolic partial differential equations, J. Comp. Physics, 53 (1984), 484–512.

J. Bergh, J. Löfström, *Interpolation Spaces, An Introduction*, Springer, 1976.

S. Bertoluzza, A-posteriori error estimates for wavelet Galerkin methods, Appl Math. Lett., 8 (1995), 1–6.

S. Bertoluzza, An adaptive collocation method based on interpolating wavelets, in: W. Dahmen, A.J. Kurdila and P.Oswald, eds., *Multiscale Wavelet Methods for PDE's* (1997), Academic Press, New York.

G. Beylkin, R. R. Coifman, V. Rokhlin, Fast wavelet transforms and numerical algorithms I, Comm. Pure and Appl. Math., 44 (1991), 141–183.

P. Binev and R. DeVore, Fast Computation in Adaptive Tree Approximation ,University of South Carolina IMI Preprint vol 2002:11, to appear in Numer. Math.

J.H. Bramble, R.D. Lazarov, and J.E. Pasciak, A least-squares approach based on a discrete minus one inner product for first order systems, Math. Comput. 66 (1997), 935–955.

F. Brezzi and M. Fortin, *Mixed and Hybrid Finite Element Methods*, Springer, 1991.

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

A. Canuto, A. Tabacco, K. Urban, The wavelet element method, part II: Realization and additional features, Appl. Comp. Harm. Anal. 8 (2000), 123–165.

J.M. Carnicer, W. Dahmen and J.M. Peña, Local decomposition of refinable spaces, Appl. Comp. Harm. Anal. 3 (1996), 127-153.

A. Cohen, Wavelet methods in numerical analysis, in the Handbook of Numerical Analysis, vol. VII, P.-G. Ciarlet et J.-L. Lions eds., Elsevier, Amsterdam, 2000.

A. Cohen, Numerical analysis of wavelet methods, Studies in Mathematics and its Applications, Vol. 32, Elsevier, Amsterdam, 2003.

A. Cohen, I. Daubechies, J.-C. Feauveau, Biorthogonal bases of compactly supported wavelets, Comm. Pure and Appl. Math., 45 (1992), 485–560.

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

A. Cohen, W. Dahmen, R. DeVore, Adaptive wavelet methods II - Beyond the elliptic case, Foundations of Comp. Math., 2, (no. 3) (2002), 203–245.

A. Cohen, W. Dahmen, R. DeVore, Adaptive Wavelet Schemes for Nonlinear Variational Problems, IGPM Report # 221, RWTH Aachen, July 2002, to appear in SIAM J. Numer. Anal.

A. Cohen, W. Dahmen, R. DeVore, Sparse evaluation of nonlinear functionals of multiscale expansions, IGPM Report # 222, RWTH Aachen, July 2002, to appear in SIAM J.Math. Anal.

A. Cohen, S.-M. Kaber, S. Müller, M. Postel, Fully adaptive multiscale methods for conservation laws, Math. Comp., 72 (2002), 183–225.

A. Cohen, R. Masson, Wavelet adaptive methods for second order elliptic problems, boundary conditions and domain decomposition, Numer. Math., 8 (1997), 21–47.

M. Costabel, E.P. Stephan, Coupling of finite and boundary element methods for an elastoplastic interface problem, SIAM J. Numer. Anal. 27 (1990), 1212–1226.

S. Dahlke: Besov regularity for elliptic boundary value problems on polygonal domains, Appl. Math. Lett., 12 (1999), 31–36.

S. Dahlke, R. DeVore, Besov regularity for elliptic boundary value problems, Comm. Partial Differential Equations, 22 (1997), 1–16.

S. Dahlke, W. Dahmen, R. Hochmuth, R. Schneider, *Stable multiscale bases and local error estimation for elliptic problems*, Appl. Numer. Maths. 8 (1997), 21–47.

S. Dahlke, W. Dahmen, K. Urban, Adaptive wavelet methods for saddle point problems – Convergence rates, SIAM J. Numer. Anal., 40 (no. 4) (2002), 1230–1262.

S. Dahlke, R. Hochmuth, K. Urban, Adaptive wavelet methods for saddle point problems, Math. Model. Numer. Anal. (M2AN), 34 (no. 5) (2000), 1003-1022.

W. Dahmen, Some remarks on multiscale transformations, stability and biorthogonality, in: *Wavelets, Images and Surface Fitting*, P.J. Laurent, A. Le M´**e**haut´**e**, L.L. Schumaker (eds.), AK Peters, Wellesley, Massachusetts, 1994, 157-188.

W. Dahmen, Stability of multiscale transformations, Journal of Fourier Analysis and Applications, 2 (1996), 341-361.

W. Dahmen, Wavelet and Multiscale Methods for Operator Equations, Acta Numerica, Cambridge University Press, 6 (1997), 55–228.

W. Dahmen, Wavelet methods for PDEs – Some recent developments, J. Comp. Appl. Math., 128 (2001), 133–185.

Encyclopedia of Computational Mechanics. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

W. Dahmen, Multiscale and Wavelet Methods for Operator Equations, C.I.M.E. Lecture Notes, Springer–Verlag, Heidelberg.

W. Dahmen, B. Gottschlich-Müller, S. Müller, Multiresolution schemes for conservation laws, Numer. Math., 88 (2001), 399-443.

W. Dahmen, H. Harbrecht, R. Schneider, Compression Techniques for Boundary Integral Equations – Optimal Complexity Estimates, IGPM Report # 218, RWTH Aachen, June 2002.

W. Dahmen, A. Kunoth, Multilevel preconditioning, Numer. Math., 63 (1992), 315–344.

W. Dahmen, A. Kunoth, Appending boundary conditions by Lagrange multipliers: Analysis of the LBB condition, Numer. Math., 88 (2001), 9–42.

W. Dahmen, A. Kunoth, Adaptive Wavelet Methods for Linear–Quadratic Elliptic Control Problems: Convergence Rates, IGPM Report # 224, RWTH Aachen, December 2002.

W. Dahmen, A. Kunoth, R. Schneider, Wavelet least squares methods for boundary value problems, Siam J. Numer. Anal. 39, (no. 6) (2002), 1985–2013.

W. Dahmen, A. Kunoth, and K. Urban, Biorthogonal spline-wavelets on the interval — Stability and moment conditions, Appl. Comp. Harm. Anal., 6 (1999), 132–196.

W. Dahmen, S. Prößdorf, R. Schneider, Multiscale methods for pseudo-differential equations on smooth manifolds, in: *Proceedings of the International Conference on Wavelets: Theory, Algorithms, and Applications*, C.K. Chui, L. Montefusco, L. Puccio (eds.), Academic Press, 1994, 385-424.

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

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

W. Dahmen, R. Schneider, Y. Xu, Nonlinear functions of wavelet expansions – Adaptive reconstruction and fast evaluation, Numer. Math., 86 (2000), 49–101.

W. Dahmen and R. Stevenson, Element-by-element construction of wavelets – stability and moment conditions, SIAM J. Numer. Anal., 37 (1999), 319–325.

W. Dahmen, K. Urban, J. Vorloeper, Adaptive wavelet methods – basic consepts and applications, in: Wavelet Analysis – Twenty Years Developments, Ding–Xuan Zhou ed., World Scientific, New Jersey, 2002, 39–80.

I. Daubechies, *Ten Lectures on Wavelets*, CBMS-NSF Regional Conference Series in Applied Math. 61, SIAM, Philadelphia, 1992.

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

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

R. DeVore, B. Jawerth, and V. Popov, Compression of wavelet decompositions, Amer. J. Math., 114 (1992), 737–785.

R. DeVore, G.G. Lorentz, *Constructive Approximation*, Grundlehren vol. 303, Springer-Verlag, Berlin, 1993.

R. DeVore, R. Sharpley, Maximal functions measuring smoothness, Memoirs of the American Mathematical Society, 47 (No 293), 1984.

V. Girault and P.-A. Raviart, *Finite Element Methods for Navier-Stokes Equations*, Springer, 1986.

R. Glowinski and V. Girault, Error analysis of a fictitious domain method applied to a Dirichlet problem, Japan J. Industr. Appl. Maths., 12 (1995), 487–514.

L. Greengard and V. Rokhlin, A fast algorithm for particle simulation, J. Comput. Phys., 73 (1987), 325–348.

Encyclopedia of Computational Mechanics. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

W. Hackbusch, A sparse matrix arithmetic based on $\mathcal{H}$-matrices. Part I: Introduction to $\mathcal{H}$-matrices, Computing, 64 (1999), 89–108.

W. Hackbusch and Z.P. Nowak. On the fast matrix multiplication in the boundary element method by panel clustering, Numer. Math., 54 (1989), 463–491.

H. Harbrecht, *Wavelet Galerkin schemes for the boundary element method in three dimensions*, *PHD Thesis, Technische Universität Chemnitz, Germany*, 2001.

H. Harbrecht, private communication.

A. Harten, Discrete multiresolution and generalized wavelets, J. Appl. Num. Math., 12 (1993), 153–193.

A. Harten, Multiresolution algorithms for the numerical solution of hyperbolic conservation laws, Comm. Pure Appl. Math., 48 (1995), 1305–1342.

A. Harten, Multiresolution representation of data II: generalized framework, SIAM J. Numer. Anal., 33 (1996), 1205–1256.

C. Lage, C. Schwab, Wavelet Galerkin algorithms for boundary integral equations, SIAM J. Sci. and Stat. Computing, 20(1998), 2195-2222.

C. Lage, Concept oriented Design of numerical Software, in: *Boundary Elements: Implementation and Analysis of Advanced Algorithms*, Proceedings of the 12th GAMM–Seminar, Kiel, Germany, January 19–21, 1996, Vieweg Notes Numer. Fluid Mech., 54 (1996), 159–170.

R. Kress, *Linear Integral Equations*, Springer-Verlag, Berlin-Heidelberg, 1989.

A. Kunoth, Multilevel preconditioning—Appending boundary conditions by Lagrange multipliers, Advances in Computational Mathematics, 4 (1995), 145–170.

A. Kunoth, *Wavelet Methods – Elliptic Boundary Value Problems and Control Problems*, Advances in Numerical Mathematics, (H.G. Bock, W. Hackbusch, M. Luskin, W. Hackbusch, eds.) Teubner, Stuttgart-Leipzig-Wiesbaden, 2001.

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

S. Müller, *Adaptive Multiscale Schemes for Conservation Laws*, Lecture Notes in Computational Science and Engineering, Vol. 27, Springer-Verlag, Berlin Heidelberg, 2003.

T. von Petersdorff, C. Schwab, Wavelet approximation for first kind integral equations on polygons, Numer. Math., 74 (1996), 479–519.

T. von Petersdorff, C. Schwab, Fully discrete multiscale Galerkin BEM, in: *Multiscale Wavelet Methods for PDEs*, W. Dahmen, A. Kurdila, P. Oswald (eds.), Academic Press, San Diego, 1997, 287–346.

T. von Petersdorff, R. Schneider, C. Schwab, Multiwavelets for second kind integral equations, SIAM J. Numer. Anal., 34 (1997), 2212–2227.

G. Schmidlin, C. Lage, C. Schwab, Rapid solution of first kind boundary integral equations in $\mathbb{R}^3$, Research Report No. 2002-07, Seminar für Angewandte Mathematik, ETH Zürich.

R. Schneider. *Multiskalen- und Wavelet-Matrixkompression: Analysisbasierte Methoden zur Lösung großer vollbesetzter Gleichungssysteme*, B.G. Teubner, Stuttgart, 1998.

R. Stevenson, Locally supported, piecewise polynomial biorthogonal wavelets on non-uniform meshes, Technical Report 1157, University of Utrecht, 2000, to appear in Constructive Approximation.

R. Stevenson, On the compressibility of operators in wavelet coordinates, Technical Report 1249, University of Utrecht, 2002, to appear in SIAM J. Math. Anal.

W. Sweldens, The lifting scheme: A custom-design construction of biorthogonal wavelets, Appl. Comput. Harm. Anal., 3 (1996), 186-200.

W. Sweldens, The lifting scheme: A construction of second generation wavelets, SIAM J. Math. Anal., 29 (1998), 511–546.

Encyclopedia of Computational Mechanics. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

H. Triebel, *Interpolation Theory, Function Spaces, and Differential Operators*, North-Holland, Amsterdam, 1978.

P.S. Vassilevski, J. Wang, Stabilizing the hierarchical basis by approximate wavelets, I: Theory, Numer. Linear Algebra Appl. 4 (1997), 103–126.

H. Yserentant, On the multilevel splitting of finite element spaces, Numer. Math. 49 (1986), 379–412.

Encyclopedia of Computational Mechanics. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.