Acta Numerica (1997), *pp.* 55–228

Wavelet and multiscale methods for operator equations

Wolfgang Dahmen
Institut für Geometrie und Praktische Mathematik
RWTH Aachen
Templergraben 55, 52056 Aachen, Germany
E-mail: dahmen@igpm.rwth-aachen.de

CONTENTS

1 Introduction 55
2 The scope of problems 72
3 Multiscale decompositions of refinable spaces 79
4 Examples 91
5 Norm equivalences and function spaces 107
6 Preconditioning 116
7 The ideal setting 138
8 Extension to more general domains 175
9 Pseudo-differential and boundary integral equations 182
10 Wavelets on manifolds and domain decomposition 197
11 Analysis of adaptive schemes for elliptic problems 206
12 What else? 216
References 218

1. Introduction

More than anything else, the increase of computing power seems to stimulate the greed for tackling ever larger problems involving large-scale numerical simulation. As a consequence, the need for understanding something like the *intrinsic complexity* of a problem occupies a more and more pivotal position. Moreover, computability often only becomes feasible if an algorithm can be found that is *asymptotically optimal.* This means that storage and the number of floating point operations needed to resolve the problem with desired accuracy remain *proportional* to the problem size when the resolution of the discretization is refined. A significant reduction of complexity

is indeed often possible, when the underlying problem admits a continuous model in terms of differential or integral equations. The physical phenomena behind such a model usually exhibit characteristic features over a wide range of scales. Accordingly, the most successful numerical schemes exploit in one way or another the interaction of different scales of discretization. A very prominent representative is the *multigrid* methodology; see, for instance, Hackbusch (1985) and Bramble (1993). In a way it has caused a breakthrough in numerical analysis since, in an important range of cases, it does indeed provide asymptotically optimal schemes. For closely related multilevel techniques and a unified treatment of several variants, such as *multiplicative* or *additive subspace correction* methods, see Bramble, Pasciak and Xu (1990), Oswald (1994), Xu (1992), and Yserentant (1993). Although there remain many unresolved problems, multigrid or multilevel schemes in the classical framework of finite difference and finite element discretizations exhibit by now a comparatively clear profile. They are particularly powerful for elliptic and parabolic problems.

1.1. Is there any vision?

Much more recently, the wavelet concept has (at least initially) raised high expectations. Traditional primary applications of wavelets have been signal analysis/processing, image processing/compression, *etc.* What are the reasons for the recent explosion of activities centred upon wavelets in connection with the numerical treatment of PDEs? Of course, anything that looks new inspires curiosity: there is certainly a bandwagon effect. Also, mathematical beauty plays a role. Perhaps it is just a fashionable new wave that will soon come to rest. In any case, comparisons of wavelet methods with conventional schemes should help in finding an answer. However, it is not that simple. First of all, the picture of wavelet concepts appears to be still quite fuzzy for several reasons. On one hand, at the present stage there simply do not yet exist complete software packages for complex real life problems, which would admit fair performance comparisons. On the other hand, the development of concepts and ideas is still far from steady state.

To find a reasonable path through the jungle, it is therefore worth spending some time on what could actually be expected.

First of all, is there any need to look at alternatives to multigrid? Of course, curiosity is a perfect reason. But looking again more closely at the multigrid methodology, its performance is best understood with respect to *uniform mesh refinements* and asymptotic optimality in the above sense refers to such settings. However, a fully refined mesh may not be necessary to resolve sufficiently the desired solution. To avoid this potential waste, *adaptive* techniques have to be and are indeed employed. There are many possibilities ranging from *a priori* local mesh refinements to fully self-adaptive

schemes. In this regard, problems of a different type are encountered. On one hand, we need a thorough analysis to control the local refinement steps. The corresponding local information is usually *implicit*: it is derived by comparing different discretizations. At this point some heuristics usually enter. On the other hand, the mesh refinements cause *geometrical* problems that have nothing to do with the underlying problem. Overall, these matters appear to be somewhat better understood for the *additive* version (Oswald 1994) (which by the way is closer to the wavelet concept), while otherwise the multiplicative version is often more efficient. So there still appears to be a strong need for a better understanding of adaptivity in this context, both with regard to the underlying analytical concepts and to the corresponding data structures.

On the other hand, by their very nature, wavelet representations have a naturally built-in adaptivity through their ability to directly express and separate components living on different scales. This, combined with the fact that many operators and their inverses have (nearly) sparse representations in wavelet coordinates, may eventually lead to competitive or even superior schemes with regard to the following goal: keep the computational work proportional to the number of *significant* terms in the wavelet expansion of the searched object, which in some sense should reflect its intrinsic complexity; see, for example, Beylkin and Keiser (1997) and Dahlke, Dahmen, Hochmuth and Schneider (1997*b*).

The potential of this point of view will be one of the main themes of subsequent discussions. Wavelets are in some sense much more sophisticated tools than conventional discretizations. It will be seen that this also facilitates a refined analysis. One central objective of this paper is to highlight some of the underlying driving analytical mechanisms.

The price of a powerful tool is the effort required to construct and understand it. Its successful application hinges on the realization of a number of requirements. Some space has to be reserved for a clear identification of these requirements as well as for their realization. This is also particularly important for understanding the severe *obstructions* that keep us at present from readily materializing all the principally promising perspectives.

These obstructions are to a great extent related to constraints imposed by domain geometries. There may be a good chance to reduce many problems to a *periodized* one (by an additional separate treatment of boundary conditions). In the periodic case *ideal wavelets* are available. Nevertheless, there will still remain important problem classes for which this strategy does not work. Therefore I will deviate from the usual way of motivating and developing wavelet concepts by means of Fourier analysis. Instead, some effort will be spent on formulating a sufficiently general and flexible framework of multiresolution decompositions that can host a variety of specializations. Moreover, appropriate substitutes for the Fourier tools have to be developed.

Wavelets are traditionally associated with *orthonormal* bases. A closer look reveals that orthogonality is often convenient but *not essential.* In presenting the material I will deviate sometimes from the original sources by formulating things in the more flexible context of *biorthogonality.* While this often supports *locality* and helps to bring out what really *is* essential, it will also be seen sometimes to be simply better, and even offers interesting new combinations of different concepts.

Of course, the acceptance of new concepts increases with their practical success. Somehow the measure is set by the existing modern multigrid techniques. The competition between different methodologies can be very stimulating. It should not be the primary point of view though. I personally believe that the additional insight gained from different, yet related, concepts will be mutually beneficial. Perhaps at a later stage, a marriage of complementary components and an enriched supply of tools will lead to true improvements.

As mentioned before, the presentation of material will necessarily be very selective. The selection criteria will not include optimal performance in existing algorithms, but will instead attempt to bring out ideas and concepts that bear some potential for future developments or, on the other hand, explain inherent limitations. Last but not least, my ignorance is to blame. I apologize to all those whose contributions do not get a proper share.

I shall next briefly discuss some simple examples in connection with admittedly trivial problems. Their purpose is only to help in identifying a few characteristic features that will then serve as a guideline for subsequent developments.

1.2. The Haar basis

The scaled shifts

$$\phi_{j,k} = 2^{j/2}\phi\left(2^j \cdot -k\right), \quad k = 0, \ldots 2^j - 1,$$

of the *box function*

$$\phi(x) = \begin{cases} 1, & 0 \le x \le 1, \\ 0, & \text{else}, \end{cases} \tag{1.1}$$

form an *orthonormal basis* of their linear span S_j relative to the standard inner product $\langle f, g\rangle = \langle f, g\rangle_{[0,1]} := \int_0^1 f(x)g(x)\,\mathrm{d}x$. Since

$$\phi(x) = \phi(2x) + \phi(2x-1),$$

so that

$$\phi_{j,k} = \frac{1}{\sqrt{2}}(\phi_{j+1,2k} + \phi_{j+1,2k+1}), \tag{1.2}$$

the S_j are nested and the closure of their union relative to $\|\cdot\|_{L_2([0,1])} := \langle\cdot,\cdot\rangle^{1/2}$ is all of $L_2([0,1])$ (the space of square integrable functions on $[0,1]$). Thus, denoting by $P_j f$ the orthogonal projection of f onto S_j, one has the representation

$$f = P_0 f + \sum_{j=1}^{\infty} (P_j - P_{j-1}) f.$$

The components $(P_j - P_{j-1})f$ represent the *'detail'* added to a given approximation when progressing to the next higher scale of discretization. In the present situation it can be conveniently encoded by the functions

$$\psi_{j,k} := \frac{1}{\sqrt{2}}(\phi_{j+1,2k} - \phi_{j+1,2k+1}), \quad k = 0, \ldots, 2^j - 1, \tag{1.3}$$

where, as before, $\psi_{j,k} := 2^{j/2}\psi(2^j \cdot - k)$, and $\psi(x) := \phi(2x) - \phi(2x-1)$. In fact, one easily verifies that

$$\langle \phi(\cdot - k), \psi(\cdot - l)\rangle = 0, \quad \langle \psi(\cdot - k), \psi(\cdot - l)\rangle = \delta_{k,l}, \quad k, l = 0, \ldots, 2^j - 1, \tag{1.4}$$

so that

$$\langle \psi_{j,k}, \psi_{n,l}\rangle = \delta_{j,n}\delta_{k,l}. \tag{1.5}$$

Thus

$$\Psi := \{\phi\} \cup \{\psi_{j,k} : k = 0, \ldots, 2^j - 1,\ j = 0, 1, 2, \ldots\} \tag{1.6}$$

constitutes an *orthonormal basis* for $L_2([0,1])$ and every $f \in L_2([0,1])$ has a unique expansion

$$f = \sum_{\psi \in \Psi} \langle f, \psi\rangle \psi, \qquad \|f\|^2_{L_2([0,1])} = \sum_{\psi \in \Psi} |\langle f, \psi\rangle|^2. \tag{1.7}$$

The equivalence between *continuous* and *discrete* norms will frequently play a pivotal role in subsequent discussions.

As a first instance, relation (1.7) suggests the following simple strategy for approximating a given f by a piecewise constant with *possibly few* pieces. Suppose that all the wavelet coefficients $\langle f, \psi\rangle$, $\psi \in \Psi$, were known, and that the set $\Lambda \subset \Psi$, such that $\#\Lambda \leq N$, contains the N *largest terms* $|\langle f, \psi\rangle|$, then the function $P_\Lambda f = \sum_{\psi \in \Lambda} \langle f, \psi\rangle \psi$ would, on account of (1.7), minimize the error among all piecewise constants on dyadic partitions with at most N pieces.

The selection of the N biggest terms is of course a nonlinear process. This aspect has been thoroughly discussed, for instance by DeVore and Lucier (1992) and DeVore, Jawerth and Popov (1992), and will be taken up in more detail again later in connection with adaptive methods. Here we add only a few comments, which are similar in spirit. Suppose that $g_J \in S_J$ is

some approximation of g. It therefore has a representation

$$g_J = \sum_{k=0}^{2^J-1} c_k \phi_{J,k}$$

in terms of the *single scale* basis functions $\phi_{J,k}$ on the highest scale J. Note that although g_J may have a very simple structure, such as a constant throughout a large part or even all of $[0,1]$, all the 2^J coefficients could be *significant* in that they are needed to preserve accuracy. On the other hand, g_J has a wavelet or *multiscale* representation

$$g_J = \langle g, \phi_{0,0}\rangle \phi_{0,0} + \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} d_{j,k}\, \psi_{j,k}.$$

If g_J were a constant, *all* the $d_{j,k}$ would vanish.

In general, one expects the $d_{j,k}$ to be very small where g_J does not vary much. In fact, if f were differentiable on the support of $\psi_{j,k}$, then, since

$$\langle \phi_{0,0}, \psi_{j,k}\rangle = \int_0^1 \psi_{j,k}(x)\, \mathrm{d}x = 0, \quad k = 0, \ldots, 2^j - 1, \quad j \geq 0, \tag{1.8}$$

one obtains

$$\begin{aligned} |\langle f, \psi_{j,k}\rangle| &= \inf_{c\in\mathbb{R}} |\langle f - c, \psi_{j,k}\rangle| \leq \inf_{c\in\mathbb{R}} \|f - c\|_{L_2([2^{-j}k, 2^{-j}(k+1)])} \\ &\leq 2^{-j} \|f'\|_{L_2([2^{-j}k, 2^{-j}(k+1)])}. \end{aligned} \tag{1.9}$$

Thus, discarding wavelet coefficients that stay below a given threshold may compress the representation significantly, while the accuracy is, in view of (1.7), still controllable. The key is (1.8), which is often referred to as *moment conditions*. Obviously, the vanishing of moments of even higher polynomial order would increase the compression effect.

Of course, to exploit these facts practically requires switching back and forth between single- and multiscale representation. This issue will be addressed later in more generality.

1.3. The Hilbert transform

The compression of functions has a counterpart for operators. The fact that differential operators admit sparse representations is not surprising. Instead, consider the Hilbert transform

$$(\mathcal{H}f)(x) := \frac{1}{\pi} \mathrm{p.v.} \int_{\mathbb{R}} \frac{f(y)}{x - y}\, \mathrm{d}y \tag{1.10}$$

as an example of a typical singular integral operator. Here *p.v.* means 'principal value', that is, $\text{p.v.}\int_{\mathbb{R}} f(x)\,dx = \lim_{\epsilon\to 0^+}\int_{\mathbb{R}\setminus(-\epsilon,\epsilon)} f(x)\,dx$. Its representation relative to the Haar basis requires the entries $H_{(j,k),(l,m)} := \langle \mathcal{H}\psi_{l,m}, \psi_{j,k}\rangle$. Suppose now that $2^{-l}(m+1) < 2^{-j}k$, and $l > j$, that is, the supports of $\psi_{l,m}$ and $\psi_{j,k}$ are disjoint. Then, by (1.8) and Taylor's expansion around $y = 2^{-l}m$, one obtains

$$\begin{aligned}
&\pi\left|H_{(j,k),(l,m)}\right| \\
&\quad= \left|\int_{2^{-j}k}^{2^{-j}(k+1)}\left\{\int_{2^{-l}m}^{2^{-l}(m+1)}\left(\frac{1}{x-y}-\frac{1}{x-2^{-l}m}\right)\psi_{l,m}(y)\,dy\right\}\psi_{j,k}(x)\,dx\right| \\
&\quad= \left|\int_{2^{-l}m}^{2^{-l}(m+1)}\left\{\int_{2^{-j}k}^{2^{-j}(k+1)}\frac{(y-2^{-l}m)}{(x-y_{l,m})^2}\psi_{j,k}(x)\,dx\right\}\psi_{l,m}(y)\,dy\right|,
\end{aligned}$$

for some $y_{l,m}$ in the support $[2^{-l}m, 2^{-l}(m+1)]$ of $\psi_{l,m}$. Repeating the same argument, one can subtract a constant in x which yields

$$\pi\left|H_{(j,k),(l,m)}\right| = \left|\int_{2^{-l}m}^{2^{-l}(m+1)}\left\{\int_{2^{-j}k}^{2^{-j}(k+1)}\left(\frac{(y-2^{-l}m)}{(x-y_{l,m})^2}-\frac{(y-2^{-l}m)}{(2^{-j}k-y_{l,m})^2}\right)\psi_{j,k}(x)\,dx\right\}\psi_{l,m}(y)\,dy\right|.$$

On account of Taylor's expansion around $x = 2^{-j}k$, the factor in front of $\psi_{j,k}(x)$ can be written as $-2(y-2^{-l}m)(x-2^{-j}k)/(x_{j,k}-y_{l,m})^3$, where $x_{j,k}$ is some point in the support $[2^{-j}k, 2^{-j}(k+1)]$ of the wavelet $\psi_{j,k}$. Noting that

$$\int_{\mathbb{R}} |\psi_{j,k}(x)|\,dx \le 2^{-j/2},$$

a straightforward estimate provides

$$\pi|H_{(j,k),(l,m)}| \le 2^{-(n+j)\frac{3}{2}}|2^{-j}k - 2^{-l}m|^{-3} = \frac{2^{-|j-l|\frac{3}{2}}}{|k-2^{j-l}m|^3}. \tag{1.11}$$

Thus the entries $H_{(j,k),(m,l)}$ exhibit a decay with increasing distance of the supports of the wavelets as well as with increasing distance of scales. In essence this behaviour persists for a large class of integral operators and is the key to *sparsify* the discretization of such operators.

1.4. A two-point boundary value problem

Consider

$$-u'' = f \quad \text{on } [0,1], \quad u(0) = u(1) = 0, \tag{1.12}$$

as a simple model for an elliptic second-order boundary value problem. Although there are, of course, much simpler ways of solving (1.12), we start from the standard *weak formulation*

$$\langle u', v' \rangle = \langle f, v \rangle, \quad v \in H_0^1([0,1]). \tag{1.13}$$

Here $H_0^1([0,1])$ is the closure of all C^∞ functions with compact support in $(0,1)$ relative to the norm $\|f\|_{H^1([0,1])} = (\|f\|^2_{L_2([0,1])} + \|f'\|^2_{L_2([0,1])})^{1/2}$. To make our point, we use a standard Galerkin approach and solve (1.13) on finite-dimensional spaces $S_j \subset H_0^1([0,1])$. The simplest conforming choice of the trial spaces S_j are the spans of scaled *tent functions*

$$\phi_{j,k}(x) = 2^{j/2}\phi(2^j \cdot - k), \quad k = 0, \dots, 2^j, \tag{1.14}$$

where

$$\phi(x) = \begin{cases} 1 + x, & -1 \le x \le 0, \\ 1 - x, & 0 \le x \le 1, \\ 0, & \text{otherwise.} \end{cases} \tag{1.15}$$

Choosing the $\phi_{j,k}$ as basis functions for S_J, the Galerkin conditions

$$\langle u_j, v \rangle = \langle f, v \rangle, \quad v \in S_j, \tag{1.16}$$

give rise to a linear system of equations

$$\mathbf{A}_J \mathbf{u} = \mathbf{f}, \tag{1.17}$$

where $\mathbf{A}_J$ is the stiffness matrix relative to the basis functions $\phi_{J,k}$ and $\mathbf{u}$, $\mathbf{f}$ are corresponding vectors with $f_k = \langle f, \phi_{J,k} \rangle$. Clearly $\mathbf{A}_J$ is tridiagonal. Hence (1.17) is very efficiently solvable. However, for higher-dimensional analogues the matrix would no longer have such a narrow bandwidth and one has to resort to *iterative methods* to preserve sparseness.

On the other hand, recalling the min-max characterization of the smallest and largest eigenvalue of a symmetric positive definite matrix, it is easy to see that the condition numbers of $\mathbf{A}_J$ grow like 2^{2J}, which renders classical iterative methods prohibitively inefficient.

To remedy this, one has to *precondition* the linear systems. One way is to exploit suitable multiscale decompositions of the trial spaces S_J. First note that, since

$$\phi(x) = \frac{1}{2}\phi(2x+1) + \phi(2x) + \frac{1}{2}\phi(2x-1), \tag{1.18}$$

that is,

$$\phi_{j,k} = \frac{1}{2\sqrt{2}}\phi_{j+1,2k-1} + \frac{1}{\sqrt{2}}\phi_{j+1,2k} + \frac{1}{2\sqrt{2}}\phi_{j+1,2k+1}, \tag{1.19}$$

the S_j are nested and, of course, their union is dense in $L_2([0,1])$.

In order to successively update solutions from coarser grids, we consider the following *hierarchical decomposition* of the trial spaces (Yserentant 1986). Instead of using orthogonal projections as in Section 1.2, we consider the *Lagrange projectors*

$$L_j f := \sum_{k=0}^{2^j} 2^{-j/2} f(2^{-j}k)\phi_{j,k}, \tag{1.20}$$

and note that the complements

$$W_j := (L_{j+1} - L_j)S_{j+1} \tag{1.21}$$

are simply spanned by the tent functions on new grid points on the next higher scale

$$\Psi_j := \{\psi_{j,k} := \phi_{j+1,2k+1} : k = 0, \ldots, 2^j - 1\}. \tag{1.22}$$

Note that neither the $\phi_{j,k}$ nor the $\psi_{j,k}$ are orthogonal but it is not hard to show that they satisfy the *stability condition*

$$c_1 \left(\sum_{k=0}^{2^j} |c_k|^2 \right)^{1/2} \leq \left\| \sum_{k=0}^{2^j} c_k \phi_{j,k} \right\|_{L_2([0,1])} \leq c_2 \left(\sum_{k=0}^{2^j} |c_k|^2 \right)^{1/2} \tag{1.23}$$

for some constants c_1, c_2 independent of the sequence $\{c_k\}_{k=0}^{2^j}$. Keeping this in mind, we now consider stiffness matrices relative to the *hierarchical bases* composed of the bases Ψ_j, and note that

$$\frac{d}{dx}\psi_{j,k}(x) = \frac{d}{dx}\phi_{j+1,2k+1}(x) = 2^{j+\frac{3}{2}}\psi^H_{j,k}(x), \tag{1.24}$$

where $\psi^H_{j,k}$ are the Haar wavelets from (1.3). Therefore one obtains from (1.5)

$$\left\langle \frac{d}{dx}\psi_{j,k}, \frac{d}{dx}\psi_{n,l} \right\rangle = 2^{j+n+3} \left\langle \psi^H_{j,k}, \psi^H_{n,l} \right\rangle = 2^{-2j+3}\delta_{j,n}\delta_{k,l}.$$

Hence $\mathbf{A}_J$ is, up to a 2×2 upper left block stemming from the coarse grid space S_0, a diagonal matrix, which is trivially *preconditioned* by symmetric *diagonal scalings.*

Now, one has to be somewhat careful when extrapolating from this observation. The fact that the hierarchical basis functions $\psi_{j,k}$ are actually orthogonal relative to the *energy inner product* is an artefact. In two dimensions this is no longer the case but it turns out that the hierarchical stiffness matrices can still be preconditioned by diagonal scaling to efficiently reduce

the growth of the condition numbers to logarithmic behaviour. Moreover, it has suggested similar strategies involving other multiscale bases which do better. Corresponding preconditioning techniques are a central theme in subsequent discussions.

1.5. Some basic ideas

Next, I would like to blend these examples into a general picture to provide some orientation and unifying structure for the subsequent discussions of a diversity of ideas.

To this end, suppose that

$$\Psi = \{\psi_\lambda : \lambda \in \nabla\} \tag{1.25}$$

is a countable *basis* of some Hilbert space H. Thus every $v \in H$ has a unique convergent expansion in terms of elements of Ψ

$$v = \sum_{\lambda \in \nabla} d_\lambda \psi_\lambda. \tag{1.26}$$

The dependence of the coefficients $\{d_\lambda\}$ on v can be expressed via the *dual basis*. This is a collection of functionals

$$\tilde{\Psi} = \{\tilde{\psi}_\lambda : \lambda \in \nabla\},$$

such that

$$\langle \psi_\lambda, \tilde{\psi}_{\lambda'} \rangle = \delta_{\lambda,\lambda'}, \quad \lambda, \lambda' \in \nabla, \tag{1.27}$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product on H. When H is infinite-dimensional, the notion of basis has to be further specified, but we will defer this issue for the moment. The collection of Haar functions $\psi^H_{j,k}$ forms such a basis for $H = L_2([0,1])$. In this case the indices $\lambda = (j,k)$ encode the information about scale and location. Of course, in the case of the Haar basis $\Psi = \Psi^H$, one has $\Psi = \tilde{\Psi}$. Equation (1.27) means that the coefficients d_λ in the expansion of v relative to Ψ are given by $d_\lambda = \langle v, \tilde{\psi}_\lambda \rangle$.

To simplify further exposition, I now introduce a compact notation for bases and their transforms that will be consistently used throughout the rest of the paper. Formally, let us view a given (countable) collection of functions Φ in H as a (column) *vector* (of possibly infinite length), so that an expansion with coefficients $c_\phi, \phi \in \Phi$ can be formally treated as an 'inner product'

$$\mathbf{c}^T \Phi := \sum_{\phi \in \Phi} c_\phi \phi.$$

The sum is always understood to converge in the norm of the underlying space, and the superscript T denotes 'transpose'. Likewise, for any $v \in H$, the quantities $\langle \Phi, v \rangle$ and $\langle v, \Phi \rangle$ mean the column-, respectively row-vector,

of coefficients $\langle \phi, v\rangle$, $\langle v, \phi\rangle$, $\phi \in \Phi$. Thus (1.26) can be written for short as $\mathbf{d}^T\Psi$. Boldface lower case or capital letters will always denote sequences or matrices acting on sequences, respectively.

To push this a little further, for any two countable collections Φ, Ξ of functions, we consider the (possibly infinite) matrix

$$\langle \Phi, \Xi\rangle = (\langle \phi, \xi\rangle)_{\phi\in\Phi,\xi\in\Xi}.$$

Specifically, the above biorthogonality relations (1.27) then become

$$\langle \Psi, \tilde{\Psi}\rangle = \mathbf{I}, \tag{1.28}$$

where $\mathbf{I}$ denotes the identity matrix (whose dimension should be clear from the context).

The examples in Sections 1.3 and 1.4 can be viewed as special cases of the following situation. Suppose that H_1 and H_2 are *Hilbert* spaces such that either one of the continuous embeddings

$$H_1 \subseteq H \subseteq H_2 \quad \text{or} \quad H_2 \subseteq H \subseteq H_1$$

holds. In many cases of interest, H_2 is the *dual* of H_1, that is, the space of bounded linear functionals on H_1 relative to the duality pairing induced by the inner product $\langle\cdot,\cdot\rangle$ on the (intermediate) space H. Furthermore, suppose that $\mathcal{L}$ is a bounded linear bijection that maps H_1 onto H_2, that is,

$$\|\mathcal{L}v\|_{H_2} \sim \|v\|_{H_1}, \qquad v \in H_1, \tag{1.29}$$

where here and below $a \sim b$ means $a \lesssim b$ and $b \lesssim a$. The latter relations express that b can be bounded by some constant times a uniformly in any parameters on which a and b may depend. Hence the equation

$$\mathcal{L}u = f \tag{1.30}$$

has a unique solution $u \in H_1$ for every $f \in H_2$. In Section 1.3 we had $\mathcal{L} = H$, $H_1 = H_2 = L_2(\mathbb{R})$, while in Section 1.4 $\mathcal{L} = -\frac{d^2}{dx^2}$, $H_1 = H_0^1([0,1])$, $H_2 = H^{-1}([0,1])$, the dual of $H_0^1([0,1])$.

The basic idea is to transform the (continuous) equation (1.30) into an infinite *discrete* system of equations. This can be done with the aid of suitable *bases* for the spaces under consideration.

Given such bases, seeking the solution u of (1.29) is equivalent to finding the expansion sequence $\mathbf{d}$ of $u = \mathbf{d}^T\Psi$. Inserting this into (1.30) yields $(\mathcal{L}\Psi)^T\mathbf{d} = f$. Now suppose that $\Theta = \{\theta_\lambda : \lambda \in \nabla\}$ is *total* over H_2, that is, $\langle v, \Theta\rangle = 0$ implies $v = 0$ for $v \in H_2$. Then $(\mathcal{L}\Psi)^T\mathbf{d} = f$ becomes the (infinite) system

$$\langle \mathcal{L}\Psi, \Theta\rangle^{\mathrm{T}}\mathbf{d} = \langle f, \Theta\rangle^{\mathrm{T}}. \tag{1.31}$$

The objective now is to find collections Ψ and Θ for which the system (1.31) is efficiently solvable. This can be approached from several different angles.

(a) *Diagonalization*
The ideal case would be to know a complete system Ψ of *eigenfunctions* so that the choice $\Theta = \Psi$ would diagonalize (1.31). Of course, in practice this is usually not feasible. However, when Ψ and $\tilde{\Psi}$ are regular enough in the sense that the collections

$$\Theta := (\mathcal{L}^{-1})^* \tilde{\Psi} \subset H_1, \quad \tilde{\Theta} := \mathcal{L}\Psi \subset H_2, \tag{1.32}$$

then *biorthogonality* (1.28) implies

$$\langle \Theta, \tilde{\Theta} \rangle = \mathbf{I}, \tag{1.33}$$

that is, biorthogonality of the pair $\Theta, \tilde{\Theta}$. Here $\mathcal{L}^*$ denotes the *dual* or *adjoint* of $\mathcal{L}$ defined by $\langle \mathcal{L}u, v \rangle = \langle u, \mathcal{L}^* v \rangle$. In this case the solution $u = \mathbf{d}^T \Psi$ is given by

$$\mathbf{d} = \langle \Theta, f \rangle, \quad u = \langle f, \Theta \rangle \Psi. \tag{1.34}$$

When Ψ is a wavelet basis, it will be seen that under certain assumptions on $\mathcal{L}$ (defined on $\mathbb{R}^n$ or the torus), the elements of Θ share several properties with wavelets. The θ_λ are then called *vaguelettes*. Truncation of $\langle f, \Theta \rangle \Psi$ would readily yield an approximation to u. Note that this can be viewed as a *Petrov–Galerkin* scheme.

(b) *Preconditioning*
One expects that vaguelettes are numerically accessible only in special cases such as for constant coefficient differential elliptic operators on $\mathbb{R}^n$ or the torus. However, these cases may be in some sense *close* to more realistic cases, which opens possibilities for preconditioning.

Alternatively, one could relax the requirements on the bases Ψ and Θ. Again one could view the eigensystem as the ideal choice. A simple diagonal scaling would then transform $\langle \mathcal{L}\Psi, \Psi \rangle^T$ into $\mathbf{I}$. Thus one could ask for bases Ψ such that for a suitable *diagonal* matrix $\mathbf{D}$,

$$\mathbf{B} := \mathbf{D} \langle \mathcal{L}\Psi, \Psi \rangle^T \mathbf{D} \asymp \mathbf{I} \tag{1.35}$$

is spectrally equivalent to the identity, in the sense that $\mathbf{B}$ and its inverse $\mathbf{B}^{-1}$ are bounded in the Euclidean norm $\|\mathbf{d}\|^2_{\ell_2(\nabla)} := \mathbf{d}^* \mathbf{d}$, where $\mathbf{d}^* := \overline{\mathbf{d}}^T$ is the usual complex conjugate transpose.

Note that the principal sections of the infinite matrix $\mathbf{B}$ correspond to the *stiffness matrices* arising from a Galerkin scheme applied to (1.30) based on trial spaces spanned by subsets of Ψ. Relation (1.35) means that these linear systems are uniformly well conditioned. Such a Ψ would be in some sense sufficiently close to the eigensystem of $\mathcal{L}$. It will be seen that for a wide class of operators wavelet bases have that property. The precise

choice of $\mathbf{D}$ depends on $\mathcal{L}$ or, more precisely, on the pair of spaces H_1, H_2 in (1.29). For instance, in Section 1.4 the diagonal entries of $\mathbf{D}$ would be 2^j for $\lambda := (j,k)$. In this context Sobolev spaces play a central role and the question of preconditioning will be seen to be intimately connected with the characterization of Sobolev spaces in terms of certain *discrete norms* induced by wavelet expansions.

(c) *Sparse representations*
The similarity between wavelet bases and eigensystems extends beyond the preconditioning effect. Indeed, for many operators the matrices $\mathbf{B}$ in (1.35), as well as their inverses, are *nearly sparse*. This means that replacing entries below a given threshold by zero yields a sparse matrix. When $\mathcal{L}$ is a differential operator and the wavelets have compact support this may not be too surprising (although the mixing of different levels creates, in general, a less sparse structure than shape functions with small support on the highest discretization level). However, it even remains true for certain integral operators as indicated by the estimate (1.11) for $\langle \mathcal{H}\Psi^H, \Psi^H \rangle$. Quantifying this sparsification will depend on $\mathcal{L}$ and on certain properties of the wavelet basis that will have to be clarified.

(d) *Significant coefficients and adaptivity*
Once you can track the wavelets in Ψ needed to represent the solution u of (1.30) accurately, one can, in principle, restrict the computations to the corresponding subspaces. Combining this with the sparse representation of operators is perhaps one of the most promising perspectives of wavelet concepts. A significant part of subsequent discussions will be initiated by this issue.

1.6. The structure of the paper

Here is a short overview of the material and the way it is organized. **Section 2** outlines the scope of problems to be treated and indicates corresponding basic obstructions to an efficient numerical solution. It is clear from the preceding discussion that, for each problem, the properties of underlying function spaces, in particular *Sobolev spaces*, have to be taken into account. A few preliminaries of this sort will therefore be collected first.

The objective of this paper is by no means the construction of wavelets. However, the performance of a wavelet scheme relies on very specific properties of the wavelet basis. I find it unsatisfactory to simply assume these properties without indicating to what extent and at what cost these properties may actually be realized. Therefore the construction of the tools also provides the necessary understanding for its limitations. Consequently some space has to be reserved for discussing properties of multiscale bases. Guided by the examples in Sections 1.2 and 1.4, **Section 3** begins by describing a general framework of multiresolution decompositions: this is to provide a

uniform platform for all the subsequent specifications, in particular, those which involve more complex domain geometries. The simple but useful concepts of *stable completions* is emphasized as a construction device that can still be used under circumstances where, for instance, classical arguments based on Fourier techniques no longer apply.

Section 4 outlines some examples of multiscale decompositions and wavelet bases, which will later be referred to frequently. So-called *hierarchical bases* on bivariate triangulations (as a straightforward generalization of the construction in Section 1.4) will serve later as a bridge to developments in the finite element context. Wavelets defined on all of Euclidean space, their periodized versions, and wavelets on cubes are by far best understood. A few facts are recorded here which are important for further extensions needed later on.

Multiresolution originates from the classical setting concerning the full Euclidean space. The shift- and scale-invariance of its ingredients provide a comfortable basis for constructions and admit in combination with Fourier techniques best computational efficiency. While wavelets are usually associated with *orthogonal bases*, the concept of biorthogonal wavelets is emphasized, because it offers much more flexibility and localization (in physical space). I will try to indicate later that this actually pays dividends in several applications.

Much of the comfort of shift- and dilation-invariance can still be retained when dealing with wavelets on the *interval* (and hence on cubes). This still looks very restrictive, but it will turn out later to be an important ingredient for extending the application of wavelet schemes, for instance to closed surfaces or other manifolds. I have collected these construction issues in one section, so that those who are familiar with this material can easily skip over this section.

Section 5 addresses the heart of the matter. Once one is willing to dispense with orthogonality, one has to understand which type of decompositions are actually suitable. A classical theme in functional analysis is to characterize function spaces through isomorphisms into sequence spaces. The discussion in Section 1.5 has already stressed this point as a basic vehicle for developing *discretizations*. Orthonormal bases naturally induce such isomorphisms. When deviating from orthogonality, the leeway is easily seen to be set by the concept of *Riesz bases*, which in turn brings in the notion of *biorthogonal* bases. Whereas biorthogonality is necessary, it is not quite sufficient for establishing the desired norm equivalences. The objective of Section 5 is to bring out what is needed in addition. In order to be able to apply this to several cases, this is formulated for a general Hilbert space setting (see Section 5.2). It should be stressed that the additional stability criteria concern properties of the underlying multiresolution spaces *not* of the particular bases. Therefore things are kept in a basis-free form. Again,

to guarantee flexible applicability, these criteria do *not* resort to Fourier techniques but are based on a pair of inequalities, describing *regularity* and *approximation* properties of the underlying multiresolution spaces (see Section 5.1). The most important application for the present purpose is the characterization of Sobolev spaces in Section 5.3 for all relevant versions of underlying domains, including manifolds such as closed surfaces. These facts will later play a crucial role in three different contexts, namely *preconditioning* (recall Section 1.5, (b)), thresholding strategies for *matrix compression* (see Section 1.5, (c)) and the analysis of *adaptive* schemes (see Section 1.5, (d)).

A first major application of the results in Section 5 is presented in **Section 6**. It is shown that (1.35), namely the transformation of a *continuous* problem into a *discrete* one, which is well-posed in the Euclidean metric, is realized for a wide class of elliptic *differential* and *integral* operators, described in Section 2.3. The entries of the diagonal matrix $\mathbf{D}$ depend on the *order* of the operator $\mathcal{L}$. Preconditioning is seen here to be an immediate consequence of the validity of norm equivalences for Sobolev spaces. It simply means that the shift in Sobolev scale caused by the operator $\mathcal{L}$ in (1.30) can be undone by a proper weighting of wavelet coefficients. Diagonal matrices act in some sense like differential or integral operators much like classical Bessel potentials.

To bring out the essential mechanism, this is formulated for a possibly abstract setting. One should look at the examples in Section 2.2 to see what it means in concrete cases. On the other hand, it is important to note that the full strength of wavelet bases is actually *not* always needed. When the order of the operator $\mathcal{L}$ is *positive*, the weaker concept of *frame* suffices. This establishes a strong link to recent, essentially independent, developments of *multilevel preconditioning techniques* in a finite element context. Both lines of development have largely ignored each other. Although the present discussion is primarily seen from the viewpoint of wavelet analysis, I will briefly discuss both schools and their interrelation.

While the concepts in Section 6 can also be realized in a finite element setting, **Section 7** confines the discussion to what will be called the *ideal setting*, meaning problems formulated on $\mathbb{R}^n$ or the torus. As detailed in Section 4.2, an extensive machinery of wavelet tools is available and much more refined properties can be exploited. Last but not least, through marriage with Fourier techniques such as FFT, this could be a tremendous support of computational efficiency. Some of the insight into local phenomena gained in this way can also be expected to help under more general circumstances. Of course, all these properties are preserved under periodization, so that wavelets still unfold their full potential for *periodic* problems.

One further reason for reserving some room for this admittedly restricted setting is to think of a *two-step approach*. Exploiting all the benefits of the

ideal setting, one aims at developing highly efficient techniques that are to cope with the bulk of computation determined by, say, the spatial dimension of the domain. This should then justify efforts to treat geometric constraints entering through boundary conditions separately, hopefully at the expense of lower-order complexity.

The common ground for all the techniques mentioned in Section 7 is that the *inverse* of an elliptic operator is fairly well accessible in wavelet coordinates. This concerns *vaguelette* techniques, which in the spirit of Section 1.5 (a), aim at *diagonalizing* the operator $\mathcal{L}$ (see Section 7.2). Several issues such as the (adaptive) *evaluation* of vaguelette coefficients, *freezing coefficient* techniques for operators with variable coefficients and relaxed notions like *energy-pre-wavelets* are discussed.

The next step is to consider a class of univariate (periodic) nonlinear evolution model equations (Section 7.7). Several different approaches such as vaguelette schemes and *best bases* methods will be discussed. The so-called *pseudo-wavelet approach* aims at a systematic development of techniques for an adaptively controlled accurate application of evolution operators and nonlinear terms. An important vehicle in this context is the so-called *non-standard form* of operators. I will try to point out the difference between several approaches which are based on a number of very interesting and fairly unconventional concepts.

These evolution equations are to be viewed as simplified models of more complex systems like the *Stokes* and *Navier–Stokes* equations. In Section 7.11 some ways of dealing with corresponding additional difficulties are discussed. It seems that biorthogonal vaguelette versions combined with (biorthogonal) compactly supported divergence-free wavelet bases offer an interesting option, which has not been explored yet.

As mentioned before, a major motivation for the developments in Section 7 was to embed problems defined on more general domains into the ideal setting and then treat boundary conditions separately. **Section 8** is devoted to a brief discussion of several such embedding strategies. I will focus on three options. The first is to use extension techniques in conjunction with the multilevel Schwarz schemes described in Section 6.5. This is particularly tailored to variational formulations of problems involving self-adjoint operators. An alternative is to correct boundary values by solving a *boundary integral* equation. Finally, one can append boundary conditions with the aid of *Lagrange multipliers.*

Section 9 deals with pseudo-differential and integral operators. As an important case, this covers *boundary integral equations.* This type of problem is interesting for several reasons. First of all, it naturally came up in Section 8 in connection with partial differential equations. Second, it poses several challenges. On one hand, boundary integral formulations frequently offer physically more adequate formulations and reduce, in the case of exter-

ior boundary value problems, the discretization of an unbounded domain to a discretization of a lower-dimensional compact domain. On the other hand, they have the serious drawback that (in some cases in addition to preconditioning issues) the resulting matrices are *dense*. However, as indicated by the example in Section 1.3, such operators have a *nearly* sparse representation relative to *appropriately* chosen wavelet bases. What exactly 'appropriate' means, and some ingredients of a rigorous analysis of corresponding *compression* techniques will be explained in this section. The issue here is twofold, namely reducing a matrix to a sparse matrix without losing *asymptotical* accuracy of the solution, and the efficient computation of the compressed matrices at costs that remain proportional to their size. Moreover, when the operator has *negative* order (see Section 2.2) preconditioning does require the full strength of wavelet decompositions. So, in principle, wavelets seem to be particularly promising for this type of problem. One expects that they offer a common platform for (i) efficiently applying operators that are otherwise dense, (ii) preconditioning the linear systems and (iii) facilitating adaptive strategies for further reducing complexity.

However, the embedding strategies from Section 8 do not apply to closed surfaces. So appropriate notions of wavelets on manifolds have to be developed. Discontinuous *multi-wavelets* have been employed so far. But according to the results in Section 6, they are not optimal for preconditioning operators of order -1. Therefore **Section 10** is devoted to the construction of wavelet bases on manifolds that have all the properties required by the analysis in Section 9. This rests on two pillars: the characterization of Sobolev spaces with respect to a *partition* of the manifold into parametric images of the unit cube (recall that the classical definition of Sobolev spaces on manifolds is based on open coverings), and certain biorthogonal wavelet bases on the unit cube that satisfy special boundary conditions. The construction of such bases, in turn, can be based on the ingredients presented in Section 4.4. This refers partly to work in progress. Some consequences with regard to domain decomposition are briefly indicated.

In **Section 11** we take up again the issue of adaptivity. The main objective is to outline a rigorous analysis for a possibly general setting that covers the previously discussed special cases. Some comments about relating this to adaptive strategies in the finite element context are included. In addition, this part should complement the intriguing algorithmic concepts discussed before. The section concludes with a brief discussion of the relation between the efficiency of adaptive approximation and *Besov regularity* of the solutions of elliptic equations.

Finally, in **Section 12** some further interesting directions of current and perhaps future research are indicated.

2. The scope of problems

The objective of this section is to put some meat on the skeleton of ideas in Section 1.5 by identifying first a list of concrete model problems satisfying (1.29) and (1.30). This requires some preparation.

2.1. Function spaces and other preliminaries

It is clear from the discussion in Section 1.5 that certain functional analytic concepts related to Sobolev spaces play an important role. This section contains corresponding relevant definitions, notation and conventions.

For any normed linear space S the norm is always denoted by $\|\cdot\|_S$. The *adjoint* or *dual* of an operator $\mathcal{L}$ is denoted by $\mathcal{L}^*$.

Important examples are L_p spaces. For $1 \le p \le \infty$ (with the usual sup-norm interpretation for $p = \infty$) and for any measure space $(\Omega, d\mu)$, the space $L_p(\Omega)$ consists of those measurable functions v such that

$$\|v\|_{L_p(\Omega)} := \left(\int_\Omega |v(x)|^p \, \mathrm{d}\mu(x) \right)^{1/p} < \infty.$$

For simplicity, we usually write $\mathrm{d}x$ instead of $\mathrm{d}\mu(x)$, since only the Lebesgue measure will matter. The case $p = 2$ is used most often. In this case $\|\cdot\|^2_{L_2(\Omega)} = \langle \cdot, \cdot \rangle_\Omega$, where

$$\langle u, v \rangle_\Omega := \int_\Omega u(x)\overline{v(x)} \, \mathrm{d}x$$

denotes the corresponding standard inner product. Here Ω may be $\mathbb{R}^n$ or a domain in $\mathbb{R}^n$ or, more generally, a manifold such as a closed surface. The latter interpretation is needed when dealing with boundary integral equations.

Partial derivatives are denoted by ∂, or ∂_x if it is stressed with respect to which variable it applies. Common multi-index notation is used, that is, $x^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$, $|\alpha| = |\alpha_1| + \cdots + |\alpha_n|$, for $\alpha \in \mathbb{N}_0$, $\mathbb{N}_0 := \{0, 1, 2, 3, \ldots\}$.

Suppose now that $\Omega \subset \mathbb{R}^n$ is a domain. We shall always assume that Ω is a bounded, open, and connected (at least) Lipschitz domain. This covers all cases of practical interest. If m is a positive integer, the *Sobolev* space $W^{m,p}((\Omega))$ consists of all functions $f \in L_p(\Omega)$, whose distributional derivatives $\partial^\nu f$, $|\nu| = m$, satisfy

$$|f|^p_{W^{m,p}(\Omega)} := \sum_{|\nu|=k} \|\partial^\nu f\|^p_{L_p(\Omega)} < \infty; \tag{2.1}$$

see, for example, Adams (1978). The pth root of (2.1) is the semi-norm for $W^{m,p}(\Omega)$, and adding to it $\|f\|_{L_p(\Omega)}$ gives the norm $\|f\|_{W^{m,p}(\Omega)}$ in $W^{m,p}(\Omega)$.

For the present purposes the most important case is again $p = 2$, which is denoted for short as $H^m(\Omega) := W^{m,2}(\Omega)$. Furthermore, Sobolev spaces with noninteger index $s \in \mathbb{R}$ are needed. There are several ways to define them. For $\Omega = \mathbb{R}^n$ one can use Fourier transforms

$$\hat{f}(y) := \int_{\mathbb{R}^n} f(x) e^{-ix\cdot y}\, dy,$$

and set

$$H^s(\mathbb{R}^n) = \{f \in L_2(\mathbb{R}^n) : \int_{\mathbb{R}^n} (1 + |y|^2)^s |\hat{f}(y)|^2\, dy < \infty\},$$

where $|\cdot|$ is the Euclidean norm on $\mathbb{R}^n$. When $\Omega \neq \mathbb{R}^n$, the Lipschitz property implies that there exist extension operators E that are bounded in H^m for any $m \in \mathbb{N}$. For $s > 0$ one can define $\|f\|_{H^s(\Omega)} := \inf\{\|g\|_{H^s(\mathbb{R}^n)} : g \mid_\Omega = f\}$. Alternatively, $H^s(\Omega)$ can be defined by interpolation between $L_2(\Omega)$ and $H^m(\Omega)$, $m > s$; see Bergh and Löfström (1976), DeVore and Popov (1988*a*) and Triebel (1978). When $s < 0$ one can use duality. For any normed linear space V, the dual space, consisting of all bounded linear functionals on V, is denoted by V^*. It is a Banach space under the norm $\|w\|_{V^*} := \sup_{\|v\|_V = 1} |w(v)|$. Specifically, when Ω is a closed manifold $(H^s(\Omega))^* = H^{-s}(\Omega)$.

We will briefly encounter *Besov* spaces $B^s_q(L_p(\Omega))$; see again Bergh and Löfström (1976), DeVore and Popov (1988*a*), DeVore and Sharpley (1993) and Triebel (1978). They arise by interpolation between $L_p(\Omega)$ and $W^{m,p}(\Omega)$. Recall that $H^s(\Omega) = B^s_2(L_2(\Omega))$.

As mentioned before, lower case boldface letters such as $\mathbf{c}, \mathbf{d}$ will always denote sequences over some (finite or infinite) index set Δ. As usual, for the same range of p as above, we set

$$\|\mathbf{c}\|_{\ell_p(\Delta)} := \left(\sum_{k\in\Delta} |c_k|^p\right)^{1/p}.$$

By convention, the elements of $\ell_p(\Delta)$ will always be viewed as *column* vectors, that is, $\mathbf{c}^T, \mathbf{c}^*$ are *rows*, the latter indicating complex conjugates when using the complex field. Analogously, for a matrix $\mathbf{M}$ the transpose is $\mathbf{M}^T$, while $\mathbf{M}^*$ denotes its complex conjugate transpose.

When there is no risk of confusion the reference to the domain or index set will sometimes be dropped, that is, we write $\langle\cdot,\cdot\rangle$, H^s, ℓ_2, *etc.*

2.2. A general class of elliptic problems

(a) *Scalar elliptic boundary value problems*
For $\Omega \subseteq \mathbb{R}^n$, an example of $\mathcal{L}$ in (1.30) is

$$\mathcal{L}u := i^{2m} \sum_{|\alpha|,|\beta|\leq m} a_{\alpha,\beta}(x)\partial^\alpha \partial^\beta u = f \quad \text{on } \Omega, \quad \mathcal{B}u = 0 \quad \text{on } \partial\Omega, \tag{2.2}$$

where $\mathcal{B}$ is a suitable trace operator, and the polynomial

$$P(\xi) := \sum_{|\alpha|,|\beta|=m} a_{\alpha,\beta}(x)\xi^{\alpha+\beta}$$

satisfies

$$P(\xi) \geq \delta > 0, \quad \xi \in \mathbb{R}^n, \ x \in \Omega. \tag{2.3}$$

Depending on the regularity of the domain, (1.29) holds with $H_1 = H^s(\Omega)$, $H_2 = H^{s-2m}(\Omega)$ for a certain range of s. An important special case is

$$-\operatorname{div}(A(x)\nabla u) + a(x)u = f \text{ on } \Omega, \quad u = 0 \text{ on } \partial\Omega, \tag{2.4}$$

where $A(x)$ is uniformly positive definite and symmetric on Ω and for a vector field v the divergence operator is defined by $\operatorname{div} v := \sum_{j=1}^n \frac{\partial}{\partial x_j} v_j$. Clearly $A = I$, $a(x) = 0$, gives Poisson's equation with Dirichlet boundary conditions. Here $H_1 = H_0^1(\Omega)$ and $H_2 = H^{-1}(\Omega) = (H_0^1(\Omega))^*$. Likewise, one could take the Helmholtz equation $\mathcal{L} = -\Delta + aI$ for $a > 0$, or $\mathcal{L} = -\Delta + \beta \cdot \nabla$. Similarly with $\mathcal{L} = \Delta^2$, $H_1 = H_0^2(\Omega)$, $H_2 = H^{-2}(\Omega)$, fourth-order problems are covered as well.

The special case that $\mathcal{L}$ is positive definite and selfadjoint is of particular interest, that is,

$$a(u, v) = \langle \mathcal{L}u, v \rangle \tag{2.5}$$

is a *symmetric bilinear* form. *Ellipticity* here means that

$$a(\cdot, \cdot) \sim \| \cdot \|_{H_1}^2, \tag{2.6}$$

which implies (1.29). Clearly (2.4) falls into this category.

Such problems can be solved approximately with the aid of finite element-based Galerkin schemes. There are several different problems that arise. For $n \geq 2$ one obtains *large* linear systems, usually with *sparse* matrices which, for instance in the case (2.4), are symmetric positive definite. Thus a major challenge lies simply in the size of such problems. Since direct solvers based on matrix factorizations would cause a significant *fill-in* of nonzero entries in the factors, and therefore prohibitively limit storage and computing time, one has to resort to iterative solvers for large problem sizes. Unfortunately, the *condition numbers* of the system matrices grow with their size N like $N^{2m/n}$. It is therefore of vital importance to *precondition* these sytems.

In fact, an asymptotically optimal scheme would require *uniformly bounded* condition numbers.

When $\mathcal{L}$ is not selfadjoint, efficient schemes such as preconditioned conjugate gradient (PCG) iterations have to be replaced by more expensive ones, whose performance is no longer a simple function of the spectral condition numbers.

Finally, the coefficients in $A(x)$ or $a(x)$ may vary rapidly. On one hand, this may adversely affect the constants in (1.29). On the other hand, the resolution of such fluctuations may require too small mesh sizes, so that questions of *homogenization* arise. In the following we will primarily address the first two issues.

(b) *Saddle point equations*

An important example for a *system* of partial differential equations is the *Stokes* problem

$$\begin{aligned} -\Delta u + \nabla p &= f, \quad \text{on } \Omega, \quad u = 0 \ \text{ on } \ \partial\Omega, \\ \operatorname{div} u &= 0, \end{aligned} \tag{2.7}$$

as a simple model for viscous incompressible flow. The vector valued function u and the scalar field p represent velocity and pressure of the fluid, respectively. Obviously, one has to factor the constants from p, for instance by requiring $\int_\Omega p(x)\,dx = 0$.

The *weak formulation* of (2.7) requires finding $(u,p) \in V \times M$, where

$$V := \left(H_0^1(\Omega)\right)^n, \quad M = L_{2,0}(\Omega) = \left\{ f \in L_2(\Omega) : \int_\Omega f(x)\,dx = 0 \right\}, \tag{2.8}$$

such that

$$\begin{aligned} a(u,v) + b(v,p) &= \langle f, v\rangle_\Omega, \quad v \in V \\ b(u,\mu) &= 0, \qquad \mu \in M, \end{aligned} \tag{2.9}$$

with

$$a(u,v) = \langle \nabla u, \nabla v\rangle_\Omega, \quad b(v,\mu) = \langle \operatorname{div} v, \mu\rangle_\Omega. \tag{2.10}$$

So-called *mixed formulations* of (2.4) for $a(x) = 0$ arise when introducing the *flux* $\sigma := -A\nabla u$ as a new variable, so that $-\operatorname{div}(A\nabla u) = f$ yields a coupled system of first-order equations

$$A\nabla u = -\sigma, \quad \operatorname{div} \sigma = f,$$

whose weak formulation is

$$\begin{aligned} a(\sigma,\tau) - b(\tau,u) &= 0, \qquad v \in V \quad := H(\operatorname{div},\Omega), \\ -b(\sigma,v) &= -\langle f, v\rangle_\Omega, \quad v \in M \quad := L_2(\Omega). \end{aligned} \tag{2.11}$$

Here $a(\cdot,\cdot) = \langle \cdot, \cdot \rangle_\Omega$, $b(\cdot,\cdot)$ is defined as before in the Stokes problem, and

$$H(\mathrm{div}, \Omega) := \{\tau \in (L_2(\Omega))^{(n}: \mathrm{div}\, \tau \in L_2(\Omega)\},$$

endowed with the graph norm $\|\tau\|_{H(\mathrm{div},\Omega)} = (\|\tau\|^2_{L_2(\Omega)} + \|\mathrm{div}\, \tau\|^2_{L_2(\Omega)})^{1/2}$.

Both cases (2.9) and (2.11) can be viewed as an operator equation of the form (1.30) with

$$\mathcal{L} = \begin{pmatrix} \mathcal{A} & \mathcal{B}^* \\ \mathcal{B} & 0 \end{pmatrix}, \tag{2.12}$$

and $\mathcal{A} : V \to V^*$, $\mathcal{B} : V \to M^*$ are defined by

$$\langle \mathcal{A}u, v\rangle_\Omega = a(u,v), \quad v \in V, \quad b(v,\mu) = \langle \mathcal{B}v, \mu\rangle_\Omega, \quad \mu \in M.$$

It is well known that in both cases $\mathcal{L}$ is an isomorphism from $H_1 := V \times M$ onto $H_2 := V^* \times M^*$, that is, (1.29) is valid (Braess 1997, Brezzi and Fortin 1991, Girault and Raviart 1986), which in this case means that

$$\inf_{\mu \in M} \sup_{v \in V} \frac{b(v,\mu)}{\|v\|_V \|\mu\|_M} \geq \beta > 0. \tag{2.13}$$

Note that in the case (2.11) the Galerkin approximation of $\mathcal{A}$ is a *mass matrix*. Introducing suitable weighted inner products on a high discretization level would precondition this part well, which is one possible strategy for dealing with fluctuating coefficients.

However, the numerical solution of (2.9) or (2.11) now poses additional difficulties. The operator $\mathcal{L}$ is no longer definite. Preconditioning therefore requires additional care. Furthermore, the discretizations of V and M must be compatible, that is, (2.13) has to hold *uniformly* in the family of trial spaces under consideration. Both issues, preconditioning as well as the construction of compatible trial spaces, will be discussed below.

(c) *Time-dependent problems*

Once elliptic problems of the above type can be handled, the next step is to consider problems of the form

$$\begin{aligned} \frac{\partial u}{\partial t} + \mathcal{L}u + \mathcal{G}(u) &= 0, \quad u(k) = u(k+l), \quad k, l \in \mathbb{Z}^n, \\ u(\cdot, 0) &= u_0, \end{aligned} \tag{2.14}$$

where $\mathcal{L}$ is an elliptic operator of the form (2.4), and $\mathcal{G}$ is a possibly nonlinear function of u or a first-order derivative of u. Prominent examples are *reaction diffusion* equations

$$\frac{\partial u}{\partial t} = \nu \frac{\partial^2}{\partial x^2} u + u^p, \quad p > 1, \quad \nu > 0, \tag{2.15}$$

or the *viscous Burgers equation*

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = \nu\frac{\partial^2}{\partial x^2}u, \tag{2.16}$$

which describes the formation of shocks. Several wavelet schemes for this type of equation will be discussed. Some will also apply to problems such as the *Korteweg-de Vries* equations

$$\frac{\partial u}{\partial t} + \alpha u\frac{\partial u}{\partial x} + \beta\frac{\partial^3}{\partial x^3}u = 0, \tag{2.17}$$

α, β constant, having special soliton solutions (Fornberg and Whitham 1978).

(d) *Boundary integral equations*
Many classical partial differential equations can be transformed into *boundary integral equations.* This includes the *Lamé–Navier equations* of linearized, three-dimensional elasticity, (Wendland 1987), the *oblique derivative problem* (Michlin 1965), arising in physical geodesy, the *exterior Stokes flow* (Ladyshenskaya 1969); see Schneider (1995) for a brief overview. Here it suffices to describe a simple example that exhibits the principal features of this class of problem. Consider the boundary value problem

$$\Delta U = 0, \text{ on } \Omega, \quad \partial_\nu U = f, \text{ on } \Gamma := \partial\Omega, \tag{2.18}$$

where Ω is a bounded domain in $\mathbb{R}^3$ and ∂_ν denotes the derivative in the direction of the outer normal to Γ. It is well known that this boundary value problem, which arises, for instance, in the computation of electrostatic fields, is equivalent to the following integral equation of second kind provided by the so-called *indirect method*

$$\mathcal{L}u = f, \tag{2.19}$$

where $\mathcal{L} = \frac{1}{2}\mathcal{I} + \mathcal{K}$ and

$$(\mathcal{K}u)(x) = \frac{1}{4\pi}\int_\Gamma \frac{\nu_y^{\mathrm{T}}(x-y)}{|x-y|^3}u(y)\,\mathrm{d}s_y. \tag{2.20}$$

Here ν_y denotes the exterior normal of Γ at y. $\mathcal{K}$ is called *double layer potential.* For smooth Γ the operator $\mathcal{K}$ is compact on $L_2(\Gamma)$ so that the principal symbol of $\mathcal{L}$ is $1/2$. Thus (1.29) holds with $H_1 = H_2 = L_2(\Gamma)$ and K is a *zero order* operator. Clearly, denoting by $G(x-y) := \frac{1}{4\pi|x-y|}$ the fundamental solution of (2.18), one has $\mathcal{K}u(x) = \int_\Gamma \partial_{\nu,y}G(x-y)u(y)\,\mathrm{d}S_y$ and the solution U of (2.18) can be obtained by evaluating $U(x) = \int_\Gamma G(x-y)u(y)\,\mathrm{d}s_y$, where u is the solution of (2.19).

This approach is particularly tempting when (2.19) is to be solved on the *exterior* $\mathbb{R}^3 \setminus \Omega$ of some bounded domain. In this case one has to append certain *radiation conditions* at infinity to determine the solution uniquely.

The so-called *direct method* arises in connection with *transmission problems* and is well suited to dealing with other boundary conditions. Problem (2.18) subject to Dirichlet conditions $U = f$ on Γ is known to be equivalent to

$$\mathcal{L}u = \mathcal{V}u = \left(\frac{1}{2}\mathcal{I} - \mathcal{K}\right) f, \tag{2.21}$$

where

$$(\mathcal{V}u)(x) = \int_\Gamma \frac{u(y)}{4\pi|x-y|}\,\mathrm{d}y \tag{2.22}$$

is the *single layer* potential. In this case (1.29) can be shown to hold for $H_1 = H^{-1/2}(\Gamma)$ and $H_2 = H^{1/2}(\Gamma)$, and $\mathcal{L}$ has order *minus one*.

In both cases the unique solvability of (1.30) and (1.29) can be established along the following lines, which work for a much wider class of *pseudo-differential operators*. In fact, for smooth Γ these operators are classical pseudo-differential operators characterized by their *symbol*; see Hildebrandt and Wienholtz (1964), Kumano-go (1981). Equation (1.29) follows from the boundedness of $\mathcal{L}$, its injectivity on H_1, and *coercivity* of the principal part of its symbol.

The advantages of the approach are obvious. A 3D discretization of a possibly unbounded domain is reduced to a 2D discretization of a compact domain. One can also argue that in many cases the integral formulation is physically more adequate.

On the other hand, there are serious drawbacks. If the order of the operator $\mathcal{L}$ is different from zero, as in the case of the single layer potential operator, the need for preconditioning remains. In addition, conventional discretizations of the integral operators lead to dense matrices, which is perhaps the most severe obstruction to the use of these concepts for realistic problem sizes N. *Appropriate* wavelet bases will be seen to realize both desired effects (b), (c) in Section 1.5 for this class of problem.

2.3. A reference class of problems

The examples in Section 2.2 illustrate the variety of problems that will be discussed in this paper. To get some structure into the diversity of existing studies of various special cases, I stress the fact that certain results, mainly concerned with (b) in Section 1.5, actually hold in remarkable generality. Presenting them in this generality will help to bring out what really matters. In all the above examples the operator $\mathcal{L}$ satisfies (1.29) where H_1, H_2 are Sobolev spaces or products of such. In order to keep the discussion

homogeneous, we will confine the formulation of a model class of problems to the scalar case. So we assume that there exist some positive constants $c_1, c_2 < \infty$ such that

$$c_1\|\mathcal{L}u\|_{H^{-t}} \leq \|u\|_{H^t} \leq c_2\|\mathcal{L}u\|_{H^{-t}}, \tag{2.23}$$

where H^s stands for a suitable (subspace of a) Sobolev space (for instance, determined by homogeneous boundary conditions) and H^{-s} for its dual space. The underlaying domain may be a bounded domain in $\mathbb{R}^n$, $\mathbb{R}^n$ itself or a more general manifold such as a closed surface according to the above examples. Thus the problem

$$\mathcal{L}u = f \tag{2.24}$$

has for every $f \in H^{-t}$ a unique solution.

The analysis that follows will also cover operators with global Schwartz kernel

$$\mathcal{L}u = \int_\Gamma K(\cdot, x)u(x)\,\mathrm{d}x,$$

as considered in Section 2.2. As in the above examples, K will always be assumed to be smooth off the diagonal $x = y$. Moreover, it is to satisfy the following asymptotic estimates, which obviously hold in the above cases as well,

$$\left|\partial_x^\alpha \partial_y^\beta K(x,y)\right| \lesssim \operatorname{dist}(x,y)^{-(n+2t+|\alpha|+|\beta|)}, \tag{2.25}$$

where $r = 2t$ is the *order* of the operator.

3. Multiscale decompositions of refinable spaces

In Section 1.5 the *transform point of view* has been stressed. As indicated there the corresponding numerical schemes can be viewed as *Galerkin* or, more generally, (generalized) *Petrov–Galerkin* schemes. The point is that these schemes are always seen in connection with a whole ascending sequence of trial spaces, often referred to as *multiresolution analysis*. This permits the interaction of different scales of discretizations. In *basis* or *transform* oriented methods this is effected with the aid of appropriate *multiscale bases* of hierarchical type. Following Carnicer, Dahmen and Peña (1996), Dahmen (1994), Dahmen (1996) and Dahmen (1995) a general framework of multiresolution and multiscale decompositions of trial spaces is described next in a form which will later host all the required specializations. The examples in Sections 1.2 and 1.4 can be used as a conceptual as well as a notational orientation.

3.1. Multiresolution

The concept of *multiresolution* analysis plays a central role in the context of classical wavelets on $\mathbb{R}^n$. The anticipated applications here require a suitable generalization. In the spirit of Section 1.5, let H be a Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and associated norm $\| \cdot \| = \| \cdot \|_H = \langle \cdot, \cdot \rangle^{1/2}$. A multiresolution sequence $\mathcal{S} = \{S_j\}_{j \in \mathbb{N}_0}$ consists of nested closed subspaces $S_j \subset H$ whose union is dense in H

$$S_j \subset S_{j+1}, \quad \operatorname{clos}_H \left(\bigcup_{j \in \mathbb{N}_0} S_j \right) = H. \tag{3.1}$$

Define for any countable subset $\Phi \subset H$

$$S(\Phi) := \operatorname{clos}_H(\operatorname{span}\{\Phi\}),$$

the closure of the linear span of Φ. In all cases of practical interest the spaces S_j have the form

$$S_j := S(\Phi_j), \quad \Phi_j = \{\phi_{j,k} : k \in \Delta_j\} \tag{3.2}$$

for some (possibly infinite) index set Δ_j, where $\{\Phi_j\} = \{\Phi_j\}_{j \in \mathbb{N}_0}$ is *uniformly stable* in the sense that (see (1.23))

$$\|\mathbf{c}\|_{\ell_2(\Delta_j)} \sim \|\mathbf{c}^{\mathrm{T}} \Phi_j\|_H. \tag{3.3}$$

The Φ_j will sometimes be called *generator bases* or *single-scale bases.* The elements $\phi_{j,k}$ typically have good localization properties such as compact supports whose size depends on the scale j.

An arbitrary but fixed highest level of discretization will usually be denoted by J, and

$$N_J := \#\Delta_J$$

abbreviates the dimension of the corresponding space $S(\Phi_J)$.

Examples are $H = L_2([0,1])$ and $\phi_{j,k}$ the box or tent functions (see Sections 1.2 and 1.4) with $\Delta_j = \{0, \ldots, 2^j - 1\}$ or $\Delta_j = \{0, \ldots, 2^j\}$, respectively.

Two-scale relations

Nestedness of the spaces $S(\Phi_j)$ combined with (3.3) means that every $\phi_{j,k} \in S(\Phi_j)$ possesses an expansion

$$\phi_{j,k} = \sum_{l \in \Delta_{j+1}} m_{l,k}^j \phi_{j+1,l}$$

with a *mask* or *filter sequence* $\mathbf{m}_k^j = \{m_{l,k}^j\}_{l \in \Delta_{j+1}} \in \ell_2(\Delta_{j+1})$; recall (1.2) and (1.19). In our compact notation this can be rewritten as

$$\Phi_j^{\mathrm{T}} = \Phi_{j+1}^{\mathrm{T}} \mathbf{M}_{j,0}, \tag{3.4}$$

where the *refinement* matrix $\mathbf{M}_{j,0}$ contains the $\mathbf{m}_k^j$ as columns.

I will make frequent use of this notation, for two reasons. First, it saves several layers of indices. Second, it clearly brings out the conceptual similarities shared by all the technically different subsequent specializations. On the other hand, a word of warning is also appropriate. The special features of the actual implementation remain somewhat obscure. For instance, it will by no means always be necessary to assemble the complete matrices $\mathbf{M}_{j,0}$. In most cases its application to a vector amounts to applying local filters. Keeping this in mind, I still grant priority to convenience.

To illustrate (3.4), recall from (1.2) that the refinement matrix for the box functions is the $2^{j+1} \times 2^j$ matrix

$$\mathbf{M}_{j,0} = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & 0 & \dots & & 0 \\ \frac{1}{\sqrt{2}} & 0 & 0 & & & \\ 0 & \frac{1}{\sqrt{2}} & 0 & & & \\ \vdots & \vdots & \vdots & & & \vdots \\ & & & 0 & \frac{1}{\sqrt{2}} & 0 \\ & & & 0 & \frac{1}{\sqrt{2}} & 0 \\ & & & 0 & 0 & \frac{1}{\sqrt{2}} \\ 0 & \dots & & 0 & 0 & \frac{1}{\sqrt{2}} \end{pmatrix}, \tag{3.5}$$

whose dependence on j concerns only its size. Likewise, (1.19) gives the $(2^{j+1} - 1) \times (2^j - 1)$ matrix

$$\mathbf{M}_{j,0} = \begin{pmatrix} \frac{1}{2\sqrt{2}} & 0 & 0 & \dots & \\ \frac{1}{\sqrt{2}} & 0 & 0 & \dots & \\ \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & 0 & \dots & \\ 0 & \frac{1}{\sqrt{2}} & 0 & \dots & \\ 0 & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & 0 & \dots \\ 0 & 0 & \dots & & \\ \vdots & \vdots & \vdots & & \vdots \\ & & & & 0 \\ & & & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} \\ 0 & \dots & & 0 & \frac{1}{\sqrt{2}} \\ 0 & \dots & & 0 & \frac{1}{2\sqrt{2}} \end{pmatrix}. \tag{3.6}$$

3.2. Stable completions

Since the union of $\mathcal{S}$ is dense in H, a basis for H can be assembled from functions which span complements between any two successive trial spaces. One may think of orthogonal complements as in Section 1.2 or of the hierarchical complements in Section 1.4 induced by Lagrange interpolation (1.21). Depending on the case at hand, different choices will be seen to be preferable. So at this point we follow Carnicer et al. (1996) and keep the specific choices open. Thus one looks for collections $\Psi_j = \{\psi_{j,k} : k \in \nabla_j\} \subset S(\Phi_{j+1})$, such that

$$S(\Phi_{j+1}) = S(\Phi_j) \oplus S(\Psi_j), \tag{3.7}$$

and $\{\Phi_j \cup \Psi_j\}$ is still uniformly stable in the sense of (3.3). Like refinability, such decompositions may be expressed equivalently in terms of *matrix relations* that will provide a convenient algebraic platform for a unified treatment of subsequent specializations. As above, (3.7) implies that there exists some matrix $\mathbf{M}_{j,1}$ such that

$$\Psi_j^{\mathrm{T}} = \Phi_{j+1}^{\mathrm{T}} \mathbf{M}_{j,1}. \tag{3.8}$$

It is easy to see that (3.7) is equivalent to the fact that the operator

$$\mathbf{M}_j := (\mathbf{M}_{j,0}, \mathbf{M}_{j,1}),$$

defined by $\mathbf{M}_j \binom{\mathbf{c}}{\mathbf{d}} := \mathbf{M}_{j,0}\mathbf{c} + \mathbf{M}_{j,1}\mathbf{d}$, for $\mathbf{c} \in \ell_2(\Delta_j)$, $\mathbf{d} \in \ell_2(\nabla_j)$, is *invertible* as a mapping from $\ell_2(\Delta_j) \times \ell_2(\nabla_j)$ onto $\ell_2(\Delta_{j+1})$. Moreover, $\{\Phi_j \cup \Psi_j\}$ is uniformly stable if and only if

$$\|\mathbf{M}_j\|, \|\mathbf{M}_j^{-1}\| = O(1), \quad j \in \mathbb{N}, \tag{3.9}$$

where $\|\cdot\|$ is the spectral norm (Carnicer et al. 1996).

It is convenient to block $\mathbf{M}_j^{-1}$ as

$$\mathbf{M}_j^{-1} =: \mathbf{G}_j = \begin{pmatrix} \mathbf{G}_{j,0} \\ \mathbf{G}_{j,1} \end{pmatrix}, \tag{3.10}$$

so that

$$\mathbf{I} = \mathbf{M}_j \mathbf{G}_j = \mathbf{M}_{j,0}\mathbf{G}_{j,0} + \mathbf{M}_{j,1}\mathbf{G}_{j,1} \tag{3.11}$$

and

$$\mathbf{G}_{j,e}\mathbf{M}_{j,e'} = \delta_{e,e'}\mathbf{I}, \quad e, e' \in \{0,1\}. \tag{3.12}$$

Of course, those who are familiar with wavelets recognize in (3.11) the classical filter relations. The matrix $\mathbf{M}_j$ describes a *change of bases* and hence the reverse change $\mathbf{G}_j$, that is, Φ_{j+1} can be expressed in terms of the coarse scale basis Φ_j and the complement basis Ψ_j. One readily concludes from (3.4), (3.8) and (3.11) the *reconstruction* relation

$$\Phi_{j+1}^{\mathrm{T}} = \Phi_j^{\mathrm{T}}\mathbf{G}_{j,0} + \Psi_j^{\mathrm{T}}\mathbf{G}_{j,1}. \tag{3.13}$$

In general it may be difficult to identify the inverse $\mathbf{G}_j$, or, better, to arrange $\mathbf{M}_{j,1}$ in such a way that also $\mathbf{G}_j$ has a nice structure such as sparseness. One rather expects that when $\mathbf{M}_j$ is sparse, $\mathbf{G}_j$ will be full. In some sense, the art of wavelet construction can be viewed as finding the exceptions.

It is again instructive to recall the examples in Section 1. The relation (1.3), defining the Haar wavelet, corresponds to the $2^{j+1} \times 2^j$ matrix

$$\mathbf{M}_{j,1} = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & 0 & \dots & & 0 \\ -\frac{1}{\sqrt{2}} & 0 & 0 & & & \\ 0 & \frac{1}{\sqrt{2}} & 0 & & & \\ \vdots & \vdots & \vdots & & & \vdots \\ & & & 0 & \frac{1}{\sqrt{2}} & 0 \\ & & & 0 & -\frac{1}{\sqrt{2}} & 0 \\ & & & 0 & 0 & \frac{1}{\sqrt{2}} \\ 0 & \dots & & 0 & 0 & -\frac{1}{\sqrt{2}} \end{pmatrix}. \tag{3.14}$$

Since the Haar system is orthonormal, one simply has in this case

$$\mathbf{G}_j = \mathbf{M}_j^{\mathrm{T}}, \quad \|\mathbf{M}_j\| = \|\mathbf{M}_j^{-1}\| = 1. \tag{3.15}$$

Adding and subtracting (1.2) and (1.3), one could also deduce directly that

$$\phi_{j+1,2k} = \frac{1}{\sqrt{2}}(\phi_{j,k} + \psi_{j,k}^H), \quad \phi_{j+1,2k+1} = \frac{1}{\sqrt{2}}(\phi_{j,k} - \psi_{j,k}^H).$$

For the hierarchical basis from Section 1.4 one obtains the $(2^{j+1} - 1) \times 2^j$ matrix

$$\mathbf{M}_{j,1} = \begin{pmatrix} 1 & 0 & 0 & & & \\ 0 & 0 & 0 & & & \\ 0 & 1 & 0 & & & \\ \vdots & \vdots & \vdots & & & \vdots \\ & & & 0 & 1 & 0 \\ & & & 0 & 0 & 0 \\ & & & 0 & 0 & 1 \end{pmatrix}. \tag{3.16}$$

Moreover, since by (1.19) and (1.22),

$$\begin{aligned} \phi_{j+1,2k} &= \sqrt{2}\,\phi_{j,k} - \frac{1}{2}\,(\phi_{j+1,2k-1} + \phi_{j+1,2k+1}) \\ &= \sqrt{2}\,\phi_{j,k} - \frac{1}{2}\,(\psi_{j,k-1} + \psi_{j,k}), \quad k = 1, \dots, 2^j - 1, \\ \phi_{j+1,0} &= \sqrt{2}\,\phi_{j,0} - \frac{1}{2}\psi_{j,0}, \quad \phi_{j+1,2^{j+1}} = \sqrt{2}\,\phi_{j,2^j} - \frac{1}{2}\psi_{j,2^j-1}, \end{aligned}$$

while

$$\phi_{j+1,2k+1} = \psi_{j,k}, \quad k = 0, \ldots, 2^j - 1,$$

one readily identifies, in view of (3.13), the inverse $\mathbf{G}_j$ as

$$\mathbf{G}_{j,0} = \begin{pmatrix} 0 & \sqrt{2} & 0 & 0 & 0 & & \ldots & 0 & 0 \\ 0 & 0 & 0 & \sqrt{2} & 0 & & \vdots & 0 & 0 \\ \vdots & \vdots & & & & & \vdots & & \vdots \\ & & & & & \sqrt{2} & 0 & 0 & 0 \\ 0 & 0 & \ldots & & & 0 & 0 & \sqrt{2} & 0 \end{pmatrix}, \tag{3.17}$$

and

$$\mathbf{G}_{j,1} = \begin{pmatrix} 1 & -\frac{1}{2} & 0 & 0 & \ldots & & \\ 0 & -\frac{1}{2} & 1 & -\frac{1}{2} & & & \\ & & & & & & \\ & & & & & -\frac{1}{2} & 0 \\ \ldots & & & & & -\frac{1}{2} & 1 \end{pmatrix}. \tag{3.18}$$

Again one trivially has $\|\mathbf{M}_j\|, \|\mathbf{G}_j\| = O(1)$, $j \in \mathbb{N}_0$, so that the hierarchical complement bases are also uniformly stable in the above sense.

Remark 3.1 Evidently, the identification of a complement basis (3.7) is equivalent to *completing* a given refinement matrix $\mathbf{M}_{j,0}$ to an invertible mapping. Any $\mathbf{M}_{j,1}$ for which the completed matrix $\mathbf{M}_j$ satisfies (3.9) will be called *stable completion* of $\mathbf{M}_{j,0}$.

3.3. Multiscale bases

Repeating the decomposition (3.7), one can write each space $S(\Phi_J)$ as a sum of complement spaces

$$S(\Phi_J) = S(\Phi_0) \bigoplus_{j=0}^{J-1} S(\Psi_j).$$

Accordingly, $g_J \in S(\Phi_J)$ can be expanded in *single-scale* form with respect to Φ_J as

$$g_J = \Phi_J^{\mathrm{T}} \mathbf{c}^J, \tag{3.19}$$

as well as in *multiscale form* as

$$g_J = \Phi_0^{\mathrm{T}} \mathbf{c}^0 + \Psi_0^{\mathrm{T}} \mathbf{d}^0 + \ldots + \Psi_{J-1}^{\mathrm{T}} \mathbf{d}^{J-1}, \tag{3.20}$$

with respect to the multiscale basis

$$\Psi^J := \Phi_0 \bigcup_{j=0}^{J-1} \Psi_j. \tag{3.21}$$

Hence, by the denseness of $\mathcal{S}$ (3.1), the union

$$\Psi := \Phi_{j_0} \cup \bigcup_{j=j_0}^{\infty} \Psi_j =: \{\psi_\lambda : \lambda \in \nabla\} \tag{3.22}$$

is a candidate for a *basis* for the whole space H. Here j_0 is some fixed *coarsest level* (which, for simplicity, will usually be assumed to be $j_0 = 0$). We will always use the convention

$$\nabla := \Delta_+ \cup \nabla_-, \tag{3.23}$$

where

$$\Delta_+ := \Delta_{j_0}, \quad \psi_\lambda = \phi_{j_0,k}, \ \lambda := (j_0, k), \quad \nabla_- := \{(j,k) : k \in \nabla_j, j \in \mathbb{N}_0\}.$$

In principle, there is no need to consider only subsets Ψ^J of Ψ defined by *levelwise* truncation. Instead one can select *arbitrary* subsets $\Lambda \subset \nabla$ to form trial spaces

$$S_\Lambda := S(\Psi_\Lambda), \quad \Psi_\Lambda := \{\psi_\lambda : \lambda \in \Lambda\},$$

to discretize (1.30), say. According to (d) in Section 1.5, the selection of Λ, depending on a particular problem at hand, is a very natural way of steering *adaptivity*. This is perhaps one of the most promising aspects of multiscale basis-oriented methods in comparison with conventional discretizations.

3.4. Multiscale transformations

On the other hand, working with arbitrary subsets $\Lambda \subset \nabla$ will be seen to cause practical problems that should not be underestimated. Adequate data structures have yet to be developed. Things are much simpler for the special case

$$\Lambda_J := \{\lambda \in \nabla : |\lambda| < J\}, \tag{3.24}$$

where

$$|\lambda| = \begin{cases} j & \text{if} \quad \psi_\lambda \in \Psi_j, \\ j_0 - 1 & \text{if} \quad \lambda \in \Delta_+, \end{cases}$$

which deserves some special attention.

To this end, both coefficient vectors $\mathbf{c}$ and $\mathbf{d}$ appearing in (3.19), (3.20), respectively, convey different information. While $\mathbf{c}^J$ in (3.19) indicates in many cases, for instance, the geometrical location of the graph of g_J, the $\mathbf{d}^j$ in (3.20) have the character of *differences*. While usually all the entries of $\mathbf{c}^J$ are needed to represent g_J accurately, many of the entries in $\mathbf{d}$ may be small,

and replacing some of them by zero may still permit a sufficiently accurate approximation to g_J (recall (1.9) in Section 1.2). On the other hand, the pointwise evaluation of g_J is much simpler in the single-scale form (3.19). These questions will be encountered repeatedly in the course of subsequent developments.

To exploit the benefits of both representations, one needs a mechanism to convert one into the other. These transformations all have a common *pyramid structure*, which is explained next. Since by (3.4) and (3.8),

$$\Phi_j^{\mathrm{T}}\mathbf{c}^j + \Psi_j^{\mathrm{T}}\mathbf{d}^j = \Phi_{j+1}^{\mathrm{T}}\left(\mathbf{M}_{j,0}\mathbf{c}^j + \mathbf{M}_{j,1}\mathbf{d}^j\right),$$

the transformation

$$\mathbf{T}_J : \mathbf{d} \to \mathbf{c} \tag{3.25}$$

is schematically given by

$$\begin{array}{ccccccccc} & \mathbf{M}_{0,0} & & \mathbf{M}_{1,0} & & & \mathbf{M}_{J-1,0} & & \\ \mathbf{c}^0 & \to & \mathbf{c}^1 & \to & \mathbf{c}^2 & \to \cdots & \to & \mathbf{c}^J & \\ & \mathbf{M}_{0,1} & & \mathbf{M}_{1,1} & & & \mathbf{M}_{J-1,1} & & \\ & \nearrow & & \nearrow & & \nearrow \cdots & \nearrow & & \\ \mathbf{d}^0 & & \mathbf{d}^1 & & \mathbf{d}^2 & & \mathbf{d}^{J-1} & & . \end{array} \tag{3.26}$$

To express this in terms of matrix multiplications, define for $j < J$ the $\#\Phi_J \times \#\Phi_J$ matrix

$$\mathbf{T}_{J,j} := \begin{pmatrix} \mathbf{M}_j & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix},$$

where $\mathbf{I}$ is the identity block of size $\#\Phi_J - \#\Phi_{j+1}$. Then (3.26) becomes

$$\mathbf{T}_J = \mathbf{T}_{J,J-1} \cdots \mathbf{T}_{J,0}. \tag{3.27}$$

As for the inverse transformation, since, by (3.13),

$$\Phi_{j+1}^{\mathrm{T}}\mathbf{c}^{j+1} = \Phi_j^{\mathrm{T}}(\mathbf{G}_{j,0}\mathbf{c}^{j+1}) + \Psi_j^{\mathrm{T}}(\mathbf{G}_{j,1}\mathbf{c}^{j+1}) = \Phi_j^{\mathrm{T}}\mathbf{c}^j + \Psi_j^{\mathrm{T}}\mathbf{d}^j,$$

$\mathbf{T}_J^{-1}$ is realized by

$$\begin{array}{ccccccccc} & \mathbf{G}_{J-1,0} & & \mathbf{G}_{J-2,0} & & & \mathbf{G}_{0,0} & & \\ \mathbf{c}^J & \to & \mathbf{c}^{J-1} & \to & \mathbf{c}^{J-2} & \to \cdots & \to & \mathbf{c}^0 & \\ & \mathbf{G}_{J-1,1} & & \mathbf{G}_{J-2,1} & & & \mathbf{G}_{0,1} & & \\ & \searrow & & \searrow & & \searrow \cdots & \searrow & & \\ & & \mathbf{d}^{J-1} & & \mathbf{d}^{J-2} & & & \mathbf{d}^0, & \end{array} \tag{3.28}$$

which, of course, has a similar product structure as (3.27) involving the blocks $\mathbf{G}_j$.

Complexity of multiscale transformations

Let us comment first on the complexity of the transformations $\mathbf{T}_J$, $\mathbf{T}_J^{-1}$. In the above two examples (see (3.5), (3.6), (3.14), (3.16)) the matrices $\mathbf{M}_j$ and $\mathbf{G}_j$ have only finitely many nonzero entries in each column and row. Thus the operations that take $\mathbf{c}^j, \mathbf{d}^j$ into $\mathbf{c}^{j+1}$ as well as $\mathbf{c}^{j+1}$ into $\mathbf{c}^j, \mathbf{d}^j$ require the order of $\#\Delta_{j+1}$ operations uniformly in j. Since in both cases $\#\Delta_{j+1}/\#\Delta_j \sim \varrho > 1$ (here $\varrho = 2$), one concludes that the execution of $\mathbf{T}_J$ and $\mathbf{T}_J^{-1}$ requires the order of $\#\Delta_J = \dim S(\Phi_J)$ operations uniformly in $J \in \mathbb{N}$. Note that one need not assemble the global transformation $\mathbf{T}_J$ but rather apply local filters like (1.2) and (1.19) which correspond to the successive application of the factors $\mathbf{T}_{J,j}$. This pattern holds in much greater generality, as long as $\#\Delta_j/\#\Delta_{j-1} \geq \varrho > 1$, and the matrices $\mathbf{M}_j$, $\mathbf{G}_j$ stay *uniformly* sparse. By this we mean that the columns (rows) of $\mathbf{M}_j$, $(\mathbf{G}_j)$ contain only a uniformly bounded number of nonzero entries. Thus one may record the following for later use.

Remark 3.2 When all $\mathbf{M}_j$ are uniformly sparse and the cardinality of Φ_j grows geometrically, then the application of $\mathbf{T}_J$ requires $\mathcal{O}(\#\Delta_J)$ operations. Under the same assumptions on the $\mathbf{G}_j$ an analogous statement holds for the inverse transformation $\mathbf{T}_J^{-1}$.

Let us see next how the transformation $\mathbf{T}_J$ may enter a numerical scheme for the approximate solution of (1.30). Suppose one wants to employ a Galerkin scheme based on $S(\Phi_J)$, that is, one has to compute $u_J \in S(\Phi_J)$ satisfying

$$\langle \mathcal{L}u_J, v\rangle = \langle f, v\rangle, \quad v \in S(\Phi_J). \tag{3.29}$$

If u_J is to be represented in single-scale form $u_J = (\mathbf{c}^J)^T\Phi_J$, this amounts to solving the linear system

$$\langle \mathcal{L}\Phi_J, \Phi_J\rangle^T\mathbf{c}^J = \langle f, \Phi_J\rangle^T \tag{3.30}$$

for the unknown coefficient vector $\mathbf{c}^J$. As pointed out in Section 2.2, the matrix $\mathbf{A}_{\Phi_J} := \langle \mathcal{L}\Phi_J, \Phi_J\rangle^T$ may be sparse but increasingly ill-conditioned when J grows. In the special situation of Section 1.4 it has been observed that the stiffness matrix relative to the hierarchical basis has more favourable properties. One readily checks that, in general, the stiffness matrix $\mathbf{A}_{\Psi^J} := \langle \mathcal{L}\Psi^J, \Psi^J\rangle^T$ relative to the multiscale basis Ψ^J (see (3.21)) has the form

$$\mathbf{A}_{\Psi^J} = \mathbf{T}_J^T\mathbf{A}_{\Phi_J}\mathbf{T}_J. \tag{3.31}$$

Hence $\mathbf{A}_{\Psi^J}$ is a *principal section* of the (infinite) matrix

$$\mathbf{A}_\Psi := \langle \mathcal{L}\Psi, \Psi\rangle^T, \tag{3.32}$$

which is often called *standard representation* of $\mathcal{L}$.

Let us assume that $\mathcal{L}$ is a differential operator, so that, when Φ_J consists of compactly supported functions, $\mathbf{A}_{\Phi_J}$ is sparse and has only $\mathcal{O}(N_J)$ nonvanishing entries, where, as before, $N_J = \#\Delta_J$. Hence its accurate computation requires only the order of N_J operations and storage. Since the basis Ψ^J contains functions defined on coarse levels, basis functions from different scales will generally interact, so that $\mathbf{A}_{\Psi^J}$ will generally be much denser. However, in the context of iterative schemes, only the *application* of a matrix to a vector matters. By (3.31), the application of $\mathbf{A}_{\Psi^J}$ to a vector reduces to applying successively $\mathbf{T}_J, \mathbf{A}_{\Phi_j}$ and $\mathbf{T}_J^T$, each requiring, on account of Remark 3.2, the order of N_J operations.

In the above form these multiscale transformations are very efficient relative to the complexity of the *full* space $S(\Phi_J)$. At this point, though, it is not clear how to deal with spaces $S(\Psi_\Lambda)$ spanned by subsets of Ψ^J.

Stability and biorthogonality

There is obviously a continuum of possible complement bases Ψ_j that yield decompositions (3.7), and the question arises whether they are all equally suitable. The Haar basis corresponds to taking orthogonal complements relative to the ℓ_2-inner product, while the hierarchical basis spans orthogonal complements relative to the inner product $a(u,v) = \langle u', v' \rangle_{[0,1]}$ in $H_0^1([0,1])$. Thus orthogonal complements appear to be a canonical choice. However, they are frequently not easy to realize. For instance, any stable completion for (3.6), which induces orthogonal complements, is either dense or gives rise to dense inverses $\mathbf{G}_j$. Moreover, we will encounter situations where orthogonal complements are actually not the best choice.

At any rate, the qualification of the complement bases Ψ_j will be seen to depend crucially on the topological properties of their union Ψ. Aside from efficiency, a first reasonable constraint on the choice of the Ψ_j is the *stability* of the multiscale transformations; see, for example, Dahmen (1994, 1996).

Theorem 3.3 The transformations $\mathbf{T}_J$ are well conditioned in the sense that

$$\|\mathbf{T}_J\|, \|\mathbf{T}_J^{-1}\| = O(1), \quad J \in \mathbb{N}, \tag{3.33}$$

if and only if the collection Ψ, defined by (3.22), is a *Riesz basis* of H. This means that every $f \in H$ has unique expansions

$$f = \sum_{\lambda \in \nabla} \langle f, \tilde{\psi}_\lambda \rangle \psi_\lambda = \sum_{\lambda \in \nabla} \langle f, \psi_\lambda \rangle \tilde{\psi}_\lambda, \tag{3.34}$$

where $\tilde{\Psi} \subset H$ is a *biorthogonal Riesz basis*, that is,

$$\langle \Psi, \tilde{\Psi} \rangle = \mathbf{I}, \tag{3.35}$$

such that

$$\|f\|_H \sim \|\langle f, \Psi \rangle\|_{\ell_2(\nabla)} \sim \|\langle f, \tilde{\Psi} \rangle\|_{\ell_2(\nabla)}. \tag{3.36}$$

Thus *biorthogonality* is as far as one can deviate from orthogonality. It will be seen that the framework of biorthogonal bases offers a much more flexible setting for constructing multiscale bases such that the matrices $\mathbf{M}_j$ as well as their inverses $\mathbf{G}_j$ are uniformly sparse and give rise to well-conditioned multiscale transformations. Moreover, several schemes that have originally been formulated for orthogonal wavelets (at the expense of infinite although decaying filters) can be adapted to the biorthogonal setting with better localization in physical space.

Remark 3.4 Biorthogonality came out as a necessary condition. In general, it is not quite a sufficient condition for the Riesz basis property (3.36). In fact, as observed by Meyer (1994), not every Schauder basis in a separable Hilbert space is a Riesz basis. Additional conditions ensuring (3.36) will be discussed later.

3.5. Stable completions continued

Constructing a stable completion in the sense of Section 3.2 does not yet guarantee that a collection Ψ of the form (3.22) is a Riesz basis in H. Since in general we cannot resort to Fourier techniques, other tools are needed. As we have seen in Sections 1.2 and 1.4, sometimes *certain* stable completions can be found that may not yet have the desired form. For instance, the hierarchical bases in Sections 1.4 and 4.1 are *not* Riesz bases. In such cases a simple device will help, that allows one to modify the complement bases (Carnicer et al. 1996). It will have several applications later. The first important observation is that, once *some* stable completion is known, *all* others can be *parametrized* as follows.

Proposition 3.5 Suppose that Φ_j are uniformly stable with refinement matrices $\mathbf{M}_{j,0}$ and let $\check{\mathbf{M}}_{j,1}$ be some (uniformly) stable completion of $\mathbf{M}_{j,0}$. Let $\check{\mathbf{G}}_j := \binom{\check{\mathbf{G}}_{j,0}}{\check{\mathbf{G}}_{j,1}}$ denote the inverse of $\check{\mathbf{M}}_j = (\mathbf{M}_{j,0}, \check{\mathbf{M}}_{j,1})$. Then $\mathbf{M}_{j,1}$ is also a stable completion of $\mathbf{M}_{j,0}$, if and only if there exist

$$\mathbf{L}_j : \ell_2(\nabla_j) \to \ell_2(\Delta_j), \quad \mathbf{K}_j : \ell_2(\nabla_j) \to \ell_2(\nabla_j)$$

such that $\mathbf{L}_j$, $\mathbf{K}_j$, $\mathbf{K}_j^{-1}$ are uniformly bounded as operators and

$$\mathbf{M}_{j,1} = \mathbf{M}_{j,0}\mathbf{L}_j + \check{\mathbf{M}}_{j,1}\mathbf{K}_j. \tag{3.37}$$

Moreover, the inverse $\mathbf{G}_j$ of $\mathbf{M}_j = (\mathbf{M}_{j,0}, \mathbf{M}_{j,1})$ is given by

$$\mathbf{G}_{j,0} = \check{\mathbf{G}}_{j,0} - \mathbf{L}_j\mathbf{K}_j^{-1}\check{\mathbf{G}}_{j,1}, \quad \mathbf{G}_{j,1} = \mathbf{K}_j^{-1}\check{\mathbf{G}}_{j,1}. \tag{3.38}$$

Thus, given $\check{\mathbf{M}}_{j,1}$, varying $\mathbf{L}_j$ and $\mathbf{K}_j$ produces a whole family of further stable completions and corresponding decompositions of the spaces $S(\Phi_j)$. The special case $\mathbf{K}_j = \mathbf{I}$ covers the *lifting scheme* proposed by

Sweldens (1996, 1997). In this case one has

$$\Psi_j^{\mathrm{T}} = \Phi_{j+1}^{\mathrm{T}}\mathbf{M}_{j,1} = \Phi_{j+1}^{\mathrm{T}}\mathbf{M}_{j,0}\mathbf{L}_j + \Phi_{j+1}^{\mathrm{T}}\check{\mathbf{M}}_{j,1} = \Phi_j^{\mathrm{T}}\mathbf{L} + \check{\Psi}_j,$$

that is, in terms of individual functions, one has

$$\psi_{j,k} = \sum_{l\in\Delta_j} (\mathbf{L}_j)_{l,k}\phi_{j,l} + \check{\psi}_{j,k}. \tag{3.39}$$

Thus the new wavelet $\psi_{j,k}$ is obtained from the initial wavelet $\check{\psi}_{j,k}$ by adding a linear combination of coarse scale generating functions.

Now the task remains to pick from the above family of stable completions a certain desired one. Specifically, we will have to identify stable completions associated with linear projectors of the form $\langle\cdot,\Xi_j\rangle\Phi_j$ where $\langle\Phi_j,\Xi_j\rangle = \mathbf{I}$.

In fact, Carnicer et al. (1996) have shown that

$$\mathbf{M}_{j,1} = (\mathbf{I} - \mathbf{M}_{j,0}\langle\Phi_{j+1},\Xi_j\rangle^{\mathrm{T}})\,\check{\mathbf{M}}_{j,1} \tag{3.40}$$

are also stable completions with

$$\mathbf{G}_{j,0} = \check{\mathbf{G}}_{j,0} + \langle\Phi_{j+1},\Xi_j\rangle^{\mathrm{T}}\check{\mathbf{M}}_{j,1}\check{\mathbf{G}}_{j,1}, \quad \mathbf{G}_{j,1} = \check{\mathbf{G}}_{j,1}. \tag{3.41}$$

This obviously corresponds to the case $\mathbf{K}_j = \mathbf{I}$ and

$$\mathbf{L}_j = -\langle\Phi_{j+1},\Xi_j\rangle^{\mathrm{T}}\check{\mathbf{M}}_{j,1}. \tag{3.42}$$

To see the relevance of this latter observation in the present context, let for any $\Lambda \subset \nabla$

$$\Psi_\Lambda := \{\psi_\lambda : \lambda \in \Lambda\}. \tag{3.43}$$

If Ψ and $\tilde{\Psi}$ are biorthogonal collections (3.35), then

$$Q_\Lambda v := \langle v, \tilde{\Psi}_\Lambda\rangle\Psi_\Lambda, \quad Q_\Lambda^* v := \langle v, \Psi_\Lambda\rangle\tilde{\Psi}_\Lambda, \tag{3.44}$$

are *projectors* onto the spaces $S(\Psi_\Lambda), S(\tilde{\Psi}_\Lambda)$, respectively, which are adjoints of each other. In particular, for $\Lambda = \Lambda_j$ we simply write $Q_j = Q_{\Lambda_j}$.

Remark 3.6 If Ψ and $\tilde{\Psi}$ are biorthogonal, then

$$Q_\Lambda Q_{\hat{\Lambda}} = Q_\Lambda \quad \text{when} \quad \Lambda \subseteq \hat{\Lambda} \subset \nabla. \tag{3.45}$$

If in addition (3.36) holds, then the Q_Λ, Q_Λ^* are uniformly bounded in H, $\Lambda \subset \nabla$.

Suppose now that the desired biorthogonal multiscale bases are not yet known. Projectors can be also represented with respect to the basis Φ_j of $S(\Phi_j) = S(\Psi^j)$. So let

$$Q_j v = \langle v, \tilde{\Phi}_j\rangle\Phi_j, \tag{3.46}$$

where

$$\langle\Phi_j,\tilde{\Phi}_j\rangle = \mathbf{I}, \tag{3.47}$$

for some $\tilde{\Phi}_j \subset S(\tilde{\Psi}^j)$. We will see next what (3.45) means for the $\tilde{\Phi}_j$.

Remark 3.7 *The Q_j defined by (3.46) satisfy (3.45), if and only if the collection $\tilde{\Phi}_j$ is refinable, that is, there exists a matrix $\tilde{\mathbf{M}}_{j,0}$ such that*

$$\tilde{\Phi}_j^T = \tilde{\Phi}_{j+1}^T \tilde{\mathbf{M}}_{j,0}, \tag{3.48}$$

and

$$\tilde{\mathbf{M}}_{j,0}^* \mathbf{M}_{j,0} = \mathbf{I}. \tag{3.49}$$

The key to constructing biorthogonal wavelet bases is the following observation (Carnicer et al. 1996). The point is that if dual pairs of refinable generator bases $\Phi_j, \tilde{\Phi}_j$ satisfying (3.47) are given and *some initial* stable completion is known, then biorthogonal wavelets can easily be obtained as follows. One infers from (3.40) and (3.41) the following.

Proposition 3.8 Under the assumptions of Proposition 3.5,

$$\mathbf{M}_{j,1} = \left(\mathbf{I} - \mathbf{M}_{j,0}\tilde{\mathbf{M}}_{j,0}^*\right) \check{\mathbf{M}}_{j,1} \tag{3.50}$$

are also stable completions with

$$\mathbf{G}_{j,0} = \tilde{\mathbf{M}}_{j,0}^*, \quad \mathbf{G}_{j,1} = \check{\mathbf{G}}_{j,1}. \tag{3.51}$$

Moreover, $\tilde{\mathbf{M}}_{j,1} := \mathbf{G}_{j,1}^*$ is a stable completion of $\tilde{\mathbf{M}}_{j,0}$ and the collections Ψ, $\tilde{\Psi}$ obtained from

$$\Psi_j^T := \Phi_{j+1}^T \mathbf{M}_{j,1}, \quad \tilde{\Psi}_j^T := \tilde{\Phi}_{j+1}^T \tilde{\mathbf{M}}_{j,1}, \tag{3.52}$$

by (3.22), are biorthogonal.

Note that when $\check{\mathbf{M}}_j, \check{\mathbf{G}}_j$ and $\tilde{\mathbf{M}}_{j,0}$ are sparse, then the biorthogonal wavelets in Ψ and $\tilde{\Psi}$ have compact support.

4. Examples

The objective of this section is to identify several specializations of the setting described in Section 3, which will be needed later.

4.1. Hierarchical bases

The first example concerns the bivariate counterpart to the construction in Section 1.4. It has attracted considerable attention in connection with the *hierarchical bases preconditioner* (Yserentant 1986).

Suppose Ω is a bounded polygonal domain in $\mathbb{R}^2$ and $\mathcal{T}_0$ is some triangulation of Ω. This means the union of triangles in $\mathcal{T}_0$ agrees with $\overline{\Omega}$ and the intersection of any two different triangles $\tau, \tau' \in \mathcal{T}_0$ is either empty or a common vertex or a common edge. A sequence of triangulations $\mathcal{T}_j$ is then obtained by subdividing each $\tau \in \mathcal{T}_{j-1}$ into four congruent triangles. With

each $\mathcal{T}_j$ we associate the space S_j of continuous piecewise linear functions on Ω. Thus, as in the univariate case (see Section 1.4), tent functions form a basis for S_j. In fact, denoting by $\phi_{j,k}$ the unique piecewise linear function which has the value 2^j at the vertex k of $\mathcal{T}_j$ while vanishing at all other vertices, one can show that the corresponding collections Φ_j are uniformly stable (3.3); see, for instance, Oswald (1990). It is clear that the union of the $S(\Phi_j)$ is dense in $H = L_2(\Omega)$.

The *hierarchical bases* are obtained by adding to Φ_j just those basis functions on the next level that correspond to the *new* vertices at the midpoints of the edges in $\mathcal{T}_j$. Thus, denoting by Δ_j the vertices in $\mathcal{T}_j$ and by ∇_j the midpoints of the edges in $\mathcal{T}_j$ or, equivalently, $\nabla_j = \Delta_{j+1} \setminus \Delta_j$, and calling $\mathcal{N}_{j+1,k}$ for $k \in \Delta_j$ the set of neighbouring vertices of k in Δ_{j+1}, one has

$$\phi_{j,k} = \sum_{m \in \{k\} \cup \mathcal{N}_{j+1,k}} 2^{-j-1} \phi_{j,k}(m) \phi_{j+1,m}, \quad k \in \Delta_j, \tag{4.1}$$

that is, the entries of $\mathbf{M}_{j,0}$ are given by

$$(\mathbf{M}_{j,0})_{m,k} = 2^{-j-1} \phi_{j,k}(m) = \begin{cases} \frac{1}{2}, & m = k, \\ \frac{1}{4}, & m \in \mathcal{N}_{j+1,k}, \\ 0, & \text{else.} \end{cases} \tag{4.2}$$

Since

$$\psi_{j,k} := \phi_{j+1,k}, \quad k \in \nabla_j \tag{4.3}$$

one has the completion

$$(\mathbf{M}_{j,1})_{m,k} = \delta_{m,k}, \quad m \in \Delta_{j+1}, \quad k \in \nabla_j. \tag{4.4}$$

On the other hand, since also for $m \in \Delta_j$ one has $\mathcal{N}_{j+1,m} \subseteq \nabla_j$, (4.1) and (4.3) imply

$$\phi_{j+1,m} = 2\phi_{j,m} - \sum_{k \in \mathcal{N}_{j+1,m}} \frac{1}{2} \psi_{j,k}, \tag{4.5}$$

so that in this case we infer from (3.13)

$$(\mathbf{G}_{j,0})_{k,m} = 2\delta_{k,m}, \quad m \in \Delta_{j+1}, \quad k \in \Delta_j, \tag{4.6}$$

and

$$(\mathbf{G}_{j,1})_{k,m} = \begin{cases} -\frac{1}{2}, & m \in \Delta_j, k \in \mathcal{N}_{j+1,m}, \\ \delta_{k,m}, & k, m \in \nabla_j, \\ 0, & \text{else.} \end{cases} \tag{4.7}$$

Since obviously $\|\mathbf{M}_j\|, \|\mathbf{G}_j\| = O(1)$, $j \in \mathbb{N}$, the $\mathbf{M}_{j,1}$, defined by (4.4), are indeed uniformly stable completions.

However, note that

$$S(\Psi_j) = (L_{j+1} - L_j)S(\Phi_{j+1}) \tag{4.8}$$

where the L_j are the *interpolation* projectors defined by

$$L_j f := \sum_{k \in \Delta_j} 2^{-j} f(k) \phi_{j,k}.$$

Hence the basis Ψ obtained in this way (see (3.22)) has no dual in $L_2(\Omega)$ and is therefore *not* a Riesz basis. However, it will serve as a convenient initial stable completion in the sense of Proposition 3.8. Some consequences of these facts will be discussed later in connection with preconditioning.

4.2. Wavelets on $\mathbb{R}^n$

The construction of wavelet bases is best understood for $H = L_2(\mathbb{R})$, where the notion of *multiresolution analysis* has originated from Mallat (1989), Meyer (1990) and Daubechies (1988).

Stationary multiresolution

Let us first consider the univariate case $n = 1$. Suppose that $\phi \in L_2(\mathbb{R})$ has *stable shifts*

$$\|\mathbf{c}\|_{\ell_2(\mathbb{Z})} \sim \left\| \sum_{k \in \mathbb{Z}} c_k \phi(\cdot - k) \right\|_{L_2(\mathbb{R})} \tag{4.9}$$

and is *refinable*, that is, there exists a *mask* $\mathbf{a} \in \ell_2(\mathbb{Z})$ such that

$$\phi(x) = \sum_{k \in \mathbb{Z}} a_k \phi(2x - k), \quad x \in \mathbb{R}, \text{ almost everywhere.} \tag{4.10}$$

Hence the collections

$$\Phi_j := \left\{ \phi_{j,k} := 2^{j/2} \phi(2^j \cdot - k) : k \in \mathbb{Z} \right\}$$

are *uniformly stable* (3.3) and satisfy (3.4) with $\mathbf{M}_{j,0} = \mathbf{M}_0 = (a_{l-2k})_{l,k \in \mathbb{Z}}$. Thus the refinement matrices are stationary, that is, they are independent of the scale j and the spatial location k. The examples from Sections 1.2 and 1.4 are obviously obtained by restricting collections Φ_j of this type to $[0, 1]$. The function ϕ is often called *scaling function* or *generator* of the multiresolution sequence $\mathcal{S} = \{S(\Phi_j)\}_{j \in \mathbb{Z}}$, which is known to be dense in $L_2(\mathbb{R})$; see, for example, de Boor, DeVore and Ron (1993) and Jia and Micchelli (1991).

Time-frequency analysis and Fourier techniques have been an indispensible source of construction tools. It is well known (de Boor et al. 1993,

Daubechies 1992, Mallat 1989, Jia and Micchelli 1991) that, in terms of the Fourier transform, stability (4.9) is equivalent to

$$\sum_{k\in\mathbb{Z}} |\hat{\phi}(y+2\pi k)|^2 \geq c > 0, \tag{4.11}$$

while the refinement relation (4.10) reads

$$\hat{\phi}(y) = 2^{-1} a(e^{-iy/2})\hat{\phi}(y/2). \tag{4.12}$$

The Laurent polynomial

$$a(z) = \sum_{k\in\mathbb{Z}} a_k z^k$$

is called the *symbol* of the mask **a**. Since under the present assumptions $\hat{\phi}$ is continuous, reiteration of (4.12) yields

$$\hat{\phi}(y) = \left\{ \prod_{j=1}^{\infty} \left(2^{-1} a(e^{-i2^{-j}y})\right) \right\} \hat{\phi}(0), \tag{4.13}$$

where the product converges uniformly on compact sets so that we always have $\hat{\phi}(0) \neq 0$. Thus we may assume that ϕ is normalized to $\hat{\phi}(0) = 1$.

An important special case arises when the shifts $\phi(\cdot - k)$ are *orthonormal* so that (4.9) becomes an equality. An example is the scaling function (see Daubechies (1992, page 137))

$$\hat{\phi}(y) := \begin{cases} 1, & |y| \leq 2\pi/3, \\ \cos\left(\frac{\pi}{2}\nu(\frac{3}{2\pi}|y| - 1)\right), & 2\pi/3 \leq |y| \leq 4\pi/3, \\ 0, & \text{otherwise}, \end{cases} \tag{4.14}$$

where ν is a smooth function satisfying

$$\nu(x) = \begin{cases} 0, & x \leq 0, \\ 1, & x \geq 1. \end{cases}$$

Another interesting example is

$$\hat{\phi}(y) = 2 + (1 - e^{-iy})q(y), \tag{4.15}$$

where the trigonometric polynomial q is chosen, so that the shifts $\phi(\cdot - k), k \in \mathbb{Z}$ are orthonormal and

$$\frac{d^l}{dy^l}\hat{\phi}\,|_{y=0} = \delta_{0,l}, \quad l = 0, \ldots, d-1.$$

This latter condition means that

$$\int_{\mathbb{R}} x^l \phi(x)\,dx = \delta_{0,l}, \quad l = 0, \ldots, d-1, \tag{4.16}$$

that is, the scaling function ϕ also has certain vanishing moments. Using

essentially the same Taylor expansion argument as in Section 1.3, condition (4.16) implies that, for smooth f, one has $\langle f, \phi_{J,k}\rangle_{\mathbb{R}} \approx 2^{J/2} f(2^{-J}k)$. In fact, one can show that, for instance,

$$\left| \sum_{k\in\mathbb{Z}} f(2^{-J}k)\phi(2^J l - k) - f(2^{-J}l) \right| \lesssim 2^{-Jd} \|f\|_{W^{d,\infty}(\mathbb{R})}, \tag{4.17}$$

so that the expansion $\sum_{k\in\mathbb{Z}} f(2^{-J}k)\phi(2^J x - k)$ *almost interpolates* f. While ϕ from (4.14) has global support, the support width of ϕ from (4.15) is $3d-1$ (Daubechies 1992, page 258).

The now famous scaling functions ϕ with orthonormal shifts of smaller support (of width $2d-1$) have been constructed by Daubechies (1988, 1992).

When the shifts of ϕ are orthonormal, it can be verified that the shifts of

$$\psi(x) := \sum_{k\in\mathbb{Z}} (-1)^k a_{1-k} \phi(2x - k) \tag{4.18}$$

form an orthonormal basis of the orthogonal complement of $S(\Phi_0)$ in $S(\Phi_1)$, so that the corresponding $\psi_{j,k}$ constitute an orthonormal basis for $L_2(\mathbb{R})$. The function ϕ from (4.14) gives rise to the *Meyer wavelet* (Meyer 1990) which has extremely good localization in Fourier space but has rather slow decay in physical space. The wavelets for (4.15) are called *coiflets* and will be referred to later again.

In general one can say that $\hat{\phi}$ and $\hat{\psi}$ act like *low pass* and *band pass* filters. For an extensive discussion of this background see Daubechies (1992).

However, orthonormality will merely be viewed as a special case of the more flexible concept of biorthogonality that came up in Section 3.4; see Cohen, Daubechies and Feauveau (1992).

Dual pairs

The scaling functions $\phi, \tilde{\phi}$ are said to form a *dual pair* if

$$\langle \phi, \tilde{\phi}(\cdot - k)\rangle_{\mathbb{R}} := \int_{\mathbb{R}} \phi(x)\overline{\tilde{\phi}(x-k)}\,\mathrm{d}x = \delta_{0,k}, \quad k \in \mathbb{Z}. \tag{4.19}$$

We will sometimes refer to ϕ and $\tilde{\phi}$ as *primal* and *dual* generator, respectively. It is easy to see that compact support of ϕ and $\tilde{\phi}$ implies that the masks $\mathbf{a}$ and $\tilde{\mathbf{a}}$ have *finite support* and that (4.19) implies stability (4.9). Moreover, it is known that the functions

$$\psi(x) := \sum_{k\in\mathbb{Z}} (-1)^k \tilde{a}_{1-k} \phi(2x - k), \quad \tilde{\psi}(x) := \sum_{k\in\mathbb{Z}} (-1)^k a_{1-k} \tilde{\phi}(2x - k) \tag{4.20}$$

satisfy

$$\langle \phi, \tilde{\psi}(\cdot - k)\rangle_{\mathbb{R}} = \langle \tilde{\phi}, \psi(\cdot - k)\rangle_{\mathbb{R}} = 0, \quad \langle \psi, \tilde{\psi}(\cdot - k)\rangle_{\mathbb{R}} = \delta_{0,k}, \quad k \in \mathbb{Z}, \tag{4.21}$$

which obviously covers (4.18) as a special case. Observe next that straightforward computations confirm that the relations (4.19) and (4.21) are equivalent to

$$\begin{pmatrix} a(z) & a(-z) \\ b(z) & b(-z) \end{pmatrix} \begin{pmatrix} \overline{\tilde{a}(z)} & \overline{\tilde{b}(z)} \\ \overline{\tilde{a}(-z)} & \overline{\tilde{b}(-z)} \end{pmatrix} = \begin{pmatrix} 4 & 0 \\ 0 & 4 \end{pmatrix}. \tag{4.22}$$

This can be used for the construction of the dual generator $\tilde{\phi}$. Given the mask $\mathbf{a}$ one can determine $\tilde{\mathbf{a}}$ satisfying the first relation in (4.22) and then show that the product (4.13) with $\tilde{\mathbf{a}}$, instead of $\mathbf{a}$, is the Fourier transform of an L_2-function.

One easily deduces from (4.10) and (4.21) that for $\Psi_j := \{\psi_{j,k} : k \in \mathbb{Z}\}$, $\tilde{\Psi}_j := \{\tilde{\psi}_{j,k} : k \in \mathbb{Z}\}$ the collections

$$\Psi := \Phi_0 \bigcup_{j \geq 0} \Psi_j, \quad \tilde{\Psi} := \tilde{\Phi}_0 \bigcup_{j \geq 0} \tilde{\Psi}_j \tag{4.23}$$

are biorthogonal.

To relate this to the discussion in Section 3.2, note that with $b_k := (-1)^k \tilde{a}_{1-k}$, $\tilde{b}_k := (-1)^k a_{1-k}$ the bi-infinite matrix $\mathbf{M}_{j,1} = \mathbf{M}_1 := (b_{l-2k})_{l,k \in \mathbb{Z}}$ is a stable completion of $\mathbf{M}_0$ above and that in this case (see Proposition 3.8),

$$\mathbf{G}_0 = \tilde{\mathbf{M}}_0^* = \left(\bar{\tilde{a}}_{l-2k}\right)_{k,l \in \mathbb{Z}}, \quad \mathbf{G}_1 = \tilde{\mathbf{M}}_1^* = \left(\bar{\tilde{b}}_{l-2k}\right)_{k,l \in \mathbb{Z}}. \tag{4.24}$$

B-splines as primal generators give rise to an important class of dual pairs where both generators have compact support. Let $\lfloor x \rfloor$ ($\lceil x \rceil$) denote the largest (smallest) integer less (greater) than or equal to x, and define $N_d = \chi_{[0,1)} * \ldots * \chi_{[0,1)}$ as the d-fold convolution of the box function (1.1). Then, for

$$\phi = {}_d\phi := N_d\left(\cdot + \left\lfloor \tfrac{d}{2} \right\rfloor\right), \quad \hat{N}_d(y) = \left(\frac{1 - e^{-iy}}{iy}\right)^d, \tag{4.25}$$

(4.10) becomes

$${}_d\phi(x) = \sum_{k=-\lfloor \frac{d}{2} \rfloor}^{\lceil \frac{d}{2} \rceil} 2^{1-d} \binom{d}{k + \lceil \frac{d}{2} \rceil} {}_d\phi(2x - k). \tag{4.26}$$

Cohen et al. (1992) have shown that for every $d, \tilde{d} \in \mathbb{N}$, $\tilde{d} \geq d$, $d + \tilde{d}$ even, there exists a compactly supported scaling function ${}_{d,\tilde{d}}\tilde{\phi}$ such that $({}_d\phi, {}_{d,\tilde{d}}\tilde{\phi})$ form a dual pair. The role of the parameters $d, \tilde{d}$ will be pointed out below.

Polynomial exactness

It is remarkable that in the present *stationary setting* the refinement equation (4.12) has further important consequences. In fact, since, by (4.12),

$$\hat{\phi}(2\pi k 2^n) = \left\{ \prod_{j=1}^{n} \left(2^{-1} a(e^{-i2^{n-j}2\pi}) \right) \right\} \hat{\phi}(2\pi k),$$

letting n tend to infinity and applying the Riemann–Lebesgue lemma yields

$$\hat{\phi}(2\pi k) = 0, \quad k \in \mathbb{Z} \setminus \{0\}. \tag{4.27}$$

By the Poisson summation formula, this means that (Cavaretta, Dahmen and Micchelli 1991)

$$1 = \sum_{k \in \mathbb{Z}} \phi(x-k), \quad x \in \mathbb{R}. \tag{4.28}$$

Similarly, a somewhat refined argument shows that $\phi \in H^r(\mathbb{R})$ implies

$$\hat{\phi}^{(l)}(2\pi k) = 0, \quad k \in \mathbb{Z} \setminus \{0\}, \quad l = 0, \ldots, r, \tag{4.29}$$

(Cavaretta et al. 1991) so that Poisson's summation formula again implies that, for any polynomial p of degree at most r, there exists some polynomial q of lower degree such that

$$p(x) = \sum_{k \in \mathbb{Z}} p(k)\phi(x-k) + q(x). \tag{4.30}$$

In particular, when the scaling function ϕ also has vanishing moments (4.16), then the polynomial q can be shown to vanish. Combining this polynomial reproduction property with arguments from the proof of Proposition 5.1 below yields estimates of the form (4.17) above. The fact that shifts of ϕ represent polynomials of degree r exactly is reflected by the fact that the symbol $a(z)$ contains a power of $(1+z)$, that is

$$a(z) = (1+z)^{r+1} q(z) \tag{4.31}$$

where $q(1) = 2^{-r}$ (Daubechies 1992).

Returning to the above family $({}_d\phi, {}_{d,\tilde{d}}\tilde{\phi})$ of dual pairs, the parameters $d, \tilde{d}$ are exactly the respective orders of polynomial reproduction. Thus (4.19) yields

$$\begin{aligned} x^r &= \sum_{k \in \mathbb{Z}} \langle (\cdot)^r, {}_{d,\tilde{d}}\tilde{\phi}(\cdot - k) \rangle_{\mathbb{R}} \, {}_d\phi(x-k), \quad r = 0, \ldots, d-1, \\ x^r &= \sum_{k \in \mathbb{Z}} \langle (\cdot)^r, {}_d\phi(\cdot - k) \rangle_{\mathbb{R}} \, {}_{d,\tilde{d}}\tilde{\phi}(x-k), \quad r = 0, \ldots, \tilde{d}-1, \end{aligned} \tag{4.32}$$

which has two important consequences. On one hand, as indicated above, the order of polynomial reproduction governs the approximation power of

the spaces $S(\Phi_j)$; see, for instance, Cavaretta et al. (1991). This will be established later in somewhat greater generality. Here we mention first the following important further implication.

Moment conditions
As an immediate consequence of (4.21) and (4.32), we state that

$$\int_{\mathbb{R}} x^r \psi(x)\,dx = 0, \quad r = 0,\dots,\tilde{d}-1, \qquad \int_{\mathbb{R}} x^r \tilde{\psi}(x)\,dx = 0, \quad r = 0,\dots,d-1, \tag{4.33}$$

when $\psi, \tilde{\psi}$ are the wavelets (4.20) relative to the dual pair $({}_d\phi, {}_{d,\tilde{d}}\tilde{\phi})$ from above. The wavelets $\psi, \tilde{\psi}$ are said to have *vanishing moments* of order $\tilde{d}, d$, respectively. Recall from Section 1.3 that the order of vanishing moments governs the compression capacity of a wavelet. The fact that in connection with biorthogonal wavelets the order of vanishing moments can be chosen *independently* of the order of exactness will play an important role later.

Integration by parts
There is an important trick for generating a dual pair from another one, essentially by *integrating up* and *differentiating down* (Dahmen, Kunoth and Urban 1996*c*, Lemarié-Rieusset 1992, Urban 1995*a*). To this end, suppose that $(\phi, \tilde{\phi})$ is a dual pair and $\phi \in H^{1+\varepsilon}(\mathbb{R})$. By the previous remarks, its symbol $a(z)$ is divisible by $(1+z)$. The new symbols

$$a^-(z) := \frac{2}{1+z} a(z), \quad \tilde{a}^+(z) := \frac{1+\bar{z}}{z} \tilde{a}(z) \tag{4.34}$$

obviously still satisfy the first relation in (4.22). Moreover, the refinement relations (4.11) relative to the masks $\mathbf{a}^-, \tilde{\mathbf{a}}^+$ can be shown still to possess solutions $\phi^-, \tilde{\phi}^+ \in L_2(\mathbb{R})$ with compact support, which are related by

$$\frac{d}{dx}\phi(x) = \phi^-(x) - \phi^-(x-1), \quad \frac{d}{dx}\tilde{\phi}^+(x) = \tilde{\phi}(x+1) - \tilde{\phi}(x). \tag{4.35}$$

Since one still has $a^-(z)\overline{\tilde{a}^+(z)} + a^-(-z)\overline{\tilde{a}^+(-z)} = 4$, $(\phi^-, \tilde{\phi}^+)$ is still a dual pair. Moreover, the corresponding wavelets $\psi^-, \tilde{\psi}^+$, defined by (4.20), are related to $\psi, \tilde{\psi}$ by

$$\frac{d}{dx}\psi(x) = 4\psi^-(x), \quad \frac{d}{dx}\tilde{\psi}^+(x) = -4\tilde{\psi}(x). \tag{4.36}$$

We will have several opportunities to make use of these facts later.

The multivariate case
The simplest way of generating orthogonal or biorthogonal wavelets on $\mathbb{R}^n$ is via tensor products. Given any dual pair $(\varphi, \tilde{\varphi})$ of univariate scaling

functions, the products

$$\phi(x) := \varphi(x_1)\cdots\varphi(x_n), \quad \tilde{\phi}(x) := \tilde{\varphi}(x_1)\cdots\tilde{\varphi}(x_n) \tag{4.37}$$

obviously form a dual pair in $L_2(\mathbb{R}^n)$.

The corresponding masks are obtained from the univariate ones in a straightforward fashion. One should note that for scalings by powers of two one now needs $2^n - 1$ different wavelets whose shifts span the complement spaces. Setting $E := \{0,1\}^n$, $E_* := E \setminus \{0\}$, it is convenient to index these *mother wavelets* as follows,

$$\psi_e(x) = \psi_{e_1}(x_1)\cdots\psi_{e_n}(x_n), \quad e \in E_*, \tag{4.38}$$

where we sometimes denote for convenience $\psi_0 := \varphi$. The $\tilde{\psi}_e$ are defined analogously. Thus, while associating the functions

$$\phi_{j,k} := 2^{nj/2}\phi(2^j \cdot -k), \quad k \in \mathbb{Z}^n, \tag{4.39}$$

with the index set or grid $\Delta_j := 2^{-j}\mathbb{Z}^n$, the wavelets $\psi_{e,j,k}$, $\tilde{\psi}_{e,j,k}$ correspond to $\nabla_{e,j} := 2^{-j}\left(\frac{e}{2} + \mathbb{Z}^n\right)$, so that $\Delta_{j+1} = \Delta_j \cup \left(\bigcup_{e\in E_*} \nabla_{e,j}\right)$.

Several alternatives have been studied. First, one might look for *genuinely* multivariate scaling functions and wavelets. The practical relevance in terms of small masks and locality seems to be confined to a few special cases; see, for instance, Cohen and Schlenker (1993). On the other hand, the tensor product structure offers numerous advantages with regard to computational efficiency, via reduction to univariate problems, and data structures, as long as the underlying grid structure is regular. However, to reduce the number of mother wavelets, one might employ scalings by suitable integer matrices M with all eigenvalues strictly greater than one. One then needs $|\det M| - 1$ mother wavelets (Gröchenich and Madych 1992, Cohen and Daubechies 1993, Dahlke, Dahmen and Latour 1995). Again, much less machinery is available in this case. Finally, instead of considering spaces generated by a single scaling function, one can use a fixed finite collection of generators. In summary, however, since none of these approaches overcomes the obstructions posed by more complex domain geometries, it is fair to say they do not offer any significant advantages for the problems considered here.

Computational issues

Obviously, the stationary setting offers a variety of computational advantages. One need not assemble any level dependent refinement or completion matrices. The multiscale transformations (3.26) and (3.28) reduce to local applications of finite filter masks which are fixed once and for all; see Barsch, Kunoth and Urban (1997) for a discussion of these issues. The main point of this section is to present some computational techniques for basic tasks

like evaluating function values, derivatives and integrals of wavelets, which have no counterpart in conventional discretization settings.

Even though many scaling functions and hence corresponding wavelets possess no closed analytic representation, all essential information can be drawn from the masks. We will briefly exemplify this fact for the computation of integrals of products of scaling functions and wavelets or their derivatives. More details of the following facts can be found in Dahmen and Micchelli (1993) and Latto, Resnikoff and Tenenbaum (1992), and corresponding implementations are documented in Kunoth (1995).

Due to the two scale relations (4.20), integrals involving wavelets can be reduced to integrals involving only scaling functions. Thus Galerkin discretization of a partial differential equation requires evaluating terms like

$$\int_{\Omega} a(x)\partial^{\alpha}\phi_{j,k}(x)\partial^{\beta}\phi_{j,l}(x)\,\mathrm{d}x, \tag{4.40}$$

where a successive application of (4.10) has been used when wavelets on different levels j, j' are involved. Assuming for simplicity that Ω is a union of rectangular domains, the above integral can be written as

$$\sum_{m\in\mathbb{Z}^n}\int_{R^n}\chi_{j,m}(x)a(x)\partial^{\alpha}\phi_{j,k}(x)\partial^{\beta}\phi_{j,l}(x)\,\mathrm{d}x, \tag{4.41}$$

where $\chi = \chi_{\square}$ is the characteristic function of the unit cube $\square = [0,1]^n$. Due to the compact support of ϕ, the sum is actually finite and involves at most $|\operatorname{supp}\phi|$ terms.

Applying quadrature to quantities like (4.40) may not always be advisable, since although $a(x)$ may be very regular the accuracy of the quadrature is limited by the factors $\partial^{\alpha}\phi_{j,k}$, which may have very low regularity. Let us therefore point out how to evaluate (4.40) up to an accuracy that *only* depends on $a(x)$. To this end, let θ be any other scaling function such as a (tensor product) B-spline. Replacing $a(x)$ by some approximation $\sum_{l\in\mathbb{Z}^n} a_l\theta_{j,l}(x) =: a_j(x)$ which could, for instance, be obtained by interpolation, the compact support of θ again ensures that, when replacing $a(x)$ by $a_j(x)$ in (4.41), the sum over $l \in \mathbb{Z}^n$ is again finite, so that one ultimately has to compute after rescaling the quantities

$$\int_{\mathbb{R}}\chi_{\square}(x)\theta(x-k^1)\partial^{\alpha}\phi(x-k^2)\partial^{\beta}\phi(x-k^3)\,\mathrm{d}x. \tag{4.42}$$

Similar expressions arise when discretizing nonlinear terms such as those appearing in Burgers equation.

Here a new idea enters. The point is now that, given *any* finite number of (possibly different) scaling functions ϕ_i, (with finitely supported masks),

$i = 0, \ldots, m$, with $\phi_i \in C^r(\mathbb{R}^n)$ say, then expressions of the form

$$I(k^1, \ldots, k^m, \mu^1, \ldots, \mu^m) := \int_{\mathbb{R}^n} \phi_0(x) \prod_{i=1}^{m} \partial^{\mu^i} \phi_i(x - k^i)\, \mathrm{d}x \tag{4.43}$$

can be computed *exactly* (up to round-off). Thus the accuracy of the quantities in (4.40) depends *only* on the approximability of the coefficient $a(x)$.

This is essentially a consequence of refinability and its close connection with *subdivision techniques*; see Cavaretta et al. (1991), Dahmen and Micchelli (1993) and Latto et al. (1992). The main ideas are now sketched. Suppose ϕ is a scaling function. Differentiating and evaluating (4.10) at (multi-)integers, yields

$$2^{-|\mu|} \partial^\mu \phi(k) = \sum_{l \in \mathbb{Z}^n} a_{2k-l} \partial^\mu \phi(l). \tag{4.44}$$

Clearly $(\partial^\mu \phi(k) : k \in \mathbb{Z}^n)$ is finitely supported. Thus (4.44) may be seen as an eigenvector relation, that is, the vector $\mathbf{V}^\mu = (\partial^\mu \phi(k) : k \in \operatorname{supp} \phi)$ is an eigenvector of a finite section of the transpose of the refinement matrix for the eigenvalue $2^{-|\mu|}$. When $n > 1$, that is, μ is a multi-integer, *every* $\mathbf{V}^\mu$ with $|\mu| = r$ is an eigenvector with eigenvalue 2^{-r}.

To exploit these relations for evaluating $\partial^\mu \phi(k)$, $k \in \mathbb{Z}^n$, one therefore has to find suitable additional conditions and show that they actually identify each $\mathbf{V}^\mu$ *uniquely*. Before we describe such conditions we point out that,

(i) once $\partial^\mu \phi|_{\mathbb{Z}^n}$ is known, successive use of (4.10) yields $\partial^\mu \phi|_{2^{-j}\mathbb{Z}^n}, j \in \mathbb{N}$
(ii) this can be used to determine the integrals (4.43).

To explain this latter fact, let us catenate $(k^1, \ldots, k^m)$, $(\mu^1, \ldots, \mu^m)$ to vectors k, μ in $\mathbb{Z}^s$, $\mathbb{Z}_+^s$, respectively, where $s = mn$. Note that

$$I(k, \mu) = (-1)^\mu \partial^\mu F(k), \tag{4.45}$$

where

$$F(y) := \int_{\mathbb{R}^n} \phi_0(x) \phi_1(x - y^1) \cdots \phi_m(x - y^m)\, \mathrm{d}x.$$

The point is that F is again a refinable function with mask coefficients

$$c_k = 2^{-n} \sum_{l \in \mathbb{Z}^n} a_l^0 \prod_{i=1}^{m} a_{l-k^i}^i, \tag{4.46}$$

where $\mathbf{a}^i$ is the mask of ϕ_i.

Theorem 4.1 (Dahmen and Micchelli 1993) Suppose that all ϕ_i are stable in the sense of (4.9) and $\phi_i \in C^r(\mathbb{R}^n)$, $i = 1, \ldots, m$. Then for any $\mu \in \mathbb{Z}_+^{mn}$, $|\mu| \leq r$, there exists a unique sequence $\mathbf{V}^\mu$ of finite support in

$\mathbb{Z}^{mn}$, satisfying

$$2^{-|\mu|}V_k^\mu = \sum_{l\in\mathbb{Z}^{mn}} c_{2k-l}V_l^\mu, \quad k \in \mathbb{Z}^{mn}, \tag{4.47}$$

and

$$\sum_{k\in\mathbb{Z}^{mn}} (-k)^\nu V_k^\mu = \mu!\,\delta_{\nu,\mu}, \quad |\nu| \le |\mu|, \quad \nu,\mu \in \mathbb{N}_0^{mn}, \tag{4.48}$$

where $\mathbf{c}$ is defined by (4.46). Moreover, one has

$$\mathbf{V}_k^\mu = \partial^\mu F(k) = (-1)^\mu I(k,\mu), \quad k \in \mathbb{Z}^{mn}. \tag{4.49}$$

The moment conditions (4.48) are implied by the polynomial reproduction (4.30), (4.32). The proof that these conditions determine the $\mathbf{V}^\mu$ uniquely, employs the concept of subdivision algorithms (Dahmen and Micchelli 1993, Cavaretta et al. 1991).

Other variants of similar nature can be found in Dahmen and Micchelli (1993) and Sweldens and Piessens (1994), among them recursions for evaluating moments like $\int_{\mathbb{R}^n} x^\beta \phi(x-\alpha)\,\mathrm{d}x$.

Remark 4.2 The efficiency of this concept deteriorates when the factors in the integrals (4.40) or (4.41) involve functions on different scales, since this requires correspondingly many prior applications of refinement matrices. This problem does not arise when working with the so-called non-standard representation, which will be introduced later. Likewise, when $\mathcal{L}$ is a differential operator and ϕ has compact support, the above scheme can be used to compute the stiffness matrix $\mathbf{A}_{\Phi_J} := \langle \mathcal{L}\Phi_J, \Phi_J\rangle^{\mathrm{T}}$ accurately and efficiently. The multiscale transformation $\mathbf{T}_J$ (3.26) can then be employed to generate the stiffness matrix $\mathbf{A}_{\Psi J}$ (3.32) at the expense of $\mathcal{O}(N_J)$ operations. Again this may not be the best strategy for dealing with matrices $\mathbf{A}_{\Psi_\Lambda}$ for arbitrary $\Lambda \subset \nabla$.

4.3. Periodization

The above setting is clearly not suitable yet for the treatment of operator equations which are usually defined on bounded domains.

A very special but nevertheless important framework is the *periodic* setting (Meyer 1990). It essentially retains all the structural and computational advantages of the stationary shift-invariant case considered above. There are at least two reasons for addressing this case with great care. First, many effects will be seen to be local in nature and hence also provide important insight for more general situations. Second, one might aim at a *two-stage* process, trying to carry out the bulk of computation via the full spatial dimension relative to a periodized problem, while treating domain-related effects like boundary conditions separately.

The simple trick is to replace the meaning of $g_{j,k} := 2^{nj/2} g(2^j \cdot - k)$, $k \in \mathbb{Z}^n$, for compactly supported or rapidly decaying $g \in L_2(\mathbb{R}^n)$ by its *periodized* counterpart

$$g_{j,k}(x) := 2^{nj/2} \sum_{l \in \mathbb{Z}^n} g\left(2^j(x+l) - k\right). \tag{4.50}$$

Given any dual pair $(\phi, \tilde{\phi})$ on $\mathbb{R}^n$, and setting $\Delta_j := \mathbb{Z}^n / 2^j \mathbb{Z}^n$, the corresponding sets

$$\Phi_j := \{\phi_{j,k} : k \in \Delta_j\}, \quad \Psi_{e,j} := \{\psi_{e,j,k} : k \in \Delta_j\}, \quad e \in E_*, \tag{4.51}$$

and likewise $\tilde{\Phi}_j, \tilde{\Psi}_{e,j}$, have finite cardinality 2^{nj} and consist of functions which are *one-periodic* in each variable. Note that this preserves orthogonality relations. One easily checks that (4.19) and (4.21) still imply that

$$\langle \phi_{j,k}, \tilde{\phi}_{j,l} \rangle_\square = \int_\square \phi_{j,k}(x) \overline{\tilde{\phi}_{j,l}(x)} \, \mathrm{d}x = \delta_{k,l}, \quad k, l \in \Delta_j, \tag{4.52}$$

and that the collections

$$\Psi := \Phi_0 \cup \bigcup_{j=0}^{\infty} \left(\bigcup_{e \in E_*} \Psi_{e,j} \right), \quad \tilde{\Psi} := \tilde{\Phi}_0 \cup \bigcup_{j=0}^{\infty} \left(\bigcup_{e \in E_*} \tilde{\Psi}_{e,j} \right), \tag{4.53}$$

are biorthogonal

$$\langle \Psi, \tilde{\Psi} \rangle_\square = \mathbf{I}. \tag{4.54}$$

Hence the $\mathcal{S} = \{S(\Phi_j)\}_{j \in \mathbb{N}_0}$, $\tilde{\mathcal{S}} = \{S(\tilde{\Phi}_j)\}_{j \in \mathbb{N}_0}$ form two biorthogonal multiresolution sequences fitting into the framework of Section 3 for $H = L_2(\mathbb{R}^n / \mathbb{Z}^n)$. One readily verifies that

$$\phi_{j,k}(x) = \sum_{l \in \Delta_{j+1}} \left(\sum_{m \in \mathbb{Z}^n} 2^{-n/2} a_{l-2k+2^{j+1}m} \right) \phi_{j+1,l}, \tag{4.55}$$

that is, the new masks are obtained by 2^{j+1}-periodization. Thus the refinement matrices $\mathbf{M}_{j,0}$ have *circulant* structure and analogously the completion $\mathbf{M}_{j,1}$ (as well as $\tilde{\mathbf{M}}_{j,0}$, $\tilde{\mathbf{M}}_{j,1}$).

Defining the discrete Fourier coefficients

$$\mathcal{F}_k(f) := \int_\square f(x) e^{-2\pi i x \cdot k} \, \mathrm{d}x, \quad k \in \mathbb{Z}^n,$$

it is clear that for the periodization $[g] := \sum_{m \in \mathbb{Z}^n} g(\cdot + m)$ one has

$$\mathcal{F}_k([g]) = \hat{g}(k), \quad k \in \mathbb{Z}^n.$$

Hence any results relative to $\mathbb{R}^n$ are readily related to corresponding results for $\mathbb{R}^n / \mathbb{Z}^n$; see also Fröhlich and Schneider (1995).

4.4. Wavelets on the interval

There is one further extension beyond $\mathbb{R}^n$ or $\mathbb{R}^n/\mathbb{Z}^n$ that is worth mentioning, namely wavelet-like bases on $[0,1]$. This still seems to be awfully restrictive. However, it will later be seen to be a key ingredient for the construction of wavelets on any domain that can be represented as a disjoint union of parametric images of cubes. This includes closed surfaces arising in connection with boundary integral equations (see Section 2.2).

Wavelets on the interval have been discussed in several papers; see, for instance Andersson, Hall, Jawerth and Peters (1994), Cohen, Daubechies and Vial (1993), Chui and Quak (1992) and Dahmen, Kunoth and Urban (1996*b*). The basic idea common to all these approaches is to construct multiresolution sequences $\mathcal{S}$ on $[0,1]$, which, up to local boundary effects, agree with the restriction of the stationary spaces defined on all of $\mathbb{R}$. Thus one retains possibly many translates $2^{j/2}\phi(2^j\cdot - k)$ whose support is strictly inside $(0,1)$. In addition, one takes fixed linear combinations of those translates interfering with the boundaries in such a way that the original order of polynomial exactness is preserved. The following discussion is based on Dahmen et al. (1996*b*), which differs somewhat from the other sources but seems to be tailored best to the needs of subsequent applications.

For any dual pair $(\phi,\tilde\phi)$ from the spline family (4.25), that is, $\phi = {}_d\phi$, $\tilde\phi = {}_{d,\tilde d}\tilde\phi$, $\tilde d \geq d$, $d+\tilde d$ even, define

$$\alpha^L_{j,m,r} := 2^{j/2}\langle (2^j\cdot)^r, 2^{j/2}\phi(2^j\cdot - m)\rangle_{\mathbb{R}} = \langle (\cdot)^r, \phi(\cdot - m)\rangle_{\mathbb{R}} =: \alpha_{m,r}, \quad (4.56)$$

and

$$\alpha^R_{j,m,r} := 2^{j/2}\langle (2^j(1-\cdot))^r, 2^{j/2}\phi(2^j\cdot - m)\rangle_{\mathbb{R}} = \langle (2^j - \cdot)^r, \phi(\cdot - m)\rangle_{\mathbb{R}}, \quad (4.57)$$

for $r = 0,\dots,\tilde d - 1$. Likewise $\tilde\alpha^L_{j,m,r}$, $\tilde\alpha^R_{j,m,r}$, $r = 0,\dots,d-1$, are defined by replacing ϕ by $\tilde\phi$. It is known that the support of $\tilde\phi$ always contains $\operatorname{supp}\phi$. It turns out that things depend somewhat on the parity $l(d) := d \bmod 2$. So fix $\tilde l \in \mathbb{N}$, such that for $j \geq j_0$, $\operatorname{supp}\tilde\phi(2^j\cdot - m) \subset (0,1)$, for $l \leq m \leq 2^j - \tilde l - l(d)$. Define left ($L$) and right ($R$) boundary functions by

$$\begin{aligned}
\tilde\phi^L_{j,\tilde l-\tilde d+r} &:= \sum_{m=-\tilde l_2+1}^{\tilde l-1} \alpha_{m,r}2^{j/2}\tilde\phi(2^j\cdot - m)\Big|_{[0,1]},\\
\tilde\phi^R_{j,\tilde l-2^j-l(d)+\tilde d-r} &:= \sum_{m=2^j-\tilde l-l(d)+1}^{2^j-\tilde l_1-1} \alpha^R_{j,m,r}2^{j/2}\tilde\phi(2^j\cdot - m)\Big|_{[0,1]}, \quad r = 0,\dots,\tilde d-1,
\end{aligned} \quad (4.58)$$

where $\operatorname{supp}\tilde\phi = [\tilde l_1, \tilde l_2]$. Since by (4.56), (4.57), the functions $\tilde\phi^L_{j,k}, \tilde\phi^R_{j,k}$ are simply truncations of the polynomial representations (4.32), it is easy to see

that the collections of left and right *boundary functions*

$$\begin{aligned}\tilde{\Phi}_j^L &= \left\{\tilde{\phi}_{j,\tilde{l}-\tilde{d}+r}^L : r = 0,\dots,\tilde{d}-1\right\},\\ \tilde{\Phi}_j^R &= \left\{\tilde{\phi}_{j,2^j-\tilde{l}-l(d)+\tilde{d}-r}^R : r = 0,\dots,\tilde{d}-1\right\}\end{aligned} \tag{4.59}$$

together with the *interior* translates

$$\tilde{\Phi}_j^I := \left\{2^{j/2}\tilde{\phi}(2^j\cdot - m) : m = \tilde{l},\dots,2^j - \tilde{l} - l(d)\right\}$$

span all polynomials $\Pi_{\tilde{d}}$ of degree $\leq \tilde{d}-1$ on $[0,1]$, that is,

$$\Pi_{\tilde{d}} \subseteq S\left(\tilde{\Phi}_j^L \cup \tilde{\Phi}_j^I \cup \tilde{\Phi}_j^R\right). \tag{4.60}$$

Setting

$$l := \tilde{l} - (\tilde{d} - d),$$

the functions $\phi_{j,k}^L$, $\phi_{j,k}^R$ are defined in exactly the same way with all tildes removed, providing

$$\Pi_d \subseteq S(\Phi_j^L \cup \Phi_j^I \cup \Phi_j^R). \tag{4.61}$$

Also, by construction,

$$\#\left(\Phi_j^L \cup \Phi_j^I \cup \Phi_j^R\right) = \#\left(\tilde{\Phi}_j^L \cup \tilde{\Phi}_j^I \cup \tilde{\Phi}_j^R\right).$$

However, while the interior functions in Φ_j^I, $\tilde{\Phi}_j^I$ are still biorthogonal, the boundary modifications have certainly destroyed biorthogonality of the elements in Φ_j^X, $\tilde{\Phi}_j^X$, $X \in \{L,R\}$. Nevertheless, it can be shown that these collections can always be *biorthogonalized*. Moreover, this is a completely local process, which need be done only once. In this and in several other respects it is very fortunate that things have been set up to exploit symmetry as much as possible. In fact, using the fact that ϕ and $\tilde{\phi}$ are symmetric around $l(d)/2$, one can show that

$$\phi_{j,2^j-l-l(d)+d-r}^R(1-x) = \phi_{j,l-d+r}^L(x), \quad r = 0,\dots,d-1, \tag{4.62}$$

and likewise for $\tilde{\phi}_{j,k}^R$, $\tilde{\phi}_{j,k}^L$. Thus one ends up with pairs of collections

$$\Phi_j = \{\phi_{j,k} : k \in \Delta_j\}, \quad \tilde{\Phi}_j = \{\tilde{\phi}_{j,k} : k \in \Delta_j\},$$

where $\Delta_j := \{l-d,\dots,2^j - l - l(d) + d\}$, with the following properties (Dahmen et al. 1996*b*).

(i) The functions in $\Phi_j, \tilde{\Phi}_j$ have small support, that is,

$$\operatorname{diam}(\operatorname{supp}\phi_{j,k}), \quad \operatorname{diam}(\operatorname{supp}\tilde{\phi}_{j,k}) \sim 2^{-j}. \tag{4.63}$$

(ii) The $\Phi_j, \tilde{\Phi}_j$ are biorthogonal

$$\langle \Phi_j, \tilde{\Phi}_j^{\mathrm{T}} \rangle = \mathbf{I}.$$

(iii) The spaces $S(\Phi_j)$, $S(\tilde{\Phi}_j)$ are exact of order $d, \tilde{d}$, respectively, that is,

$$\Pi_d \subset S(\Phi_j), \quad \Pi_{\tilde{d}} \subset S(\tilde{\Phi}_j). \tag{4.64}$$

(iv) The spaces $S(\Phi_j)$, $S(\tilde{\Phi}_j)$ are nested. This can be verified by exploiting trivial scaling properties of polynomials and refinability of the interior translates.

It is worth commenting on the structure of the corresponding refinement matrices in

$$\Phi_j^{\mathrm{T}} = \Phi_{j+1}^{\mathrm{T}} \mathbf{M}_{j,0}, \quad \tilde{\Phi}_j^{\mathrm{T}} = \tilde{\Phi}_{j+1}^{\mathrm{T}} \tilde{\mathbf{M}}_{j,0}. \tag{4.65}$$

Each $\mathbf{M}_{j,0}$, $\tilde{\mathbf{M}}_{j,0}$ consists of a stationary interior block, whose size grows like 2^j, as well as an upper left and lower right block, which are completely independent of j and of fixed size. The interior blocks are just finite sections of the bi-infinite refinement matrices $(a_{k-2m})_{k,m\in\mathbb{Z}}$, $(\tilde{a}_{k-2m})_{k,m\in\mathbb{Z}}$. Moreover, symmetry surfaces again. Denoting for a given matrix $\mathbf{M}$ by $\mathbf{M}^{\updownarrow}$ the matrix obtained from $\mathbf{M}$ by reversing the order of rows and columns, one can show (see also (4.62)) that

$$\mathbf{M}_{j,0}^{\updownarrow} = \mathbf{M}_{j,0}, \quad \tilde{\mathbf{M}}_{j,0}^{\updownarrow} = \tilde{\mathbf{M}}_{j,0}. \tag{4.66}$$

The next step, namely to construct corresponding biorthogonal bases, is somewhat more involved. Using tools from spline theory, one can first construct suitable *initial* stable completions. Then Proposition 3.8 can be applied providing new (sparse) stable completions $\mathbf{M}_{j,1}$, $\tilde{\mathbf{M}}_{j,1}$ of the above refinement matrices, which have completely analogous structure and satisfy

$$\mathbf{M}_{j,0}\tilde{\mathbf{M}}_{j,0}^{\mathrm{T}} + \mathbf{M}_{j,1}\tilde{\mathbf{M}}_{j,1}^{\mathrm{T}} = \mathbf{I}, \quad \tilde{\mathbf{M}}_{j,e}^{\mathrm{T}}\mathbf{M}_{j,e'} = \delta_{e,e'}\mathbf{I}, \quad e, e' \in \{0,1\}. \tag{4.67}$$

Thus the wavelet bases

$$\Psi_j^{\mathrm{T}} = \Phi_{j+1}^{\mathrm{T}} \mathbf{M}_{j,1}, \quad \tilde{\Psi}_j^{\mathrm{T}} := \tilde{\Phi}_{j+1}^{\mathrm{T}} \tilde{\mathbf{M}}_{j,1} \tag{4.68}$$

satisfy $\langle \Psi_j, \tilde{\Psi}_{j'} \rangle_{[0,1]} = \delta_{j,j'}\mathbf{I}$ and hence

$$\langle \Psi, \tilde{\Psi} \rangle_{[0,1]} = \mathbf{I}, \quad \Psi := \Phi_{j_0} \cup \bigcup_{j\geq j_0} \Psi_j, \quad \tilde{\Psi} := \tilde{\Phi}_{j_0} \cup \bigcup_{j\geq j_0} \tilde{\Psi}_j, \tag{4.69}$$

are biorthogonal.

All filters have finite length so that the $\psi_{j,k}, \tilde{\psi}_{j,k}$ also satisfy (4.63). The filters are stationary in the above sense. Thus the multiscale transformations $\mathbf{T}_J$ (3.26) and $\mathbf{T}_J^{-1}$ (3.28) are still efficient and require the order of N_J operations. Finally, observe that the techniques described in Section 4.2 still apply, since all operations ultimately reduce to restrictions of $\phi(2^j \cdot -k)$ to $[0,1]$ which can be realized by choosing $\chi_{j,k}$ as an additional factor in (4.43).

5. Norm equivalences and function spaces

One of the most important properties of wavelets is that they can be used to characterize function spaces; see, for example, (DeVore et al. 1992, Meyer 1990). The Riesz basis property (3.36) which came up in connection with the stability of multiscale transformations (Theorem 3.3) is a special case in a whole scale of similar relations. This will be seen to play a vital role for preconditioning, matrix compression and adaptive techniques (recall (b), (c) and (d) in Section 3.1).

In the classical stationary shift invariant or periodic setting such results are established by making heavy use of Fourier techniques. They no longer apply in a straightforward manner for other domains such as the interval or more complex cases such as closed surfaces yet to come. Recall from Section 3.4 that the Riesz basis property as one instance of such norm equivalences naturally leads to the concept of biorthogonal bases. When these bases correspond to *orthogonal* complements between successive spaces $S(\Phi_j)$, $S(\Phi_{j+1})$, the Riesz basis property reduces to the Pythagorean theorem, once the complement bases Ψ_j are uniformly stable relative to each level. However, orthogonal decompositions are often difficult to realize, lead to dense matrices $\mathbf{G}_j$, and in some cases are not optimal for the application at hand. Thus understanding the general class of biorthogonal multiscale bases is vital. However, while being *necessary*, biorthogonality by itself is *not* quite *sufficient* to imply the Riesz basis property (Meyer 1994). The developments in this section are therefore guided by the following point:

- find criteria for the validity of the Riesz basis property and other norm equivalences for biorthogonal bases, which can still be employed in situations where Fourier techniques no longer work.

A key ingredient is a pair of direct and inverse estimates; these are also known to play an important role in convergence theory of multigrid algorithms.

5.1. Direct and inverse estimates

The type of estimate we are aiming at is rooted in approximation theory, concerning approximation and regularity properties of the trial spaces. To formulate versions suitable for the present purpose, suppose that $\Omega \subseteq \mathbb{R}^n$ is an open connected domain (the case $\Omega = \mathbb{R}^n$ included). If Ω has a boundary we assume that it has some minimal regularity such as the uniform cone condition; see, for instance, DeVore and Sharpley (1993) and Johnen and Scherer (1977). Thus there exists an extension operator $E : L_p(\Omega) \to L_p(\mathbb{R}^n)$ that is bounded in $W_p^m(\Omega)$ for any $m \in \mathbb{N}$. The estimates we require are

$$\inf_{v \in S(\Phi_j)} \|f - v\|_{L_p(\Omega)} \lesssim 2^{-dj} \|f\|_{W_p^d(\Omega)}, \quad f \in W_p^d(\Omega). \tag{5.1}$$

We will refer to such estimates as *direct* or *Jackson* estimates. By interpolation, one derives from (5.1) a scale of similar estimates with the right-hand side replaced by $2^{-sj}\|f\|_{B_q^s(L_p(\Omega))}$, $s < m$, $1 \le q \le \infty$, where $B_q^s(L_p(\Omega))$ are corresponding Besov spaces (see Section 2.1).

There is often a counterpart called *inverse* or *Bernstein* estimate

$$\|v\|_{B_q^s(L_p(\Omega))} \lesssim 2^{sj}\|v\|_{L_p(\Omega)}, \quad v \in S(\Phi_j). \tag{5.2}$$

We next give a simple criterion for verifying (5.1) that will apply in all cases of interest.

Proposition 5.1 Let $\Phi_j \subset L_p(\Omega)$ and $\Xi_j \subset L_{p'}(\Omega)$ with $\frac{1}{p} + \frac{1}{p'} = 1$ have the following properties:

(i) Ξ_j and Φ_j are biorthogonal,

$$\langle \Phi_j, \Xi_j \rangle_\Omega = \mathbf{I} \tag{5.3}$$

where $\langle \cdot, \cdot \rangle_\Omega$ denotes the dual pairing for $L_p(\Omega) \times L_{p'}(\Omega)$.

(ii) The elements of Φ_j and Ξ_j are uniformly bounded, that is,

$$\|\phi_{j,k}\|_{L_p}, \|\xi_{j,k}\|_{L_{p'}(\Omega)} = O(1), \quad j \in \mathbb{N}, \quad k \in \Delta_j. \tag{5.4}$$

(iii) The collections Φ_j, Ξ_j are locally finite, that is, there exists a constant $C < \infty$ such that

$$\#\{k' : \square_{j,k} \cap \square_{j,k'} \neq \emptyset\} \le C, \quad \operatorname{diam} \square_{j,k} \lesssim 2^{-j} \tag{5.5}$$

where $\square_{j,k}$ is the smallest cube containing $\operatorname{supp} \phi_{j,k}$ and $\operatorname{supp} \xi_{j,k}$.

(iv) The spaces $S(\Phi_j)$ contain all polynomials of order d (degree $\le d-1$) on Ω,

$$\Pi_d \subseteq S(\Phi_j). \tag{5.6}$$

Then one has

$$\|f - \langle f, \Xi_j \rangle_\Omega \Phi_j\|_{L_p(\Omega)} \lesssim 2^{-dj}\|f\|_{W_p^d(\Omega)}. \tag{5.7}$$

The type of argument needed here is essentially folklore. Since it plays a central role we sketch a proof. By (5.6), one has for any $P \in \Pi_d$

$$\|f - \langle f, \Xi_j \rangle_\Omega\|_{L_p(\square_{j,k})}^p \lesssim \|f - P\|_{L_p(\square_{j,k})}^p + \sum_{\square_{j,k'} \cap \square_{j,k} \neq \emptyset} |\langle f - P, \xi_{j,k'} \rangle_\Omega|^p \|\phi_{j,k'}\|_{L_p(\Omega)}^p. \tag{5.8}$$

On account of (5.5), the sum involves a uniformly bounded number of summands. Using (5.4) gives

$$|\langle f - P, \xi_{j,k'} \rangle_\Omega|^p \|\phi_{j,k'}\|_{L_p(\Omega)}^p \lesssim \|f - P\|_{L_p(\square_{j,k'})}^p \lesssim 2^{-dpj}\|f\|_{W_p^d(\square_{j,k'})}^p, \tag{5.9}$$

where a Bramble–Hilbert-type argument has been used in the last step. A little care has to be taken near the boundary. In order to employ the

scaling argument needed for the Bramble–Hilbert argument, one can employ extension techniques; see Oswald (1997) for details. Bearing (5.5) in mind, and summing over $k \in \Delta_j$, yields (5.7). □

Estimates of the type (5.1) readily lead to estimates without *regularity* assumptions. Consider the *K-functional* (see, for instance, Bergh and Löfström (1976))

$$K_d(f,t) = K(f,t,L_p,W_p^d) := \inf_{g \in W_p^d(\Omega)} \left\{ \|f-g\|_{L_p(\Omega)} + t^d \|g\|_{W_p^d(\Omega)} \right\}. \quad (5.10)$$

One immediately infers from (5.1) that

$$\inf_{v \in S_j} \|f-v\|_{L_p(\Omega)} \lesssim K_d(f, 2^{-j}). \quad (5.11)$$

Remark 5.2 Under assumptions (5.4) and (5.5), the Φ_j are uniformly stable (relative to $\|\cdot\|_{L_p(\Omega)}$ and $\|\cdot\|_{\ell_p(\Delta_j)}$).

Remark 5.3 Obviously Proposition 5.1 applies to all the above examples of biorthogonal multiresolution sequences (with $\Xi_j = \tilde{\Phi}_j$). In fact, for wavelets on $\mathbb{R}^n$ recall (4.32), the multiresolution on $[0,1]$ was constructed in Section 4.4 so that (5.6) holds, while all other conditions are obviously satisfied. Thus we will assume from now on that the direct estimate (5.1) is valid for the order d of polynomial exactness.

Remark 5.4 Suppose $\phi \in L_2(\mathbb{R}^n)$ is a (compactly supported stable) scaling function. Let

$$\gamma := \sup\{s : \phi \in H^s(\mathbb{R}^n)\}.$$

Then

$$\|v\|_{H^s(R^n)} \lesssim 2^{sj} \|v\|_{L_2(\mathbb{R}^n)}, \quad v \in S(\Phi_j), \quad (5.12)$$

holds for any $s < \gamma$. It is also known that $\phi \in L_2(\mathbb{R}^n)$ implies $\phi \in H^s(\mathbb{R}^n)$ for some $s > 0$ (Villemoes 1993).

For a proof see, for instance, Dahmen (1995). One can show that when $\phi, \tilde{\phi}$ is a dual pair of compactly supported generators, then their integer shifts are *locally linearly independent*. Then $\|\cdot\|_{H^s(\square)}$ and $\|\cdot\|_{L_2(\square)}$ are equivalent norms on $S(\Phi_0)$ and the claim for integer s follows from summing the local norms and rescaling.

Remark 5.5 (Dahmen 1995, Dahmen et al. 1996*b*) Let $\Phi_j \subset L_2([0,1])$ denote the generator bases constructed in Section 4.4. Then

$$\|v\|_{H^s([0,1])} \lesssim 2^{sj} \|v\|_{L_2([0,1])}, \quad v \in S(\Phi_j), \quad s < \gamma. \quad (5.13)$$

Finally, the inverse inequalities can be expressed in terms of the K-functional as well. In fact, from (5.2), one can deduce that

$$K_d(v,t) \lesssim (\min\{1, t2^j\})^\gamma \|v\|_{L_p(\Omega)}, \quad v \in S(\Phi_j). \quad (5.14)$$

The form (5.10) and (5.14) of the direct and inverse estimates will guide the subsequent discussion.

5.2. The Riesz basis property

As pointed out in Section 3.4, the stability of multiscale transformations (3.26), (3.28) is equivalent to the Riesz basis property of the basis Ψ (and $\tilde{\Psi}$). It turns out that sufficient conditions which apply in our cases of interest can be formulated in a general Hilbert space setting. This will shed some light on the essential mechanisms. An important point is that, as one will see, once a biorthogonal pair $\Psi, \tilde{\Psi}$ is given, additional conditions implying the Riesz basis property *only* concern properties of the *spaces* spanned by subsets of Ψ and $\tilde{\Psi}$, *not* of the particular bases. These properties can be formulated in terms of the estimates (5.10), (5.14).

In order to stress this point, we will first reformulate the problem somewhat, which, by the way, corresponds also to the strategy of constructing Riesz bases employed in Section 4.4. First of all, it is usually not so difficult to assure stability of a complement basis $\Psi_j = \{\psi_{j,k} : k \in \nabla_j\}$ in the space Φ_{j+1}. We will therefore assume in the following that

$$\|\mathbf{d}^T \Psi_j\|_H \sim \|\mathbf{d}\|_{\ell_2(\nabla_j)}. \tag{5.15}$$

Moreover, recall from (3.45) that biorthogonality is equivalent to

$$Q_j Q_l = Q_j \quad \text{for} \quad j \le l, \tag{5.16}$$

where the Q_j are the projectors $Q_j v = \langle v, \tilde{\Phi}_j \rangle \Phi_j = \langle v, \tilde{\Psi}^j \rangle \Psi^j$ of (3.44), (3.46), which, by Remark 3.6, have to be uniformly bounded when Ψ and $\tilde{\Psi}$ are Riesz bases. Then, by (5.15), the norm equivalence (3.36) can be equivalently expressed as

$$\|f\|_H \sim N_{\mathcal{Q}}(f) \sim N_{\mathcal{Q}^*}(f), \tag{5.17}$$

where, for $Q_{-1} := 0$,

$$N_{\mathcal{Q}}(f)^2 := \sum_{j=0}^{\infty} \|(Q_j - Q_{j-1}) f\|_H^2. \tag{5.18}$$

The objective now is to establish the validity of (5.18) for a given sequence $\mathcal{Q}$ of projectors satisfying the necessary conditions of uniform boundedness and (5.16). It is important to note that in this form the result applies when the Q_j are given only in the form $Q_j v = \langle v, \tilde{\Phi}_j \rangle \Phi_j$, that is, without explicit knowledge of the *right* complement bases Ψ_j yet. Note also that the condition (5.16) implies that the ranges $\tilde{S}_j$ of the adjoints Q_j^* are also *nested*. Moreover, these spaces are also dense in H (Dahmen 1994, 1996). Let us denote the corresponding sequence by $\tilde{\mathcal{S}}$. The following result says that the Riesz basis property holds when, in addition to biorthogonality, the primal

and dual multiresolution sequences $\mathcal{S}, \tilde{\mathcal{S}}$ both have some approximation and regularity properties expressed in terms of pairs of direct and inverse estimates (Dahmen 1996).

Theorem 5.6 Let $\mathcal{S}$ be an ascending dense sequence of closed subspaces of H and let $\mathcal{Q}$ be a sequence of uniformly H-bounded projectors with ranges $\mathcal{S}$ satisfying (5.16). Let $\tilde{\mathcal{S}}$ be the ranges of the adjoint sequence $\mathcal{Q}^*$. Suppose there exists a family of uniformly bounded subadditive functionals $\omega(\cdot,t) : H \to \mathbb{R}_+$, $t > 0$, such that $\lim_{t\to 0^+} \omega(f,t) = 0$ for each $f \in H$ and that the pair of estimates

$$\inf_{v\in V_j} \|f - v\|_H \lesssim \omega(f, 2^{-j}), \tag{5.19}$$

and

$$\omega(v_j, t) \lesssim (\min\{1, t2^j\})^\gamma \|v_j\|_H, \quad v_j \in V_j, \tag{5.20}$$

holds for $\mathcal{V} = \mathcal{S}$ and $\mathcal{V} = \tilde{\mathcal{S}}$ with some $\gamma, \tilde{\gamma} > 0$, respectively. Then

$$\|\cdot\|_H \sim N_{\mathcal{Q}}(\cdot) \sim N_{\mathcal{Q}^*}(\cdot). \tag{5.21}$$

Here is an immediate consequence of Theorem 5.6.

Remark 5.7 Note that the K-functional $K_d(\cdot,t)$ defined by (5.10) has, by (5.11), (5.14),(5.13), (5.12), all the properties of $\omega(\cdot,t)$ required above. Thus the biorthogonal bases constructed in Sections 4.2 and 4.4 are indeed Riesz bases.

A few comments on the proof of Theorem 5.6 are in order; see Dahmen (1996) for details. First one observes that (Cohen 1994)

$$N_{\mathcal{Q}}(\cdot) \lesssim \|\cdot\|_H \quad \text{if and only if} \quad \|\cdot\|_H \lesssim N_{\mathcal{Q}^*}(\cdot).$$

Thus it suffices to prove that

$$\|\cdot\|_H \lesssim N_{\mathcal{Q}}(\cdot) \quad \text{and} \quad \|\cdot\|_H \lesssim N_{\mathcal{Q}^*}(\cdot), \tag{5.22}$$

or the corresponding pair of opposite inequalities. To prove estimates of the form (5.22) one can employ a technique which is also familiar in the analysis of multilevel preconditioners.

Strengthened Cauchy inequalities

To this end, suppose there is a (dense) subspace $U \subset H$ with a (stronger) norm $\|\cdot\|_U$ such that, for some $\varepsilon > 0$,

$$\|f - Q_j f\|_{U^*} \lesssim 2^{-j\varepsilon}\|f\|_H, \quad \|f - Q_j f\|_H \lesssim 2^{-j\varepsilon}\|f\|_U, \tag{5.23}$$

and

$$\|v_j\|_U \lesssim 2^{j\varepsilon}\|v_j\|_H, \quad \|v_j\|_H \lesssim 2^{j\varepsilon}\|v_j\|_{U^*}, \quad v_j \in S_j. \tag{5.24}$$

Then one can estimate

$$\langle (Q_j - Q_{j-1})f, (Q_i - Q_{i-1})f\rangle \leq \begin{cases} \|(Q_j - Q_{j-1})f\|_{U^*}\|(Q_i - Q_{i-1})f\|_U & \text{if } i \leq j, \\ \|(Q_j - Q_{j-1})f\|_U \ \|(Q_i - Q_{i-1})f\|_{U^*} & \text{if } i > j. \end{cases}$$

Thus by (5.23), (5.24),

$$\begin{aligned} \|f\|_H^2 &= \sum_{i,j=0}^{\infty} \langle (Q_j - Q_{j-1})f, (Q_i - Q_{i-1})f\rangle \\ &\lesssim \sum_{i,j=0}^{\infty} 2^{-\varepsilon|i-j|}\|(Q_j - Q_{j-1})f\|_H\|(Q_i - Q_{i-1})f\|_H \lesssim N_{\mathcal{Q}}(f)^2. \end{aligned}$$

When $\mathcal{Q}$ is uniformly bounded on U^* the estimates (5.23) and (5.24) can be shown to hold by duality also for $\mathcal{Q}^*$. Thus the same argument also yields $\|f\|_H^2 \lesssim N_{\mathcal{Q}^*}(f)^2$.

A scale of interpolation spaces

So, it remains to find such a subspace U. Natural candidates are the spaces $\mathcal{A}_{\mathcal{Q}}^s$ which are defined for $s > 0$ as the collection of those $f \in H$ for which

$$\|f\|_{\mathcal{A}_{\mathcal{Q}}^s}^s = \sum_{j=0}^{\infty} 2^{2sj}\|(Q_j - Q_{j-1})f\|_H^2 < \infty.$$

They are dense reflexive subspaces of H, and, with a proper understanding of continuously extended projectors, one has a representation of their duals in terms of the dual projectors $\mathcal{Q}^*$ (Dahmen 1996)

$$(\mathcal{A}_{\mathcal{Q}}^s)^* = \mathcal{A}_{\mathcal{Q}^*}^{-s}. \tag{5.25}$$

Moreover, these spaces are defined so that, again under assumption (5.16), a pair of direct and inverse inequalities hold, namely

$$\|f - Q_j f\|_H \lesssim 2^{-js}\|f\|_{\mathcal{A}_{\mathcal{Q}}^s}, \quad \|f - Q_j f\|_{(\mathcal{A}_{\mathcal{Q}^*}^s)^*} \lesssim 2^{-js}\|f\|_H, \tag{5.26}$$

and

$$\|v_j\|_{\mathcal{A}_{\mathcal{Q}}^s} \lesssim 2^{js}\|v_j\|_H, \quad \|v_j\|_H \lesssim 2^{js}\|v_j\|_{(\mathcal{A}_{\mathcal{Q}^*}^s)^*}. \tag{5.27}$$

Finally, by (5.16), $\mathcal{Q}$ is trivially uniformly bounded on $\mathcal{A}_{\mathcal{Q}}^s$ for all s, that is,

$$\|Q_j\|_{\mathcal{A}_{\mathcal{Q}}^s} = 1, \quad j \in \mathbb{N}_0, s \in \mathbb{R}. \tag{5.28}$$

Thus one *almost* has the pair of inequalities (5.23), (5.24) without any assumption on $\mathcal{Q}$ beyond (5.16) and uniform boundedness. What is missing is the relation between the spaces $(\mathcal{A}_{\mathcal{Q}^*}^s)^*$ and $(\mathcal{A}_{\mathcal{Q}}^s)^*$. If they were equivalent, (5.26) and (5.27), together with the strengthened Cauchy inequality

argument, would confirm (5.22) and hence the claim of Theorem 5.6. This is where the direct and inverse inequalities (5.19), (5.20) come into play. In fact, with the aid of these inequalities one can prove that

$$\|\cdot\|_{\mathcal{A}^s_{\mathcal{Q}}} \sim \|\cdot\|_{\mathcal{B}^s_\omega} \sim \|\cdot\|_{\mathcal{A}^s_{\mathcal{Q}^*}} \quad \text{for } 0 < s < \min\{\gamma, \tilde{\gamma}\}, \tag{5.29}$$

where

$$\|f\|^s_{\mathcal{B}^s_\omega} := \|f\|^2_H + \sum_{j=0}^{\infty} 2^{2sj}\omega(f, 2^{-j})^2, \tag{5.30}$$

which closes the gap.

These results are closely related to interpolation theory. In fact, the $\mathcal{A}^s_{\mathcal{Q}}$ are *interpolation spaces* obtained by the *real method*; see, for instance, Bergh and Löfström (1976), DeVore and Popov (1988*a*), DeVore and Sharpley (1993) and Peetre (1978). A detailed discussion of this point of view can be found in Dahmen (1995).

As mentioned before, the role of $\omega(\cdot, t)$ is typically played by a K-functional or a modulus of smoothness, which under our assumptions on the underlying domain are equivalent seminorms (Johnen and Scherer 1977). In that sense the spaces $\mathcal{B}^s_\omega$ can be viewed as generalized *Besov* spaces. Thus, in addition to the Riesz basis property, the above criteria *automatically* establish norm equivalences for a whole *scale* of spaces. The equivalence of the artificial spaces $\mathcal{A}^s_{\mathcal{Q}}$ with the Besov-type spaces $\mathcal{B}^s_\omega$ in some range of s immediately yields norm equivalences for these (classical) function spaces, which will be addressed next.

5.3. Characterization of Sobolev spaces

When Ω is a domain in $\mathbb{R}^n$ as above and $\omega(\cdot, t)$ is an L_2 modulus of smoothness

$$\omega(f, t) = \omega_d(f, t)_{L_2(\Omega)} := \sup_{|h| \le t} \|\Delta^d_h f\|_{L_2(\Omega_{d,h})}$$

where $\Delta^d_h = \Delta_h \Delta^{d-1}_h$, $\Delta_h f = f(\cdot + h) - f(\cdot)$ and $\Omega_{d,h} = \{x : x + lh \in \Omega, l = 0, \ldots, d\}$, or when $\omega(\cdot, t)$ is the K-functional from (5.10), the norm in (5.30) is equivalent to $\|\cdot\|_{H^s(\Omega)}$ for $0 < s < d$, and

$$H^s(\Omega) \sim B^s_2(L_2(\Omega)). \tag{5.31}$$

We will now apply the above results for $H = H^s(\Omega)$. For simplicity we focus on $H^0(\Omega) = L_2(\Omega)$. Moreover, let us denote for $s > 0$ by H^s some closed subspace of $H^s(\Omega)$ (or $H^s(\Omega)$ itself) which is, for instance, determined by some homogeneous boundary conditions. The key role is again played by a pair of direct and inverse inequalities

$$\inf_{v_j \in V_j} \|v - v_j\|_{L_2(\Omega)} \lesssim 2^{-sj}\|v\|_{H^s(\Omega)}, \quad v \in H^s, \ 0 \le s \le d_V, \tag{5.32}$$

and

$$\|v_j\|_{H^s(\Omega)} \lesssim 2^{sj}\|v_j\|_{L_2(\Omega)}, \quad v_j \in V_j, \quad s < \gamma_{\mathcal{V}}. \tag{5.33}$$

Recall from Section 5.1 under which circumstances such inequalities hold. The direct inequality may be affected by homogeneous boundary conditions incorporated in $\mathcal{S}$. If this is done properly the argument stays essentially the same, since near the boundary not all polynomials are needed.

From (5.29), (5.28) and duality (5.25) one infers the following fact.

Theorem 5.8 Let $\mathcal{Q}$ be uniformly bounded with range $\mathcal{S}$ and suppose that (5.16) holds. Moreover, assume that $\mathcal{S}$ and the range $\tilde{\mathcal{S}}$ of $\mathcal{Q}^*$ satisfy (5.32) and (5.33) for some $d := d_{\mathcal{S}}, \tilde{d} := d_{\tilde{\mathcal{S}}}$, $0 < \gamma := \min\{\gamma_{\mathcal{S}}, d\}$ and $0 < \tilde{\gamma} := \min\{\gamma_{\tilde{\mathcal{S}}}, \tilde{d}\}$, respectively. Then

$$\left(\sum_{j=0}^{\infty} 2^{2sj}\|(Q_j - Q_{j-1})f\|^2_{L_2(\Omega)}\right)^{1/2} \sim \|f\|_{H^s(\Omega)}, \quad s \in (-\tilde{\gamma}, \gamma), \tag{5.34}$$

where it is to be understood that $H^s(\Omega) = (H^{-s}(\Omega))^*$ for $s < 0$. Moreover, $\mathcal{Q}$ is uniformly bounded in H^s for that range

$$\|Q_j v\|_{H^s(\Omega)} \lesssim \|v\|_{H^s(\Omega)}, \quad v \in H^s(\Omega). \tag{5.35}$$

When $H^s = H^s(\Omega)$ and both sequences $\mathcal{S}, \tilde{\mathcal{S}}$ have a high order of exactness $d, \tilde{d}$, respectively, the above range may have a significant part for $s < 0$. There is, however, always some $\tilde{\gamma} > 0$ reaching into the negative range. It could be small if H^s is a true subspace of $H^s(\Omega)$ and the corresponding boundary conditions are incorporated in $\mathcal{S}$. What matters, though, is that, by (5.25), (5.31) and the above result applied to $s \geq 0$, one still has

$$\left(\sum_{j=0}^{\infty} 2^{-2sj}\|(Q_j^* - Q_{j-1}^*)f\|^2_{L_2(\Omega)}\right)^{1/2} \sim \|f\|_{H^{-s}}, \quad s \in [0, \gamma). \tag{5.36}$$

It is convenient to express these relations in terms of the operators

$$\Sigma_s f := \sum_{j=0}^{\infty} 2^{js}(Q_j - Q_{j-1})f, \tag{5.37}$$

which act as a *shift* in the Sobolev scale

$$\|\Sigma_s f\|_{H^t(\Omega)} \sim \|f\|_{H^{t+s}(\Omega)}, \quad t + s \in (-\tilde{\gamma}, \gamma), \tag{5.38}$$

just like classical Bessel potential operators in harmonic analysis. Due to (5.16), one has

$$\Sigma_s^{-1} = \Sigma_{-s}, \quad \Sigma_s^* = \sum_{j=0}^{\infty} 2^{js}\left(Q_j^* - Q_{j-1}^*\right). \tag{5.39}$$

It is important to note that one-sided estimates of type (5.34) hold for a wider range of s. In fact, the uniform boundedness of the Q_j ensures that

$$\|Q_j f - f\|_{L_2(\Omega)} \lesssim \inf_{v \in S(\Phi_j)} \|f - v\|_{L_2(\Omega)}.$$

Thus by Proposition 5.1 and (5.11) one obtains

$$\|(Q_j - Q_{j-1})f\|_{L_2(\Omega)} \lesssim K\left(f, 2^{-j+1}, L_2, H^d\right).$$

Since

$$\left(\|f\|^2_{L_2(\Omega)} + \sum_{J=0}^{\infty} 2^{2sj} K\left(f, 2^{-j+1}, L_2, H^d\right)^2\right)^{1/2}$$

is known to be a norm for the Besov space $B^s_2(L_2(\Omega)) = H^s(\Omega)$, one obtains, for instance,

$$\sum_{j=0}^{\infty} 2^{2js} \|(Q_j - Q_{j-1})f\|^2_{L_2(\Omega)} \lesssim \|f\|^2_{H^s(\Omega)}, \quad -\tilde{\gamma} < s < d. \tag{5.40}$$

If corresponding wavelet bases $\Psi, \tilde{\Psi}$ are known, Σ_s can be written as

$$\Sigma_s f = \sum_{\lambda \in \nabla} 2^{s|\lambda|} \langle f, \tilde{\psi}_\lambda \rangle_\Omega \psi_\lambda, \tag{5.41}$$

and (5.34) becomes

$$\|f\|_{H^s} \sim \|\mathbf{D}^s \langle f, \tilde{\Psi} \rangle^{\mathrm{T}}_\Omega\|_{\ell_2(\nabla)}, \quad s \in (-\tilde{\gamma}, \gamma), \tag{5.42}$$

where $\mathbf{D}^s$ denotes the diagonal matrix

$$(\mathbf{D}^s)_{\lambda,\lambda'} = 2^{s|\lambda|} \delta_{\lambda,\lambda'}. \tag{5.43}$$

Remark 5.9 In view of Proposition 5.1, Remark 5.4, Remark 5.5, Theorem 5.8 implies that the wavelet bases constructed in (4.23) for $L_2(\mathbb{R})$, in (4.53) for the periodic case and (4.69) for the interval $[0,1]$ all satisfy (5.42) for $s \in (-\tilde{\gamma}, \gamma)$.

Frames

It is important to note that norm equivalences of the type (5.34) for $s > 0$ do *not* require knowlege of concrete bases for decompositions $(Q_j - Q_{j-1})S_j$. Instead one can prove that (Dahmen 1995, Oswald 1994, Oswald 1992, Oswald 1990)

$$\|f\|^2_{H^s(\Omega)} \sim \inf\left\{\sum_{j=0}^{\infty} 2^{2sj} \|f_j\|^2_{L_2(\Omega)} : f = \sum_{j=0}^{\infty} f_j\right\}. \tag{5.44}$$

In terms of interpolation theory, norms of this type correspond to the J-method (Bergh and Löfström 1976, Peetre 1978). Such norm equivalences will play a crucial role in preconditioning.

These results have further natural extensions to other (reflexive) Banach spaces like L_p-spaces ($1 < p < \infty$). Interpolation between $L_p(\Omega)$ and $W_p^m(\Omega)$, say, leads to Besov spaces $B_q^s(L_p(\Omega))$ endowed with the norms

$$\|f\|^q_{B^s_q(L_p(\Omega))} = \|f\|^q_{L_p(\Omega)} + \sum_{j=0}^{\infty} 2^{qsj} K\left(f, 2^{-j}, L_p, W_p^d\right)^q \tag{5.45}$$

for $d \geq s > 0$. Assuming that Ψ, $\tilde{\Psi}$ are biorthogonal wavelet bases (in $L_2(\Omega)$) one still obtains norm equivalences of the form

$$\|f\|_{B^s_q(L_p(\Omega))} \sim \left(\|\langle f, \tilde{\Phi}_0\rangle_\Omega\|^q_{\ell_p(\Delta_{j_0})} + \sum_{j=j_0}^{\infty} 2^{jq(s+\frac{n}{2}-\frac{n}{p})} \|\langle f, \tilde{\Psi}_j\rangle_\Omega\|^q_{\ell_p(\nabla_j)} \right)^{1/q} \tag{5.46}$$

which, of course, reduce to (5.42) for $s > 0$, $p = q = 2$. These norm equivalences play an important role in *nonlinear approximation* (DeVore and Lucier 1992, DeVore et al. 1992). This, in turn, will be of interest in connection with adaptive schemes (see Section 11).

6. Preconditioning

This section is only concerned with preconditioning systems arising from discretizations of operator equations, which in a loose sense may be termed *elliptic.* In particular, all the examples in Section 2 are covered (see also (b) in Section 1.5). I would like to stress the following points.

- Once the norm equivalences discussed in Section 5.3 are available, the principal argument is rather simple and applies to a relatively wide range of cases, represented by the reference problem in Section 2.3. To bring out the basic mechanism, I will address it first in this generality, which will cover various special cases treated in the literature.
- The strongest interrelation between rather independent developments in the area of wavelets on one hand and finite element discretizations on the other hand occurs in connection with preconditioning. Since these developments usually ignore each other, I will comment on both. In view of the existing excellent treatments of *multilevel subspace correction methods* seen through the finite element eye, the main focus here will be on the wavelet or basis oriented point of view.
- In the present generality the results are purely asymptotical. The actual performance of corresponding schemes depends very much on the concrete case at hand. In general, it is hard to say which concept is best able do cope with near degeneracies or strong isotropies.

6.1. Discretization and projection methods

Consider problem (2.24) for the spaces H^t, H^{-t}, L_2 as described in Section 2.3. Throughout the following we will assume that $\Psi = \{\psi_\lambda : \lambda \in \nabla\}$, $\tilde{\Psi} = \{\tilde{\psi}_\lambda : \lambda \in \nabla\}$, with $\nabla = \Delta_+ \cup \nabla_-$, are biorthogonal wavelet bases in L_2 such that the norm equivalences

$$\|f\|_{H^s} \sim \|\mathbf{D}^s \langle f, \tilde{\Psi}\rangle^{\mathrm{T}}\|_{\ell_2(\nabla)}, \quad s \in (-\tilde{\gamma}, \gamma), \tag{6.1}$$

hold.

The numerical schemes to be used for the solution of (2.24) may be viewed as *generalized Petrov–Galerkin schemes.* To describe this, we adhere to notation (3.43) and suppose that Θ_Λ is a collection of functionals that is defined and total over $S(\mathcal{L}\Psi_\Lambda)$. To solve (2.24), the objective is to determine $u_\Lambda \in S(\Psi_\Lambda)$ such that

$$\langle \mathcal{L}u_\Lambda, \Theta_\Lambda\rangle = \langle f, \Theta_\Lambda\rangle. \tag{6.2}$$

Of course, $\Theta = \Psi$ gives rise to a classical Galerkin scheme, while collocation is obtained when Θ involves Dirac functionals. In the latter case the right-hand side has to be taken from a sufficiently smooth space. This is appropriate when $\mathcal{L}$ is also known to be boundedly invertible as an operator from H^s into H^{s-2t} for some larger $s \in \mathbb{R}$. To explain what is meant by *stability* of the scheme, it is convenient to reinterpret (6.2) as a *projection method.* Suppose that $\tilde{\Theta}$ is a sufficiently regular dual set for Θ and let $P_\Lambda := \langle \cdot, \Theta_\Lambda\rangle \tilde{\Theta}_\Lambda$ be an associated projector. Then (6.2) is equivalent to

$$P_\Lambda \mathcal{L} Q_\Lambda u = P_\Lambda f. \tag{6.3}$$

The scheme (6.2) is said to be $(s, 2t)$-*stable* if for $\#\Lambda$ large

$$\|P_\Lambda \mathcal{L} v\|_{H^{s-2t}} \sim \|v\|_{H^s}, \quad v \in S(\Psi_\Lambda), \tag{6.4}$$

that is, the finite-dimensional operators $\mathcal{L}_\Lambda := P_\Lambda \mathcal{L} Q_\Lambda$ are uniformly bounded invertible mappings from $H^s \cap S(\Psi_\Lambda)$ onto $H^{s-2t} \cap S(\tilde{\Theta}_\Lambda)$. In terms of linear systems, substituting $u_\Lambda = \mathbf{d}^T \Psi_\Lambda$ into (6.2) yields the linear system

$$\mathbf{d}^T \langle \mathcal{L}\Psi_\Lambda, \Theta_\Lambda\rangle = \langle f, \Theta_\Lambda\rangle. \tag{6.5}$$

In particular, for the Galerkin case, (6.3) becomes

$$Q_\Lambda^* \mathcal{L} Q_\Lambda u_\Lambda = Q_\Lambda^* f. \tag{6.6}$$

The most important case for the subsequent discussion is $(t, 2t)$-stability, in brief *stability*, which then means

$$\|Q_\Lambda^* \mathcal{L} v\|_{H^{-t}} \sim \|v\|_{H^t}, \quad u \in S_\Lambda. \tag{6.7}$$

In general not much is known about stability for the above general class of Petrov–Galerkin schemes. For (nonconstant coefficient) pseudo-differential operators on the torus, stability conditions are established in Dahmen,

Prößdorf and Schneider (1994*c*); see also Dahmen, Kleemann, Prößdorf and Schneider (1996*a*) for an application to collocation.

When $\mathcal{L}$ is a pseudo-differential operator, its injectivity, boundedness and coercivity of the principal part of its symbol also imply stability (6.7) of the Galerkin scheme (Dahmen et al. 1994*c*, Dahmen, Prößdorf and Schneider 1994*b*, Hildebrandt and Wienholtz 1964). Of course, when $\mathcal{L}$ is *selfadjoint* in the sense that

$$a(u,v) := \langle \mathcal{L}u, v\rangle \tag{6.8}$$

is a *symmetric bilinear form*, ellipticity (2.23) means that

$$\|\cdot\|^2 := a(\cdot,\cdot)^2 \sim \|\cdot\|_{H^t}, \tag{6.9}$$

and the Galerkin scheme is trivially stable.

We think of the trial spaces having large dimension so that direct solvers based on factorization techniques are prohibitively expensive in storage and computing time. On the other hand, in the symmetric case (6.9), for instance, the speed of convergence of *iterative methods* is known to be governed by the condition numbers

$$\kappa_2(\mathcal{L}_\Lambda) := \lambda_{\max}(\mathcal{L}_\Lambda)/\lambda_{\min}(\mathcal{L}_\Lambda), \tag{6.10}$$

where

$$\lambda_{\max}(\mathcal{L}_\Lambda) := \sup_{v\in S(\Psi_\Lambda)} \frac{\langle \mathcal{L}v, v\rangle}{\langle v, v\rangle}, \quad \lambda_{\min}(\mathcal{L}_\Lambda) := \inf_{v\in S(\Psi_\Lambda)} \frac{\langle \mathcal{L}v, v\rangle}{\langle v, v\rangle}. \tag{6.11}$$

Note that when $t \neq 0$, the condition numbers grow with increasing $\#\Lambda$. In fact, on account of the norm equivalence (6.1) and (6.9), one obtains

$$\lambda_{\min}(\mathcal{L}_\Lambda) \leq \langle \mathcal{L}\psi_\lambda, \psi_\lambda\rangle / \|\psi_\lambda\|_{L_2} \sim 2^{2t|\lambda|},$$

while

$$\lambda_{\max}(\mathcal{L}_\Lambda) \geq \langle \mathcal{L}\psi_\lambda, \psi_\lambda\rangle / \|\psi_\lambda\|_{L_2} \sim 2^{2t|\lambda|}.$$

Thus choosing $|\lambda|$ as the lowest or highest level in Λ, depending on the sign of t, it is clear that

$$\kappa_2(\mathcal{L}_\Lambda) \gtrsim 2^{2|t||\Lambda|}, \tag{6.12}$$

where $|\Lambda| := \max\{|\lambda| - |\lambda'| : \lambda, \lambda' \in \Lambda\}$.

Thus, in such cases the objective is to find a symmetric positive definite operator $\mathcal{C}_\Lambda$ such that $\kappa_2(\mathcal{C}_\Lambda \mathcal{L}_\Lambda)$ remains possibly uniformly bounded, so that schemes like

$$u_\Lambda^{l+1} := u_\Lambda^l + \mathcal{C}_\Lambda(\mathcal{L}_\Lambda u_\Lambda^l - f),$$

or, better, correspondingly preconditioned conjugate gradient iterations, would converge rapidly.

6.2. An application of norm equivalences

With the results of Section 5.3 at hand, the task of preconditioning has become relatively easy. Since, under the assumption (2.23), $\mathcal{L}$ acts as a shift in the Sobolev scale, it is reasonable to exploit the fact that Σ_s from (5.37) does that too. Hence Σ_s should have the capability of undoing the effect of $\mathcal{L}$. By the previous remarks, the stiffness matrix

$$\mathbf{A}_\Lambda := \langle \mathcal{L}\Psi_\Lambda, \Psi_\Lambda \rangle^{\mathrm{T}} \tag{6.13}$$

relative to the wavelet basis Ψ_Λ is ill conditioned for $t \neq 0$ and large $\#\Lambda$. However, a diagonal symmetric scaling suffices to remedy this. This observation has been made on various different levels of generality in several papers; see, for instance Beylkin (1993), Dahmen and Kunoth (1992), Dahmen et al. (1996*c*), Dahmen et al. (1994*b*), Jaffard (1992) and Oswald (1992).

Theorem 6.1 (Dahmen et al. 1994*b*) Suppose that the Galerkin scheme (6.6) is stable (6.7) and that the parameters $\gamma, \tilde{\gamma}$ in (6.1) satisfy

$$|t| < \gamma, \tilde{\gamma}. \tag{6.14}$$

Let $\mathbf{D}_\Lambda^s$ be the diagonal matrix defined by (5.43). Then the matrices

$$\mathbf{B}_\Lambda := \mathbf{D}_\Lambda^{-t} \mathbf{A}_\Lambda \mathbf{D}_\Lambda^{-t} \tag{6.15}$$

have uniformly bounded spectral condition numbers

$$\|\mathbf{B}_\Lambda\| \, \|\mathbf{B}_\Lambda^{-1}\| = O(1), \quad \Lambda \subset \nabla. \tag{6.16}$$

Proof. Consider any $v \in S_\Lambda$ and set $w := \Sigma_t v$ (see (5.37)). Thus, by (6.14) and (5.38), one obtains

$$\|w\|_{L_2} = \|\Sigma_t v\|_{L_2} \sim \|v\|_{H^t} \sim \|Q_\Lambda^* \mathcal{L} Q_\Lambda v\|_{H^{-t}},$$

where we have used the stability (6.7) in the last step. Employing the norm equivalence (5.38), now relative to the dual basis, and bearing (5.39) in mind, yields

$$\|w\|_{L_2} \sim \|\Sigma_{-t}^* Q_\Lambda^* \mathcal{L} Q_\Lambda \Sigma_{-t} w\|_{L_2}.$$

This means that the operators

$$\mathcal{L}_{t,\Lambda} := \Sigma_{-t}^* Q_\Lambda^* \mathcal{L} Q_\Lambda \Sigma_{-t} : S_\Lambda \to \tilde{S}_\Lambda$$

are uniformly boundedly invertible, that is,

$$\|\mathcal{L}_{t,\Lambda}\| \, \|\mathcal{L}_{t,\Lambda}^{-1}\| = O(1), \quad \#\Lambda \to \infty. \tag{6.17}$$

It is now a matter of straightforward calculation to verify that the matrix representation of $\mathcal{L}_{t,\Lambda}$ relative to Ψ_Λ is

$$\langle \mathcal{L}_{t,\Lambda} \Psi_\Lambda, \Psi_\Lambda \rangle^{\mathrm{T}} = \mathbf{D}_\Lambda^{-t} \mathbf{A}_\Lambda \mathbf{D}_\Lambda^{-t}, \tag{6.18}$$

which proves the claim. □

Letting $\#\Lambda$ tend to infinity, the original equation $\mathcal{L}u = f$ can be viewed as an infinite discrete system (recall (1.35) in Section 1.5)

$$\mathbf{D}^{-t}\mathbf{A}\mathbf{D}^{-t}\mathbf{d} = \mathbf{D}^{-t}\mathbf{f} \tag{6.19}$$

where $\mathbf{D}^{-t}$ and $\mathbf{A}$ are the infinite counterparts of $\mathbf{D}_\Lambda^{-t}$, $\mathbf{A}_\Lambda$, respectively, and $\mathbf{f} := \langle f, \Psi\rangle^{\mathrm{T}}$ is the coefficient sequence of f expanded relative to the *dual* basis $\tilde{\Psi}$. The sequence $\mathbf{d}$ then consists of the wavelet coefficients (relative to Ψ) of the solution

$$u = \mathbf{d}^{\mathrm{T}}\Psi$$

of (2.24). The infinite matrix $\mathbf{B} := \mathbf{D}^{-t}\mathbf{A}\mathbf{D}^{-t}$ is, on account of Theorem 6.1, a boundedly invertible mapping from $\ell_2(\nabla)$ onto $\ell_2(\nabla)$.

It is remarkable that similar techniques also lead to preconditioners for *collocation* matrices (Schneider 1995). In brief, recall that (6.3) defines a collocation scheme, when the P_Λ in (6.3) are *interpolation* projectors. Let us consider full sets Λ_J, defined in (3.24), which means $S(\Phi_J) = S_{\Lambda_J}$, and assume that for a suitable mesh of points $\{x_{J,k}\}_{k\in\Delta_J}$, the corresponding projectors have the form

$$P_{\Lambda_J} f = L_J f = \sum_{k\in\Delta_J} 2^{-nJ/2} f(x_{J,k})\tilde{\theta}_{J,k} =: \langle f, \delta_J\rangle\tilde{\theta}_J,$$

that is, $\delta_{J,k}\theta_{J,m} = \theta_{J,m}(x_{J,k}) = 2^{Jd/2}\delta_{k,m}$, $k, m \in \Delta_J$. For instance, $\tilde{\theta}$ could be a spline function interpolating the Kronecker sequence. Moreover, assume that

$$(L_{j+1} - L_j)\, f = \langle f, \vartheta_j\rangle\gamma_j, \tag{6.20}$$

where

$$\vartheta_j^{\mathrm{T}} = \delta_{j+1}^{\mathrm{T}}\mathbf{M}_{j,1}^{\delta}, \quad \gamma_j = \tilde{\theta}_{j+1}^{\mathrm{T}}\tilde{\mathbf{M}}_{j,1}^{\delta}, \tag{6.21}$$

are corresponding stable completions.

Theorem 6.2 (Schneider 1995) Suppose that the collocation method (6.3) relative to $P_{\Lambda_J} = L_J$ is $(s, 2t)$-stable in the sense of (6.4), and assume that

$$\frac{n}{2} < s - 2t, \quad \frac{n}{2} < \gamma_{\tilde{\theta}}, \quad \tilde{\gamma} > 0, \quad s < \gamma. \tag{6.22}$$

Then the matrices

$$\mathbf{D}_J^{s-2t}\left\langle \mathcal{L}\Psi^J, \vartheta^J\right\rangle^{\mathrm{T}}\mathbf{D}_J^{-s} \tag{6.23}$$

have uniformly bounded spectral condition numbers.

Details of the proof can be found in Schneider (1995). It uses continuity of $\mathcal{L}$ as a mapping on Sobolev spaces, the stability (6.4) and the fact that

$$\|f\|_{H^\tau}^2 \sim \sum_{j=0}^{\infty} 2^{2\tau j} \|(L_{j+1} - L_j) f\|_{L_2}^2$$

for $n/2 < \tau < \gamma_{\tilde{\theta}}$. As for this norm equivalence, note that the L_j do satisfy (5.16). But they are only bounded in higher Sobolev spaces, which causes conditions (6.22). Keeping this in mind, the above equivalence can be deduced from the general results in Sections 5.2 and 5.3. For details see Dahmen (1996).

The above simple argument is designed to show the qualitative role of norm equivalences in connection with preconditioning. In practice, the constants involved will matter. However, in principle, it should be noted that, in Theorem 6.1, neither

- selfadjointness of $\mathcal{L}$, nor
- positive order $2t > 0$

is required for the validity of (6.16).

For unsymmetric problems, (6.16) alone is not sufficient to imply the efficiency of corresponding variants of the preconditioned conjugate gradient method, such as GMRES. But behind the validity of (1.29) or (2.23) there is usually a symmetric principal part of the operator $\mathcal{L}$, in which case GMRES will perform well, provided that the condition numbers stay small. Alternatively, if the constant in (6.16) stays moderate, one can square the preconditioned system and the conjugate gradient scheme works well. We can summarize this under the following purely asymptotic result.

Remark 6.3 Suppose that every matrix vector multiplication with $\mathbf{A}_\Lambda$ can be carried out in $O(\#\Lambda)$ operations uniformly in $\Lambda \subset \nabla$, and assume that the (exact) Galerkin solution u_Λ of (6.6) in S_Λ satisfies

$$\|u - u_\Lambda\|_{H^t} = \varepsilon_\Lambda.$$

Then an approximate solution $\hat{u}_\Lambda$ of (6.6) satisfying $\|u - \hat{u}_\Lambda\|_{H^t} = O(\varepsilon_\Lambda)$ uniformly in Λ can be computed at the expense of $O(\#\Lambda)$ operations.

The argument is based on standard *nested iteration.* Solving first on a small $\Lambda_0 \subset \nabla$, then doubling $\#\Lambda_0$ to Λ_1, say, and noting that $\varepsilon_{\Lambda_0}/\varepsilon_{\Lambda_1} \leq C$, only $O(\Lambda_1)$ iterations on the preconditioned system are needed to reduce the error from ε_{Λ_0} to ε_{Λ_1}, when using $\hat{u}_{\Lambda_0}$ as a starting solution. Repeating this argument confirms the assertion.

Next, let us address some algorithmic issues. When $\mathcal{L}$ is a differential operator, the stiffness matrices

$$\mathbf{A}_{\Phi_J} = \langle \mathcal{L}\Phi_J, \Phi_J \rangle^{\mathrm{T}}$$

relative to the fine scale (nodal) bases Φ_J are sparse under the usual assumption (5.5). Due to the larger supports of wavelets from coarser scales the corresponding stiffness matrices $\mathbf{A}_{\Lambda_J}$ relative to the wavelet bases of $S(\Phi_J)$ are less sparse. Thus, assembling the wavelet stiffness matrix exactly would increase computational and storage complexity. However, when working with the fully refined sequence of spaces $S(\Phi_j)$, this can be remedied as follows. All that is needed in an iterative scheme is the *application* of the preconditioned matrix. Since by (3.31),

$$\mathbf{A}_{\Lambda_J} = \mathbf{T}_J^{\mathrm{T}} \mathbf{A}_{\Phi_J} \mathbf{T}_J, \tag{6.24}$$

where $\mathbf{T}_J$ is the multiscale transformation from (3.25), (3.26), the application of the preconditioned matrix $\mathbf{B}_J := \mathbf{D}_{\Lambda_J}^{-t} \mathbf{A}_{\Lambda_J} \mathbf{D}_{\Lambda_J}^{-t}$ to a vector $\mathbf{v}$ can be carried out as follows.

ALGORITHM 1 (CB: CHANGE OF BASES)

(1) Compute $\mathbf{w} = \mathbf{T}_J \mathbf{D}_{\Lambda_J}^{-t} \mathbf{v}$. Due to the pyramid structure of $\mathbf{T}_J$ (3.26) and the geometrical increase of $\#\Phi_j$, this requires $O(\#\Phi_J)$ operations, where the constant depends on the length of the masks in $\mathbf{M}_j$.
(2) Compute $\mathbf{z} := \mathbf{A}_{\Phi_J} \mathbf{w}$, which, due to the sparseness of $\mathbf{A}_{\Phi_J}$ is again a $O(\#\Phi_J)$ process.
(3) Compute $\mathbf{D}_{\Lambda_J}^{-t} \mathbf{T}_J^{\mathrm{T}} \mathbf{z}$, which corresponds to the first step.

Remark 6.4 When $\mathcal{L}$ is a differential operator, the application of the preconditioned matrix $\mathbf{D}_{\Lambda_J}^{-t} \mathbf{A}_{\Lambda_J} \mathbf{D}_{\Lambda_J}^{-t}$ relative to the full spaces $S(\Phi_J)$ to a vector requires the amount of $O(\#\Phi_J) = \mathcal{O}(N_J)$ operations and storage.

Remark 6.5 In the periodic case, or when working on the interval, $\mathbf{A}_{\Phi_J}$ can be computed very efficiently (even for variable coefficients) by the methods described in Section 4.2.

Remark 6.6 It is also important to note that the above preconditioner only requires knowlege of the transformation $\mathbf{T}_J$ in (3.26) *not* of the inverse $\mathbf{T}_J^{-1}$ (see Section 3.4). Recall that $\mathbf{T}_J$ involves the refinement matrices for Φ_j and the stable completions $\mathbf{M}_{j,1}$, $j < J$, that is, the masks of the wavelets (see (3.27)). Hence this method can still be used in the present context with the same efficiency when only the matrices $\mathbf{M}_j$ are sparse while the inverses $\mathbf{G}_j$ are fully populated. This is the case for many *pre-wavelets*, that is, for stable complement bases Ψ_j, which span the *orthogonal* complement of $S(\Phi_j)$ in $S(\Phi_{j+1})$.

So far this strategy refers to *fully refined* spaces $S(\Phi_J)$. Things change when the trial spaces are to be adapted *during* the solution process. This means that one actually wants to compute a solution from spaces S_Λ where Λ is a much smaller *lacunary* subset of Λ_J, $J = \max\{|\lambda| + 1 : \lambda \in \Lambda\}$. To take full advantage of the corresponding principal reduction of complexity,

any steps requiring the computational complexity of the full space $S(\Phi_J)$ should be avoided. This suggests building up the matrix $\mathbf{A}_\Lambda$ directly but only relative to the elements in Λ. How to do this efficiently depends very much on the particular operator $\mathcal{L}$. We will comment on this issue later in more detail for operators satisfying (2.25). In this case one can exploit certain decay properties of the entries of $\mathbf{A}_\Lambda$ to compute $\mathbf{A}_\Lambda$ approximately to any desired accuracy.

6.3. Hierarchical bases preconditioner

The change of bases preconditioner has already been employed in connection with hierarchical bases (Yserentant 1986). The corresponding setting of piecewise linear bivariate finite element and associated hierarchical bases was described in Section 4.1. Due to the simple and very sparse structure of the matrices $\mathbf{M}_j$ (see (4.2) and (4.4)), Algorithm 1 above is very efficient. However, the hierarchical bases are invariant under the application of $L_j - L_{j-1}$ as in (6.20), where the L_j are Lagrange interpolation operators relative to the triangulation $\mathcal{T}_j$. Hence they are *not* bounded in $L_2(\Omega)$ and the collection $\Phi_0 \cup \bigcup_{j=0}^{\infty} \Psi_j$ is *not* a Riesz basis for $L_2(\Omega)$. Moreover, $\|\cdot\|_{H^1}$ is *not* equivalent to the discrete norm ($L_{-1} := 0$) $\left(\sum_{j=0}^{\infty} 2^{2j} \|(L_{j+1} - L_j)f\|_{L_2}^2\right)^{1/2}$ for $n \geq 2$. Hence the *hierarchical* basis preconditioner, based on Algorithm 1, is *not* asymptotically optimal. For $n = 2$, the condition numbers grow like the square of the number of levels, while for $n = 3$ they already exhibit an exponential growth. Nevertheless, its extreme simplicity accounts for its attractiveness for $n = 2$. Ways of *stabilizing* it, for instance with the aid of the techniques in Section 3.5, will be presented later.

6.4. BPX scheme

Although, as it stands, the simple hierarchical complement bases do not provide an asymptotically optimal scheme with regard to preconditioning, it turns out that the full power of wavelet decompositions is needed only for operators of *non-positive* order. Throughout this section we will assume that $\mathcal{L}$ is selfadjoint positive definite (6.8), (6.9) and that the order $2t$ of $\mathcal{L}$ is *positive*. In this case one gets away with much less. So, suppose the bases Ψ_J are stable in the sense of (3.3) and give rise to a hierarchy of nested spaces $S_J = S(\Phi_J) \subset H^t$ as before. The following discussion reflects an approach to multilevel preconditioners developed in the context of finite element discretizations (Bramble et al. 1990, Oswald 1992, Yserentant 1990, Xu 1992, Zhang 1992). The objective is to find a positive definite selfadjoint operator $\mathcal{C}_J$ on S_J such that

$$\langle \mathcal{C}_J^{-1} v, v\rangle \sim \langle \mathcal{L}_J v, v\rangle = a(v,v), \quad v \in S_J, \tag{6.25}$$

which means that $\mathcal{C}_J$ and $\mathcal{L}_J$ are *spectrally equivalent.* In fact, the uniformity of (6.25) in J implies, in view of the min-max characterization of eigenvalues, that

$$\frac{\lambda_{\max}\left(\mathcal{C}_J^{1/2}\mathcal{L}_J\mathcal{C}_J^{1/2}\right)}{\lambda_{\min}\left(\mathcal{C}_J^{1/2}\mathcal{L}_J\mathcal{C}_J^{1/2}\right)} \sim 1. \tag{6.26}$$

To describe a candidate for $\mathcal{C}_J$, let P_J denote the orthogonal projector onto S_J. Clearly the P_J satisfy (5.16) and $P_J^* = P_J$. Thus Theorem 5.8 applies and (5.38) means that

$$\hat{\mathcal{C}}_J^{-1} = \sum_{j=0}^{J} 2^{2tj}(P_j - P_{j-1}), \quad P_{-1} := 0,$$

satisfies

$$\langle \hat{\mathcal{C}}_J^{-1}v, v\rangle = \langle \hat{\mathcal{C}}_J^{-1/2}v, \hat{\mathcal{C}}_J^{-1/2}v\rangle = \|\Sigma_t v\|_{L_2} \sim \|v\|_{H^t}. \tag{6.27}$$

Hence, by ellipticity,

$$\langle \hat{\mathcal{C}}_J^{-1}v, v\rangle \sim a(v,v), \quad v \in S_J, \tag{6.28}$$

so that $\hat{\mathcal{C}}_J P_J \mathcal{L} P_J$ have uniformly bounded condition numbers. This corresponds to the situation assumed in Theorem 6.1, since the evaluation of $\hat{\mathcal{C}}_J$ seems to require knowledge of explicit bases for the orthogonal complements. However, since, clearly, by (5.39),

$$\hat{\mathcal{C}}_J = \sum_{j=0}^{J} 2^{-2tj}(P_j - P_{j-1}),$$

and $t > 0$, $\hat{\mathcal{C}}_J$ is easily seen to be spectrally equivalent to $\overline{\mathcal{C}}_J := \sum_{j=0}^{J} 2^{-2tj} P_j$, which, by the uniform stability of the Φ_j, is spectrally equivalent to

$$\mathcal{C}_J v := \sum_{j=0}^{J} 2^{-2tj} \sum_{k\in\Delta_j} \langle v, \phi_{j,k}\rangle \phi_{j,k}. \tag{6.29}$$

Combining the spectral equivalence of $\mathcal{C}_J$ and $\hat{\mathcal{C}}_J$ with (6.27) and (6.29) yields

$$\langle \mathcal{C}_J P_J \mathcal{L} P_J v, v\rangle \sim \langle v, v\rangle. \tag{6.30}$$

Hence (Dahmen and Kunoth 1992, Oswald 1992, Zhang 1992),

$$\kappa_2\left(\mathcal{C}_J P_J \mathcal{L} P_J\right) = O(1), \quad j \in \mathbb{N}. \tag{6.31}$$

Note that application of $\mathcal{C}_J$ does *not* require explicit knowledge of any complement basis. It also requires only the order of $\#\Phi_J$ operations. For more details about the actual implementation, the reader is referred to Bramble et al. (1990) and Xu (1992).

Exact decompositions in terms of complement bases have been replaced by *redundant* spanning sets, which consist here of properly weighted nodal basis functions on *each* level. In brief, the collections $\{2^{-2jt}\phi_{j,k} : k \in \Delta_j, j = 0, \ldots, J\}$ form *frames* for $H^t(\Omega)$. Here $\{g_j\}$ is called a frame for H, if

$$\|v\|_H^2 \sim \sum_j |\langle v, g_j\rangle_H|^2, \quad v \in H. \tag{6.32}$$

It is perhaps worth stressing the relation to the wavelet transforms.

Remark 6.7 The corresponding wavelet preconditioner looks like

$$C_J^w v = \sum_{j=-1}^{J-1} 2^{-2jt} \sum_{k \in \nabla_j} \langle v, \psi_{j,k}\rangle \psi_{j,k}.$$

Since the $\psi_{j,k}$ are linear combinations of the $\phi_{j+1,m}$, its evaluation always seems to be *more* expensive than that of (6.29). The cost of each iteration increases with the lengths of the masks of the wavelets (Griebel and Oswald 1995*b*, Oswald 1994).

Adaptive grids

Theorem 6.1 has been formulated for *arbitrary* subsets $\Lambda \subset \nabla$. Thus adaptivity can be based, in principle, on adapting the choice of Λ to the problem at hand. It will be explained later how to arrange that. Roughly speaking, the behaviour of the wavelet coefficients themselves is an indication for the selection of relevant indices. The point is that this kind of adaptation essentially requires managing *index sets.*

So far, in a finite element context, the above discussion of the BPX scheme refers to spaces generated by *uniform* refinements. Adaptivity usually requires *mesh refinement* strategies based on monitoring the current solution through additional local comparisons. It is interesting to see how preconditioning is affected when working with adaptively refined meshes. Employing *hanging* or *slave* nodes, that is, adding locally further nodal basis functions, corresponds, roughly speaking, to considering submatrices of those stemming from uniform refinements. Since the convex hull of the spectrum of the latter matrices contains the spectrum of the submatrices, the BPX scheme is trivially adapted to nonuniform refinements and the condition numbers remain bounded.

Slave nodes require a little care retrieving stable bases for the resulting finite element spaces. If one wants to avoid slave nodes, the nonuniform refinements have to be *closed* by introducing suitable transition elements; see, for instance, Bank, Sherman and Weiser (1983). In this case, the submatrix argument does not work in a strict sense. Nevertheless, one can prove that for such adaptive refinements resulting in highly nonuniform meshes, the BPX scheme still produces uniformly bounded condition numbers. This has

been shown first in Dahmen and Kunoth (1992), where further details can be found; see also Bornemann and Yserentant (1993).

An analogous result holds for fourth-order problems. As a model case, one could consider $\mathcal{L} = \Delta^2$ with homogeneous Dirichlet boundary conditions. A convenient conforming finite element discretization can be based on certain piecewise cubic macro patches generated by suitable subdivisions of (non rectangular) quadrilaterals. These are obtained by connecting the intersection of diagonals with the midpoint of the edges of the quadrilateral. The nodal basis functions are fundamental interpolators relative to point values and gradients at the corners of the quadrilaterals and normal derivatives at the midpoints of edges. The resulting spaces are nested and the underlying mesh refinements stay regular, in the sense that smallest angles are bounded away from zero. See Dahmen, Oswald and Shi (1993*a*) for more details. Adaptive refinements analogous to the piecewise linear case are discussed by Kunoth (1994), where the corresponding result about uniformly bounded condition numbers is also established. One should note that the classical cubic Clough–Tocher macro element is not suited for refinements. Since the quintic C^1-Argyris element requires higher smoothness at the vertices, its refinement leads to nonnested trial spaces.

The hierarchical basis and BPX preconditioner are special instances of the following more general class of schemes that have a long tradition in the finite element context.

6.5. Multilevel Schwarz schemes

We will briefly indicate how the above material ties into the more general setting of *Schwarz schemes* and *stable splittings*, which is also a convenient framework for incorporating domain decomposition and multigrid techniques. For a more extensive treatment of these issues, as well as further details concerning the following discussion, we refer, for example, to Griebel and Oswald (1995*a*), Oswald (1994), Xu (1992) and Yserentant (1993). As above, $\mathcal{L}$ will be selfadjoint positive definite on some separable Hilbert space $H = H_1$, that is, $a(u, v) := \langle \mathcal{L}u, v\rangle$ is a symmetric bilinear form and we assume that (6.9) holds with H^t replaced by H. We wish to find $u \in H$ such that

$$a(u, v) = f(v), \quad v \in H, \tag{6.33}$$

where f is a linear functional on H. In fact, at this point one can think of H being some Sobolev space H^t as above but also of the finite dimensional trial space $S(\Phi_J)$ of highest resolution. Let $\{V_j\}$ be an at most countable collection of closed nested subspaces of H such that every $v \in H$ has at least one expansion

$$v = \sum_j v_j,$$

which converges in H, in brief $H = \sum_j V_j$.

The basic idea is to solve for each V_j the problem restricted to V_j and then add these solutions up. This corresponds to (block) Gauss–Seidel or Jacobi relaxation. The solution of the subproblems will be based on auxiliary inner products $b_j(\cdot,\cdot)$ on V_j which approximate $a(\cdot,\cdot)$. Following Oswald (1994), we write $\{H;a\}$, $\{V_j;b_j\}$ to express that each V_j, H, are Hilbertian relative to the scalar products b_j, a. The *subspace splitting*

$$\{H;a\} = \sum_j \{V_j;b_j\}$$

is called *stable* if

$$\|v\|^2_{\{b_j\}} := \inf\left\{ \sum_j b_j(v_j,v_j) : v_j \in V_j, v = \sum_j v_j \right\} \sim a(v,v). \tag{6.34}$$

Taking $V_j = S(\Phi_j)$, $b_j(v,v) := 2^{2sj}\langle v,v\rangle$, $H = H^s(\Omega)$, we see that the norm equivalence (5.44) is a special case of (6.34). Alternatively, setting $V_{-1} := S(\Phi_0)$, $V_j = S(\Psi_j)$, $j \geq 0$, $v_j = (Q_j - Q_{j-1})v$, $b_j(v,v)$ as before, it is clear that $\|v\|_{\{b_j\}} \lesssim \|v\|_{\mathcal{A}^s_Q}$. Now consider the following Riesz operators $\mathcal{T}_j$, g_j which interrelate the scalar products.

$$\begin{aligned} &\mathcal{T}_j : H \to V_j : \quad b_j(\mathcal{T}_j v, v_j) = a(v,v_j), \\ &g_j \in V_j : \qquad\quad b_j(g_j,v_j) = f(v_j), \quad v_j \in V_j. \end{aligned} \tag{6.35}$$

Defining $B_j : H \to V_j$ by $b(v,v_j) = \langle B_j v, v_j\rangle, v_j \in V_j$, and recalling that $a(u,v) = \langle \mathcal{L}u, v\rangle$, we infer that $\mathcal{T}_j = B_j^{-1}\mathcal{L}$, so that each application of $\mathcal{T}_j$ corresponds to the solution of a restricted (typically small) problem. A central observation in this context is the following result; see Oswald (1994) for the background and further references.

Theorem 6.8 Let

$$\mathcal{T} := \sum_j \mathcal{T}_j, \quad g = \sum_j g_j. \tag{6.36}$$

Then equation (6.33) is equivalent to the operator equation

$$\mathcal{T}u = g, \tag{6.37}$$

which is called the *additive* Schwarz formulation of (6.33). Moreover, $\mathcal{T}$ is selfadjoint positive definite and when H is finite-dimensional its smallest and largest eigenvalue are $\lambda_{\min}(\mathcal{T})$, $\lambda_{\max}(\mathcal{T})$, where

$$\lambda_{\max}(\mathcal{T}) = \sup_{v\in H} \frac{a(v,v)}{\|v\|^2_{\{b_j\}}}, \quad \lambda_{\min}(\mathcal{T}) = \inf_{v\in H} \frac{a(v,v)}{\|v\|^2_{\{b_j\}}},$$

respectively.

Since

$$\kappa_2(\mathcal{T}) := \frac{\lambda_{\max}(\mathcal{T})}{\lambda_{\min}(\mathcal{T})}, \tag{6.38}$$

the condition of (6.37) is bounded by the ratio of the constants in the upper and lower bound in (6.34). A corresponding asymptotic statement requires the ratio of the upper and lower bound to remain uniformly bounded when the dimension of H increases. The quantitative relevance of such statements again depends, of course, on the problem at hand.

Theorem 6.8 can be deduced from the following result, which is interesting in its own right and has several further applications mentioned below.

Theorem 6.9 (Nepomnyaschikh 1990) Let $H, \tilde{H}$ be two Hilbert spaces with scalar products $\langle\cdot,\cdot\rangle_H$, $\langle\cdot,\cdot\rangle_{\tilde{H}}$, respectively, and with bilinear forms $a, \tilde{a}$ induced by symmetric positive definite operators $\mathcal{L} : H \to H$, $\tilde{\mathcal{L}} : \tilde{H} \to \tilde{H}$. Suppose that there exists a surjective bounded linear operator $\mathcal{R} : \tilde{H} \to H$ such that

$$a(v,v) \sim \inf_{\tilde{v}\in\tilde{H}, v=\mathcal{R}\tilde{v}} \tilde{a}(\tilde{v},\tilde{v}) := |\!|\!| v |\!|\!|, \quad v \in H.$$

Then $\mathcal{P} := \mathcal{R}\tilde{\mathcal{L}}^{-1}\mathcal{R}^*\mathcal{L} : H \to H$ is symmetric positive definite with

$$\lambda_{\max}(\mathcal{P}) = \sup_{v\in H} \frac{a(v,v)}{|\!|\!| v |\!|\!|^2}, \quad \lambda_{\min}(\mathcal{P}) = \inf_{v\in H} \frac{a(v,v)}{|\!|\!| v |\!|\!|^2}.$$

Of course, Theorem 6.8 is obtained by taking $\tilde{H} := \{\tilde{v} = \{v_i\} : v_i \in V_i, \sum_i b_i(v_i,v_i) < \infty\}$; see, for example, Oswald (1994).

An interesting case is, for example, $H = S(\Phi_J)$ for some (large J) where the splitting consists of *one-dimensional* subspaces $V_{j,k} := S(\{\phi_{j,k}\})$. Thus the corresponding Riesz operator $\mathcal{T}_{j,k}$ has the form $\mathcal{T}_{j,k}v = c_{j,k}\phi_{j,k}$ where

$$c_{j,k} = a(v,\phi_{j,k})/b_{j,k}(\phi_{j,k},\phi_{j,k}). \tag{6.39}$$

Thus

$$\mathcal{T}v = \sum_{j=0}^{J}\sum_{k\in\Delta_j} \frac{a(v,\phi_{j,k})}{b_{j,k}(\phi_{j,k},\phi_{j,k})}\phi_{j,k}. \tag{6.40}$$

Hence (6.34) means that $\{\phi_{j,k}/b_{j,k}(\phi_{j,k},\phi_{j,k})\}$ forms a frame for $\{H_j a\}$.

The BPX scheme and the hierarchical basis preconditioner are examples of this type. To see this, let $f(v) = \langle f, v\rangle$, where $\langle\cdot,\cdot\rangle$ is induced by the L_2 inner product. Let $B_{j,k}$ be defined by $b_{j,k}(u_{j,k},v_{j,k}) = \langle B_{j,k}u_{j,k}, v_{j,k}\rangle$, so that $\mathcal{T}_{j,k} = B_{j,k}^{-1}P_{j,k}\mathcal{L}$, $P_{j,k}$ the orthogonal projection onto $V_{j,k}$. Then one can write

$$\mathcal{T} = \left(\sum_{j=0}^{J}\sum_{k\in\Delta_j} B_{j,k}^{-1}P_{j,k}\right)\mathcal{L} =: \mathcal{C}^{-1}\mathcal{L}.$$

Thus, choosing $b_{j,k}(\cdot,\cdot) := 2^{2jt}\langle\cdot,\cdot\rangle$ and $V_{j,k} = S(\{\phi_{j,k}\}), k \in \Delta_j$, $V_{j,k} = S(\{\phi_{j,k}\}), k \in \Delta_j \setminus \Delta_{j-1}$, yields the BPX and hierarchical basis preconditioner, respectively (see (6.29)).

There is obviously great flexibility in choosing the subspaces V_j. In general, realization of the operators $\mathcal{T}_j$ requires solving linear systems of the size determined by V_j. In the context of parallel computing one might accept spaces V_j of growing dimension. For instance, the V_j could be chosen as subspaces of H associated with a decomposition of the underlying domain. A survey of applications of this type can be found in Chan and Mathew (1994).

The proofs establishing the critical equivalence (6.34) are, of course, related to the concepts discussed in Section 5.3. The norm $\|\!|\cdot|\!\|_{\{b_j\}}$ can in certain cases be evaluated exactly, which greatly simplifies the analysis; compare Griebel and Oswald (1995*a*), and Nießen (1995). The bounds $\|\!|\cdot|\!\|^2_{\{b_j\}} \lesssim a(\cdot,\cdot)$ typically involve approximation theory tools. The converse estimates are often reduced to the validity of a *strengthened Cauchy–Schwarz inequality* (recall Section 5.2), which in this context has the form

$$a(v_j, v_k) \lesssim \gamma_{j,k} b_j(v_j, v_j) b_k(v_k, v_k), \quad v_j \in V_j, v_k \in V_k, \tag{6.41}$$

where $(\gamma_{j,k})$ should be bounded in ℓ_2. For a detailed discussion of these issues, compare Griebel and Oswald (1995*a*), Yserentant (1993).

Again following Oswald (1997), we mention an interesting extension of the splitting concept which aims at relaxing the assumption of nestedness $V_j \subset V_{j+1}$, as well as of conformity $V_j \subset H$. This, again, is an application of Theorem 6.9. It requires introducing mappings $R_j : V_j \to H$ such that $R := \sum_j R_j : \prod_j V_j \to H$ is *onto* and

$$\|\!|v|\!\|_{\{b_j;R_j\}} := \inf\left\{\sum_j b_j(v_j, v_j) : v_j \in V_j, v = \sum_j R_j v_j\right\} \sim a(v,v), \quad v \in H. \tag{6.42}$$

In this case, $\mathcal{T}$ in (6.37) has to be replaced by $\mathcal{T}' := \sum_j (R_j B_j^{-1} R_j^*)\mathcal{L}$, where $B_j : V_j \to V_j$ is now defined by $b(u_j, v_j) = \langle B_j u_j, v_j\rangle_{V_j}$, $v_j \in V_j$. Equivalently, one can write $\mathcal{T}' = \sum_j R_j \mathcal{T}_j'$, where $\mathcal{T}_j' : H \to V_j$ is given by $b(\mathcal{T}_j' v, v_j) = a(v, R_j v_j)$, $v_j \in V_j$. See Oswald (1994), Griebel and Oswald (1995*a*).

We now indicate a typical iteration based on (6.35) and (6.36). The *additive version* **A** creates a sequence of approximations $\{u^l\}$ given by

$$u^{l+1} = u^l + \omega \sum_{j=0}^{J} (g_j - \mathcal{T}_j u^l). \tag{6.43}$$

Here ω plays the same role of a relaxation parameter as in the Jacobi or Richardson iteration. Recall that each iteration requires the solution of

variational problems on the spaces V_j. Perhaps the structure of the iteration (6.43) becomes more transparent on recalling that $\mathcal{T}_j = B_j^{-1}\mathcal{L}$, so that $g_j - \mathcal{T}_j u^l = B_j^{-1} r^l$, where $r^l = f - \mathcal{L}u^l$ is the residual from the last step. Likewise, in the more general version (6.42), $g_j - \mathcal{T}_j u^l$ has to be replaced by $R_j B_j^{-1} R_j^* r^l$.

The *multiplicative* version **M** reads

$$\begin{aligned} v^0 &:= u^l, \\ v^{j+1} &= v^j + \omega(g_{J-j} - \mathcal{T}_{J-j} v^j), \quad j = 0, \dots, J, \\ u^{l+1} &= v^{J+1}, \end{aligned} \tag{6.44}$$

generalizing SOR. The corresponding iteration operators are

$$M_{\mathbf{A}} = I - \omega\mathcal{T}, \quad M_{\mathbf{M}} = (I - \omega\mathcal{T}_0)(I - \omega\mathcal{T}_1)\cdots(I - \omega\mathcal{T}_J).$$

The convergence theory is given by Bramble (1993), Griebel and Oswald (1995*a*), Xu (1992) and Yserentant (1993). For the interpretation of these schemes in the multigrid context, see Bramble (1993) and Griebel (1994). Here we quote the following result from Griebel and Oswald (1995*a*).

Theorem 6.10 Assume that H is finite-dimensional and the algorithms **M** and **A** are given by (6.43) and (6.44).

(i) **A** converges for $0 < \omega < 2/\lambda_{\max}(\mathcal{T})$. The optimal rate is achieved for $\omega^* = 2/(\lambda_{\max}(\mathcal{T}) + \lambda_{\min}(\mathcal{T}))$ and equals

$$\rho_{\mathbf{A}} = \min_{0<\omega<2/\lambda_{\max}} \|M_{\mathbf{A}}\|_a = 1 - \frac{2}{1 + \kappa_2(\mathcal{T})}.$$

(ii) Suppose (6.41) holds with $\gamma_{jj} = 1$. Then **M** converges for $0 < w < 2$. The optimal rate is bounded by

$$\rho_M \le 1 - \frac{\lambda_{\min}(\mathcal{T})}{2\lambda_{\max}(\mathcal{T}) + 1}.$$

For various modifications see Oswald (1997) and the literature cited there.

6.6. Finite element-based wavelets

The previous discussion shows that preconditioning matrices stemming from Galerkin discretizations of elliptic operators of positive order does *not* require explicit knowledge of wavelet bases. Nevertheless, a number of recent studies have addressed the construction and application of wavelets in a finite element context, to obtain wavelet-based stable splittings for Schwarz schemes. Let us briefly postpone giving reasons why the additional effort might still pay in this context, and first outline some ingredients of the various approaches.

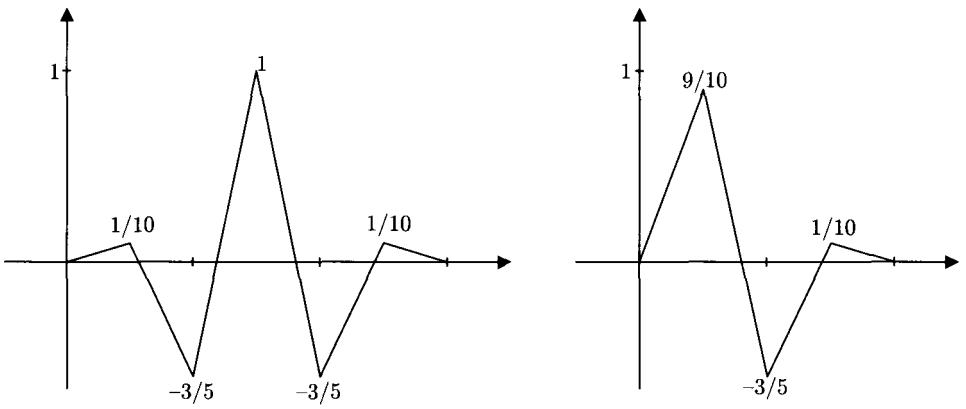

Fig. 1. Pre-wavelets for $H_0^1([0,1])$

Pre-wavelets

Recall from (6.24) that preconditioning based on a change of basis does *not* require the transform $\mathbf{T}_J^{-1}$. Thus it suffices to make sure that the scheme (3.26) is efficient, which means that the masks of the wavelets have possibly small support. Most of the presently known FE-based wavelets still refer to an underlying uniform grid structure for multilinear finite elements on regular lattices $h\mathbb{Z}^d$ (type-1 mesh) or to regular triangulations of the plane which are generated from the standard uniform rectangular mesh by inserting in each square element the southwest–northeast diagonal (type-2 mesh).

When dealing with meshes of type 1 restricted to the unit square $\square = [0,1]^2$, say, one can employ tensor products of biorthogonal wavelets on $[0,1]$ discussed in Section 4.4. For $k \in \mathbb{Z}$, $e \in \{0,1\}^2$, $j \geq j_0$, they have the form

$$\psi_{e,j,k}(x) = \psi_{e_1,j,k_1}(x_1)\psi_{e_2,j,k_2}(x_2)$$

where $\psi_{0,j,k} = \phi_{j,k}$, $\psi_{1,j,k} = \psi_{j,k}$ are the corresponding univariate generator and wavelet functions. When ϕ is the standard piecewise linear tent function (1.15) and the dual bases are exact of order 2 as well, the mask coefficients can be found in Dahmen et al. (1996*b*), and Dahmen and Schneider (1997*a*).

In most cases, however, so-called piecewise linear pre-wavelets have been used; see, for example, Griebel and Oswald (1995*b*). Interior and boundary wavelets are shown below in Figure 1.

Here, pre-wavelet means that these wavelets form uniformly L_2-stable bases for *orthogonal* complements between two successive trial spaces. Hence they also form a Riesz basis for $L_2([0,1])$. In this case the masks in the inverse transformations are not local but, as mentioned before, this is harmless

here. Obviously the masks for the tensor product wavelets have 15 or 25 nonzero coefficients.

Piecewise linear pre-wavelets for meshes of type 2 have been constructed by Kotyczka and Oswald (1996). Those of smallest possible support have 13 nonzero coefficients. The construction principle is to make an ansatz of a linear combination of tent functions on the fine scale so that, for possibly few nonzero coefficients, orthogonality to the tent functions on the coarser scale holds. Usually, the difficult part is to verify that three such linear combinations form a stable basis on each given level. Also, the adaptation to the boundary is in this case more difficult than in the tensor product case.

The resulting pre-wavelets still have relatively large support. Therefore several alternatives have been proposed resulting in complement spaces that are no longer orthogonal but are spanned by functions of smaller support, while still exhibiting better stability properties across levels than the hierarchical bases.

For instance, the discretization of the double layer potential equation on a polyhedron in Dahmen, Kleemann, Prößdorf and Schneider (1994*a*) involves piecewise linear wavelet type functions of the form shown in Figure 2.

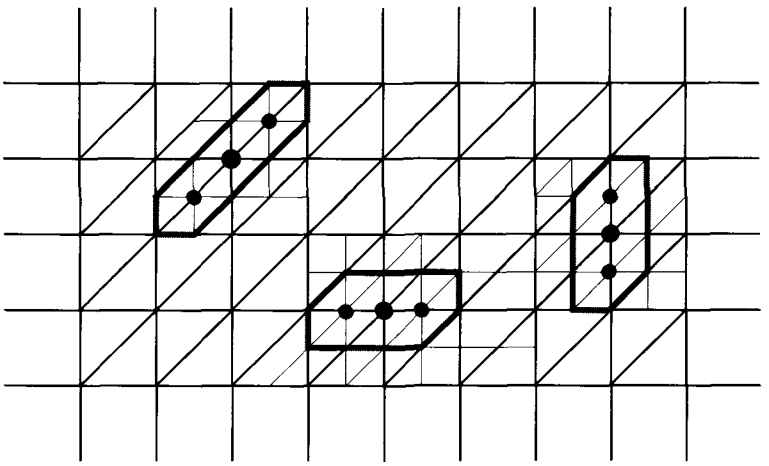

Fig. 2. Short support wavelets

The stencils in this case are

$$\begin{pmatrix} 0 & 0 & -1/2 \\ 0 & 1 & 0 \\ -1/2 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ -1/2 & 1 & -1/2 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & -1/2 & 0 \\ 0 & 1 & 0 \\ 0 & -1/2 & 0 \end{pmatrix}.$$

Here the central coefficient refers to a point in the coarse mesh, while all neighbours refer to points in the next finer mesh. Its univariate counterpart

for ϕ defined by (1.15) is

$$\psi(x) = -\frac{1}{2}\phi(2x+1) + \phi(2x) - \frac{1}{2}\phi(2x-1), \tag{6.45}$$

used by Dahmen et al. (1996*a*) to discretize the Helmholtz equation on a closed curve. The fact that these functions actually span Riesz bases in L_2 was shown by Stevenson (1995*b*). More precisely, as pointed out by Lorentz and Oswald (1996, 1997), the functions give rise to Sobolev norm equivalences (6.1) for $n \leq 3$ in the range $s \in (-0.992036, 3/2)$.

Motivated by earlier work by Hackbusch (1989) about *frequency filtering*, an interesting systematic approach to constructing L_2-stable finite element wavelet bases was proposed by Stevenson (1995*b*, 1996, 1995*a*). Again let S_j denote the space of piecewise linear finite elements on meshes of type 2 with scale $2^{-j} = h$. The central idea is to employ level *dependent discrete* scalar products which on those spaces are uniformly equivalent to the standard L_2-inner product. For instance, writing

$$\int_\Omega f(x)g(x)\,\mathrm{d}x = \langle f, g\rangle_\Omega = \sum_{\tau\in\mathcal{T}_j} \langle f, g\rangle_\tau, \tag{6.46}$$

where $\mathcal{T}_j$ is the triangulation of level j of Ω, the terms $\langle f, g\rangle_\tau$ are replaced by a *quadrature rule*. The rationale is that orthogonality with respect to discrete inner products is often easier to realize and corresponding masks are shorter, which gives rise to functions with smaller support. Thus, when $\tau = [x^1, x^2, x^3] \in \mathcal{T}_j$ has vertices x^1, x^2, x^3 one can set

$$\langle f, g\rangle_{\tau,j} = \frac{1}{|\tau|}\sum_{i=1}^{3} f(x^i)g(x^i). \tag{6.47}$$

Now, given the usual tent functions $\phi_{j,k}$ from (4.1) as generators for S_j, one then seeks for a biorthogonal collection $\Xi_j \subset S_{j+1}$ of linear combinations on the next higher level, that is,

$$\langle \Phi_j, \Xi_j\rangle = \mathbf{I}, \tag{6.48}$$

where the $\zeta_{j,k} \in \Xi_j$ have possibly small support. These auxiliary collections Ξ_j are then used to construct complement functions in S_{j+1} which are orthogonal to S_j relative to the level dependent inner product. As mentioned before, one exploits the fact that orthogonality with respect to the discrete inner products is much easier to realize than for the standard inner product. Details and concrete examples can be found in Stevenson (1995*b*, 1996, 1995*a*). In light of Section 5.2, the discrete inner products have been used here to construct a Riesz basis in L_2 without identifying the dual basis relative to the standard inner product. Compared with orthogonal splittings, one takes advantage of significantly smaller filters.

Stabilization of hierarchical bases

A further alternative has been proposed by Carnicer et al. (1996) for refinements $\mathcal{T}_j$ of *arbitrary* triangulations described in Section 4.1. Again denoting by Φ_j the L_2-normalized tent functions relative to $\mathcal{T}_j$, a biorthogonal collection Ξ_j is constructed which consists of (discontinuous) piecewise linear functions. To describe this briefly, let Δ_j again denote the set of vertices of triangulation $\mathcal{T}_j$ and let $\tau = [k, m, p]$ be a triangle in $\mathcal{T}_j$. Then there exist unique affine functions $\sigma^\tau_{j,q}$ such that $\int_\tau \phi_{j,q}(x)\sigma^\tau_{j,q'}(x)\,dx = \delta_{q,q'}$, $q, q' \in \{k, m, p\}$. Set

$$\zeta_{j,k}(x) := \begin{cases} \dfrac{1}{n_k}\sigma^\tau_{j,k}(x), & x \in \tau, \tau \subset \operatorname{supp}\phi_{j,k}, \\ 0, & x \notin \operatorname{supp}\phi_{j,k}, \end{cases} \tag{6.49}$$

where n_k is the number of triangles having k as a vertex. Thus (6.48) $\langle \Phi_j, \Xi_j\rangle = \mathbf{I}$ again holds, and the question arises of how to identify a stable basis Ψ_j for the complement space

$$W_j := \{\langle g, \Xi_{j+1}\rangle_\Omega \Phi_{j+1} - \langle g, \Xi_j\rangle_\Omega \Phi_j : g \in S_{j+1}\}, \tag{6.50}$$

induced by the projectors $\langle \cdot, \Xi_j\rangle_\Omega \Phi_j$. As in the previously discussed case, this can be done by exploiting the fact that some *initial* complement space is available, namely the one spanned by the hierarchical basis described in Section 4.1. To distinguish it, it will be denoted here as

$$\check{\Psi}_j := \{\phi_{j+1,k} : k \in \Delta_{j+1} \setminus \Delta_j\}. \tag{6.51}$$

At this point the techniques described in Section 3.5 come into play. In particular, (3.40) applies. In fact, $\mathbf{M}_{j,0}$, $\check{\mathbf{M}}_{j,1}$, $\check{\mathbf{G}}_{j,0}$, $\check{\mathbf{G}}_{j,1}$ are given by (4.2), (4.4), (4.6), (4.7) respectively. Thus, with Ξ_j defined by (6.49), the new stable completions $\mathbf{M}_{j,1}$ defined by (3.40) are readily computable. In view of the form (3.39) of the new basis functions, this process may be viewed as a *coarse grid stabilization*. The construction is *not* restricted to regular triangulations. In the case $n = 2$ and the special case of regular triangulations of type 2 (see above), the stabilized complement basis functions have the form

$$\psi_{j,k} = \phi_{j+1,k} - \sum_{l=1}^{4} a_l \phi_{j,k(l)}, \quad k \in \Delta_{j+1} \setminus \Delta_j, \tag{6.52}$$

where k are the midpoints of the edges in the triangulation $\mathcal{T}_j$, and the $k(l)$ denote the vertices of the parallelepiped having the edge associated with k as a diagonal. The construction obtained above through (6.49) is a special case of a whole family of stabilizations (6.52) of the form (Lorentz and Oswald 1996, Lorentz and Oswald 1997)

$$a_1 = a_2 = a, \quad a_3 = a_4 = \frac{1}{8} - a, \quad a \in \mathbb{R}, \tag{6.53}$$

namely for $a = 1/6$. The corresponding basis for the uniform setting is shown by Lorentz and Oswald (1996, 1997) to satisfy (6.1) for $s \in (-0.35768, 3/2)$. The choice $a = 3/16$ from Lorentz and Oswald (1996) gives a somewhat larger range $s \in (-0.440765, 3/2)$, which is maximal in this class.

There is a closely related but slightly different approach to such coarse grid stabilizations, proposed by Vassilevski and Wang (1997*a*). Recall that the multiscale transformations $\check{\mathbf{T}}_j$ associated with the hierarchical bases $\check{\Psi}_j$ from (6.51) are extremely efficient. Since it only involves nodal basis functions, not with respect to the full bases on each level but with respect to the complements only, it is even more efficient than BPX. The objective is to stabilize the hierarchical basis while retaining as much of its efficiency as possible. When switching to another stable completion of the form (3.40), the efficiency of $\check{\mathbf{T}}_j$ can still be exploited. In fact, the multiscale transformation $\mathbf{T}_j$ can be performed in two stages. First perform a step of $\check{\mathbf{T}}_j$ and then correct it, on account of (3.39) or (6.52), by terms involving only the coarse scale generator basis functions. Relevant algorithmic details are given in Sweldens (1996, 1997). The idea is to construct complement functions, that are *close* to functions which span the *orthogonal* complement between two successive trial spaces. I would like to deviate from the original approach and phrase this here in terms of the stable completions described in Section 3.5. Again, straightforward computations show that, given $\check{\mathbf{M}}_{j,1}$ as above,

$$\mathbf{M}_{j,1} = \Big(\mathbf{I} - \Big(\mathbf{M}_{j,0}\langle\Phi_j, \Phi_j\rangle^{-1}\langle\Phi_j, \Phi_{j+1}\rangle\Big)\Big)\,\check{\mathbf{M}}_{j,1} \tag{6.54}$$

gives rise to a basis $\Psi_j^{\mathrm{T}} = \Phi_{j+1}^{\mathrm{T}}\mathbf{M}_{j,1}$ spanning the orthogonal complement of $S(\Phi_j)$ in $S(\Phi_{j+1})$. In other words,

$$\mathbf{L}_j = -\langle\Phi_j, \Phi_j\rangle^{-1}\langle\Phi_j, \Phi_{j+1}\rangle\check{\mathbf{M}}_{j,1}, \quad \mathbf{K}_j = \mathbf{I}, \tag{6.55}$$

(see (3.37)) yield a suitable new stable completion. Note that, by (3.4), $\langle\Phi_j, \Phi_{j+1}\rangle = \mathbf{M}_{j,0}\langle\Phi_{j+1}, \Phi_{j+1}\rangle$. Of course, the matrix $\langle\Phi_j, \Phi_j\rangle^{-1}$ is dense and so is $\mathbf{M}_{j,1}$. However, to compute $\mathbf{M}_{j,1}\mathbf{d}$ for any coefficient vector $\mathbf{d}$ it suffices to compute $\check{\mathbf{M}}_{j,1}\mathbf{d} =: \hat{\mathbf{d}}$ and $\langle\Phi_j, \Phi_{j+1}\rangle\hat{\mathbf{d}} =: \mathbf{b}$. Next, instead of computing $\langle\Phi_j, \Phi_j\rangle^{-1}\mathbf{b}$ exactly, one performs only a few relaxation sweeps for the linear system

$$\langle\Phi_j, \Phi_j\rangle\mathbf{y} = \mathbf{b},$$

followed by $\hat{\mathbf{d}} - \mathbf{M}_{j,0}\mathbf{y}$. Note that $\langle\Phi_j, \Phi_j\rangle$ is positive definite and uniformly well conditioned, since the Φ_j are uniformly stable. Further details are found in Vassilevski and Wang (1997*a*), Vassilevski and Wang (1997*b*). Special cases again lead to (6.52) with coefficients a_l as in (6.53) with $a = 5/48$ and $s \in (0.248994, 3/2)$ (Lorentz and Oswald 1996).

Efficiency and robustness

Which of these options should be preferred? There is, of course, no uniform answer. The decision would depend on the precise problem, on the mesh and on many other side constraints. As soon as one fixes a particular model problem, some aspects become prominent. Nevertheless, there have been some recent comparisons that provide interesting information. These comparisons address two important issues, namely *efficiency* and *robustness*.

Firstly, efficiency comparisons are discussed by Ko, Kurdila and Oswald (1997) for the model problem

$$-\Delta u + qu = f \text{ on } \Omega, \quad u = 0 \text{ on } \partial\Omega, \tag{6.56}$$

where $\Omega \subset \mathbb{R}^2$ is a simple domain such as a rectangle, so that wavelet-based preconditioners can compete without struggling too much with technicalities.

Poisson-like problems

The tests in Ko et al. (1997), Lorentz and Oswald (1996) and Lorentz and Oswald (1997) indicate that, for the Poisson problem $q = 0$ in (6.56), the BPX scheme is superior to the wavelet-based methods, both with regard to the number of iterations needed to ensure a desired accuracy and to the cost of each iteration. As for the cost, this is obvious (recall Remark 6.7). Several types of wavelets, such as Daubechies wavelets, the above finite element-based wavelets, and so-called multi-wavelets, were included in the comparisons. For this kind of problem, the scheme based on Daubechies wavelets appears to be the weakest, since the cost per iteration is higher due to larger masks, while a careful study of corresponding condition numbers(Lorentz and Oswald 1997) shows that the frame bounds for the H_0^1-frames behind the BPX scheme are tighter than those of all wavelet bases. However, the finite element-based wavelets with small support come quite close. The condition numbers produced by the BPX- and by the coarse grid corrections (6.52) with $a = 1/6$ and $a = 3/16$ are reported to stay below 11 (Lorentz and Oswald 1997). The fact that the wavelets also form a Riesz basis in L_2 is not crucial in this case.

Helmholtz problems

The situation changes when $q > 0$ in (6.56) is increased. Now the additional zero-order term starts to affect stability. The efficiency of the BPX scheme in its original form starts to deteriorate. However, a suitable (inexpensive) modification, namely including a properly weighted zero-order term in the auxiliary form $b_{j,k}(\cdot,\cdot)$ has been observed to stabilize it (Oswald 1994). The condition numbers for the wavelet schemes are now smaller and the finite element-based wavelets with small support do quite well (Stevenson 1996, 1995*a*. These schemes are in that sense more *robust* for the class of prob-

lems (6.56). The reason is that, in contrast to the H_0^1-frames behind the BPX scheme, the wavelets form a Riesz basis for a larger range of Sobolev spaces including L_2, so that the zero-order term $q\,u$ is handled better. In particular, when q gets very large, the condition numbers for orthonormal wavelets tend to one, simply because the operator approaches (a multiple of) the identity. Eventually this starts to offset the higher cost per iteration. This is of particular importance for implicit discretizations of parabolic problems

$$\frac{\partial}{\partial t}u = \Delta u, \tag{6.57}$$

where for each time step an elliptic problem $\mathcal{L} = I - \Delta t\,\Delta$ has to be solved. Here wavelet preconditioners work well for a wide range of time steps without additional tuning.

The same robustness issue is also treated by Stevenson (1995*b*), using the wavelet (6.45) derived from the frequency decomposition approach mentioned in Section 6.6. It is shown to be superior to the BPX scheme for this type of problem with regard to efficiency and robustness.

Of course, when the solution is very smooth, the higher cost of a higher-order wavelet scheme per iteration may well be offset by the better approximation. Also, the effect of adaptive refinements has not been taken into account in the above comparisons.

Anisotropies

A similar observation can be made for problems of the type

$$-\varepsilon\frac{\partial^2 u}{\partial {x_1}^2} - \frac{\partial^2 u}{\partial {x_2}^2} + qu = f \text{ on } \Omega, \quad u = 0 \text{ on } \partial\Omega, \tag{6.58}$$

where Ω is again a rectangle for simplicity and ε is small. Such anisotropies aligned with coordinated lines arise, for instance, when employing boundary-adapted grids with high-aspect ratios in flow computations. Griebel and Oswald (1995*b*) compare multilevel Schwarz preconditioners based on tensor product pre-wavelets (see Figure 1) with nodal basis oriented splittings for problems of the type (6.58). Again, the latter method is typically twice as efficient as the wavelet scheme when using proper tuning, while the wavelet scheme is clearly more robust relative to varying ε and q in (6.58). Moreover, in the 3D case it still works in combination with *sparse grid* techniques.

The same issue is treated by Stevenson (1996) (for $q = 0$) with the aid of the modified frequency decomposition multilevel schemes discussed in Section 6.6. Moreover, triangular-based wavelets constructed via discrete inner products are applied by Stevenson (1995*a*) to several types of second-order elliptic boundary value problems with leading term div (A grad). In particular, the case where A is a piecewise constant diagonal matrix with large jumps is considered. The main result is to show that the proposed

wavelet-based multilevel scheme is robust for the class of problem where A exhibits possible anisotropies along the three directions of a type-2 grid, when this mesh is used for the discretization.

More detailed information about comparisons can be found in Ko et al. (1997), Lorentz and Oswald (1996) and Lorentz and Oswald (1997). In summary, it seems that the robustness issue is in favour of wavelet-based discretizations. One should not forget, though, that the comparisons did not include multiplicative multigrid schemes, which are usually more efficient than additive counterparts such as the BPX method. Of course, the story changes again when nonuniform grids and complicated domain geometries are considered. Classical wavelets no longer apply directly (see below for a two-stage approach), while the above coarse grid corrected wavelets are still defined.

Finally, a more interesting question occurs when adaptivity is employed, for instance, for a domain with reentrant corners. To my knowledge, direct comparisons have not yet been made. It will be seen later that wavelets seem to have a great potential in this regard.

7. The ideal setting

7.1. Preliminary remarks

Preconditioning is only *one* aspect of wavelet schemes. At least for positive order operators it does not require the full sophistication of wavelets, since simpler suitable frames are seen to work as well, often even better, and for more flexible meshes. In this sense, preconditioning puts only weak demands on the wavelet as a discretization tool. To exploit the full potential of wavelets one is led to ask for more. Two possible directions are, firstly, to consider operators such as integral operators whose conventional discretization gives rise to dense matrices, or, secondly, when dealing with differential operators, to try to *diagonalize* $\mathcal{L}$ in the sense of Section 1.5 (a). We will first address the latter issue. Of course, one cannot expect such an objective to be feasible under the most general circumstances. The basic rationale is to develop a two-stage process. First design highly efficient schemes for an *ideal setting* (ideal with regard to highest efficiency and availability of the tools) and then try to reduce realistic problems to the ideal case at an additional expense which, however, should be of lower order. As indicated in Sections 4.2 and 4.3, wavelets unfold their full potential when working on the whole Euclidean space $\mathbb{R}^n$ or on the n-torus $\mathbb{R}^n/\mathbb{Z}^n$. I will refer to this as the *ideal setting.*

A great deal of effort has been spent on studying elliptic operators $\mathcal{L}$ of the form (2.4), where $a > 0$ and $A(x)$ is a symmetric matrix satisfying

$$\sup_{x\in\mathbb{R}^n} \|A(x)\| < \infty, \quad \xi^{\mathrm{T}} A(x)\xi \geq \delta|\xi|^2, \quad x, \xi \in \mathbb{R}^n, \tag{7.1}$$

for some $\delta > 0$, and whose coefficients satisfy certain (weak) regularity properties. The subsequent discussion mainly reflects related work by Angeletti, Mazet and Tchamitchian (1997), Liandrat and Tchamitchian (1997) and Tchamitchian (1996). Typical questions to be studied are

(i) the continuity of $\mathcal{L}^{-1}$ in Sobolev scales $W^{s,p}$ depending on the coefficients of $A(x)$
(ii) the boundedness of associated Galerkin projections in this scale
(iii) efficient numerical procedures for the approximate solution of

$$\mathcal{L}u = f \quad \text{on } \mathbb{R}^n \text{ or } \mathbb{R}^n/\mathbb{Z}^n \tag{7.2}$$

with the aid of wavelet bases.

These questions are clearly closely interrelated. Our main concern here will be (iii), while recent studies of (i) and (ii) are given by Angeletti et al. (1997).

Recall from Section 4.2 the format of the biorthogonal wavelet bases $\Psi = \{\psi_\lambda : \lambda \in \nabla\}$ and likewise $\tilde{\Psi}$ to be used when $\Omega = \mathbb{R}^n$. By (4.38), one has to deal with $2^n - 1$ mother wavelets ψ_e, $e \in E_* := \{0,1\}^n \setminus \{0\}$. In this case it is convenient to take $j_0 = -\infty$ as the coarsest level so that $\nabla = \nabla_-$. When working on the torus it will always be tacitly assumed that $j_0 \geq 0$ is fixed. In either case, the indices $\lambda \in \nabla$ have the geometric interpretation $\lambda = 2^{-j}(k + \frac{e}{2})$, $k \in \mathbb{Z}^n$, $j \in \mathbb{Z}, j \geq j_0$, $e \in E_*$. We continue denoting by $d, \tilde{d}$ the order of polynomial exactness of the spaces $S(\Phi_j)$, $S(\tilde{\Phi}_j)$, so that the direct inequalities (5.1) hold with $d, \tilde{d}$, respectively.

The principal goal is to *diagonalize* the operator $\mathcal{L}$ in (7.2) as indicated in Section 1.5 (a). The relevant theoretical background is the theory of Calderón–Zygmund operators. To support the understanding of the subsequent developments I include some brief comments in this regard, mainly following Tchamitchian (1996).

7.2. Vaguelettes and Calderón–Zygmund operators

The subsequent discussion first follows the original development that has been tailored to orthonormal wavelets and therefore admits wavelets with global support as long as there is enough decay. Orthonormality is not crucial, though, and analogous statements can be made for the biorthogonal case as well. So assume that the ψ_e decay rapidly along with their derivatives up to some order $r_0 \geq 3$. For $|\alpha| = r_0$, $\partial^\alpha \psi_e$ is supposed to be defined almost everywhere while $\psi_e \in C^{r-1}(\mathbb{R}^n)$, $e \in \{0,1\}^n$.

In addition to the wavelet basis Ψ we consider another family $\Theta = \{\theta_\lambda : \lambda \in \nabla\}$ which is related to Ψ by

$$\mathcal{T}\Psi = \Theta, \tag{7.3}$$

for some linear operator $\mathcal{T}$. Ultimately, we will be interested in $\mathcal{T} = (\mathcal{L}^{-1})^*$. Unfortunately, the θ_λ will *not* arise from finitely many mother functions by means of dilation and translation. However, in cases of interest to us, the θ_λ still share the following properties with the ψ_λ, $\lambda \in \nabla$: there exist positive numbers C, q, r and a non-negative integer $\tilde{d} \in [0, q)$, such that for $\lambda \in \nabla_-$,

$$|\theta_\lambda(x)| \le C 2^{jn/2} \left(1 + 2^j |x - \lambda|\right)^{-n-q}, \quad x \in \mathbb{R}^n. \tag{7.4}$$

Furthermore, $\theta_\lambda \in C^{\lfloor r \rfloor}(\mathbb{R}^n)$, ($\lfloor a \rfloor$ being the largest integer less than or equal to a), and for all $\alpha \in \mathbb{Z}_+^n$, $|\alpha| \le \lfloor r \rfloor$ one has

$$|\partial^\alpha \theta_\lambda(x)| \le C 2^{j\left(\frac{n}{2} + |\alpha|\right)} \left(1 + 2^j |x - \lambda|\right)^{-n-q}, \quad x \in \mathbb{R}^n. \tag{7.5}$$

Moreover, for $|\alpha| = \lfloor r \rfloor$ the $\partial^\alpha \theta_\lambda$ are Hölder continuous, that is,

$$|\partial^\alpha \theta_\lambda(x + h) - \partial^\alpha \theta_\lambda(x)| \le C 2^{j\left(\frac{n}{2} + r\right)} |h|^{r - \lfloor r \rfloor} \left(1 + 2^j |x - \lambda|\right)^{-n-q}, \quad x \in \mathbb{R}^n, \tag{7.6}$$

and for every polynomial P of order at most $\tilde{d}$ one has

$$\int_{\mathbb{R}^n} P(x) \theta_\lambda(x)\, \mathrm{d}x = 0, \quad \lambda \in \nabla_-. \tag{7.7}$$

The set Θ is called a family of *vaguelettes* with index $(\tilde{d}, q, r)$. Note that the kernel of the operator $\mathcal{T}$ defined by (7.3) has the form

$$K(x, y) = \sum_{\lambda \in \nabla} \theta_\lambda(x) \overline{\psi_\lambda(y)}. \tag{7.8}$$

One can then show (see, for instance, Tchamitchian (1996)) that the estimates (7.4), combined with corresponding standard estimates for the wavelets, imply that there exist constants $C, \delta > 0$, such that

$$|K(x, y)| \le \frac{C}{|x - y|}, \quad x \ne y, \tag{7.9}$$

and

$$|K(x, y) - K(x', y)| + |K(y, x) - K(y, x')| \le C \frac{|x - x'|^\delta}{|x - y|^{1+\delta}}, \tag{7.10}$$

when $|x - x'| \le |x - y|/2$. An operator $\mathcal{T}$, such that for any two test functions f, g with disjoint compact supports

$$\langle \mathcal{T} f, g \rangle = \iint K(x, y) f(y) \overline{g(x)}\, \mathrm{d}y\, \mathrm{d}x,$$

is called a *Calderón–Zygmund operator* (CZO) if $\mathcal{T}$ is continuous on L_2 and K satisfies the so-called *standard estimates* (7.9), (7.10). The above notions are now interrelated by the following result, whose proof can be found in Meyer (1990).

Theorem 7.1 Suppose Θ is a family of vaguelettes with index $(\tilde{d}, q, r)$. Then $\mathcal{T}$ defined by (7.3) extends to a CZO. It is continuous on $W^{s,p}(\mathbb{R}^n)$ for $1 < p < \infty$ and $|s| < \inf(r_0, r, \tilde{d}+1, q)$. The corresponding operator norms depend on the parameters and on the constant C in (7.4)–(7.7).

Let us denote by $\mathcal{LMA}$ the collection of operators $\mathcal{T}$ such that $\Theta = \mathcal{T}\Psi$ is a family of vaguelettes.

Theorem 7.2

(i) $\mathcal{LMA}$ is independent of the choice of Ψ (provided that $r_0 > 0$).
(ii) $\mathcal{LMA}$ is an algebra which is stable under taking adjoints.
(iii) $\mathcal{LMA}$ is exactly the set of CZOs $\mathcal{T}$ such that $\mathcal{T}(1) = \mathcal{T}^*(1) = 0$.
(iv) $\mathcal{LMA}$ is not stable under taking inverses.

Assertions (i) to (iii) are due to Lemarié (1984); (iii) is related to the T1-Theorem by G. David and J. L. Journé, which characterizes the continuity of an operator $\mathcal{T}$ satisfying the estimates (7.9), (7.10). This result will be mentioned later again. As for (iv), we refer the reader to Tchamitchian (1996), and its implications will become clearer later on. Since a CZO takes $L_\infty(\mathbb{R}^n)$ into the space of functions of bounded mean oscillation (BMO), $\mathcal{T}(1)$ and $\mathcal{T}^*(1)$ are indeed defined. Recall that a locally integrable function f belongs to BMO if and only if, for any cube C, there exists a constant a such that

$$\frac{1}{|C|}\int_C |f(x) - a|\,\mathrm{d}x \lesssim 1,$$

where $|C|$ denotes the volume of C.

The following characterization of $\mathcal{LMA}$ in terms of matrices will be important in the present context. Let $\mathcal{M}$ denote the set of matrices $\mathbf{A}$ such that for some $\gamma > 0$ one has

$$|\mathbf{A}_{\lambda,\lambda'}| \lesssim \frac{2^{-\left||\lambda|-|\lambda'|\right|\left(\gamma+\frac{n}{2}\right)}}{\left(1+2^{\min\{|\lambda|,|\lambda'|\}}|\lambda-\lambda'|\right)^{1+\gamma}}. \tag{7.11}$$

Thus the entries of $\mathbf{A}$ decay with increasing difference in scale and spatial location of the indices λ, λ'. Recall that this estimate is of the type (1.11). In fact, essentially the same argument can be employed to show that, when $\int \theta_\lambda = 0, \quad \lambda \in \nabla$, then $\langle \Theta, \Theta \rangle \in \mathcal{M}$. Moreover, the following result can be found in Angeletti et al. (1997).

Theorem 7.3 $\mathcal{T}$ belongs to $\mathcal{LMA}$ if and only if the matrix $\langle \Psi, \mathcal{T}\Psi \rangle$ belongs to $\mathcal{M}$.

Next, we will describe an approach, initiated by Tchamitchian (1987), which is based on the above concepts and aims at avoiding the solution of linear systems essentially by *diagonalizing* the operator.

7.3. Constant coefficient operators

The basic strategy is to proceed again in two steps: first treat carefully the case of *constant coefficient* operators and then use a *freezing coefficient* technique. We will specify the operator $\mathcal{L}$ from (2.2) first to the special case where $a_{\alpha\beta}(x) = a_{\alpha\beta}$ are constant so that

$$\mathcal{L}u = -\sum_{|\alpha|,|\beta|=1} a_{\alpha\beta}\partial^\alpha\partial^\beta + a. \tag{7.12}$$

The operator $\mathcal{L}$ will always be assumed to be elliptic, which here means that the principal part $\sigma_0(y)$ of its *symbol* σ is strictly positive on $\mathbb{R}^n$, that is,

$$\sigma_0(y) := \sum_{|\alpha|,|\beta|=1} a_{\alpha\beta}y^{\alpha+\beta} \geq \delta > 0, \quad y \in \mathbb{R}^n, \quad \sigma(y) = \sigma_0(y) + a. \tag{7.13}$$

We follow Liandrat and Tchamitchian (1997) and try to solve $\mathcal{L}u = f$ conceptually by applying the inverse $\mathcal{L}^{-1}$ to the right-hand side f. Although at first glance this may contradict basic principles in numerical analysis, it does have tempting aspects, as shown next.

Suppose that $\Psi, \tilde{\Psi}$ are biorthogonal Riesz bases in $L_2(\mathbb{R}^n)$. Then the solution u of (7.2) has the form $u = \mathbf{d}^T\Psi$ with unknown coefficient sequence $\mathbf{d}^T = \langle u, \tilde{\Psi}\rangle$. Inserting $u = \mathcal{L}^{-1}f$, one obtains

$$\mathbf{d}^T = \langle u, \tilde{\Psi}\rangle = \langle \mathcal{L}^{-1}f, \tilde{\Psi}\rangle = \langle f, (\mathcal{L}^{-1})^*\tilde{\Psi}\rangle = \langle f, \Theta\rangle, \tag{7.14}$$

that is, the roles of $\mathcal{T}$ and Ψ in (7.3) are played here by $(\mathcal{L}^{-1})^*$ and $\tilde{\Psi}$, respectively. Thus the solution u of $\mathcal{L}u = f$ is formally given as

$$u = \langle f, \Theta\rangle\Psi. \tag{7.15}$$

Proposition 7.4 (Angeletti et al. 1997) The collection Θ, defined by $\Theta = (\mathcal{L}^{-1})^*\tilde{\Psi}$, is a family of vaguelettes.The constant C in (7.4)–(7.7) depends on the $a_{\alpha\beta}$ and $\Psi, \tilde{\Psi}$ but *not* on a.

Thus the image of $\tilde{\Psi}$ under $(\mathcal{L}^{-1})^*$ still has nice localization properties reflected by estimates (7.4)–(7.7). This suggests the following approach (Liandrat and Tchamitchian 1997).

A projection scheme

A natural idea is to compute an approximate solution of (7.2) by truncating $\langle f, \Theta\rangle$. Fixing *any* finite $\Lambda \subset \nabla$, this corresponds to projecting u into the finite-dimensional space $S_\Lambda = S(\Psi_\Lambda)$ $\Psi_\Lambda = \{\psi_\lambda : \lambda \in \Lambda\}$, (see (3.43)), that is,

$$u_\Lambda = \langle u, \tilde{\Psi}_\Lambda\rangle\Psi_\Lambda = \langle f, \Theta_\Lambda\rangle\Psi_\Lambda. \tag{7.16}$$

Note that this is a Petrov–Galerkin approximation (6.2).

Convergence
Under assumptions (7.13), $\mathcal{L}$ is also a boundedly invertible mapping from $H^{2+s}(\mathbb{R}^n)$ to $H^s(\mathbb{R}^n)$, $s \in \mathbb{R}$. Thus, noting that, by (7.15), $u_\Lambda = Q_\Lambda u = Q_\Lambda \mathcal{L}^{-1} f$, one has $u_\Lambda - u = (Q_\Lambda - I)\mathcal{L}^{-1} f$. Employing our direct estimates (5.1) (for $p = 2$), we obtain

$$\|u_\Lambda - u\|_{L_2} = \left\|(Q_\Lambda - I)\mathcal{L}^{-1} f\right\|_{L_2} \lesssim 2^{-dj(\Lambda)} \|\mathcal{L}^{-1} f\|_{H^d}, \tag{7.17}$$

where $j(\Lambda) = \max\{j : |\lambda| < j \Rightarrow \lambda \in \Lambda\}$. Continuity of $\mathcal{L}^{-1}$ in the Sobolev scale gives

$$\|u_\Lambda - u\|_{L_2} \lesssim 2^{-dj(\Lambda)} \|f\|_{H^{d-2}}. \tag{7.18}$$

Analogous estimates for the spaces $W^{s,p}$ are obtained in exactly the same way as long as one has continuity of $\mathcal{L}^{-1}$ (see question (ii) in the Section 7.1). Moreover, standard interpolation arguments yield

$$\|u_\Lambda - u\|_{L_2} \lesssim 2^{-sj(\Lambda)} \|f\|_{H^{s-2}}, \quad d \geq s \geq 2. \tag{7.19}$$

Estimates of this type are very crude. They guarantee convergence as long as the spaces S_Λ include sufficiently many low frequencies, that is, $j(\Lambda)$ grows with $\#\Lambda$. The interesting part, of course, concerns the adaptation of Λ to the problem at hand, which may result in a selection of highly nonuniformly shaped subsets $\Lambda \subset \nabla$. In view of (7.16), this is closely related to the next point.

7.4. *Evaluation of* $\langle f, \Theta_\Lambda \rangle$

By (7.15), the success of the approach hinges on identifying and computing the *significant* coefficients of $\langle f, \Theta \rangle$, represented here by the finite array $\langle f, \Theta_\Lambda \rangle$. The important point is that, by Proposition 7.4, $\mathcal{L}^{-*}$ is a CZO; here and elsewhere, we shall write $\mathcal{L}^{-*}$ instead of $(\mathcal{L}^*)^{-1}$. Hence, according to (7.11), $\mathcal{L}^{-*}$ has a quasi sparse matrix representation.

Noting that $(\mathcal{L}u)\hat{}(y) = \sigma(y)\hat{u}(y)$, where σ is by (7.13) strictly positive on $\mathbb{R}^n$, the definition (7.14) of Θ means that

$$\langle f, \theta_\lambda \rangle = \frac{1}{(2\pi)^n} \langle \hat{f}, \hat{\theta}_\lambda \rangle = \frac{1}{(2\pi)^n} \langle \hat{f}, \bar{\sigma}^{-1} \hat{\tilde{\psi}}_\lambda \rangle.$$

Since the wavelets are well localized in Fourier space, one could employ quadrature to compute $\langle \hat{f}, \bar{\sigma}^{-1} \hat{\tilde{\psi}}_\lambda \rangle$ up to any desired precision.

Projections into $\mathcal{S}$ and convolutions
We now describe an alternative approach proposed by Liandrat and Tchamitchian (1997). Due to the vaguelette estimates, θ_λ belongs up to a desired tolerance to $S(\tilde{\Phi}_{|\lambda|+p})$ for some $p \in \mathbb{N}$, which depends on $\mathcal{L}$, Ψ and $\tilde{\Psi}$ but *not* on $|\lambda|$. This suggests projecting f into the space $S(\Phi_{|\lambda|+p})$. As before,

we will always denote by Q_j or Q_Λ the canonical projectors onto the spaces $S(\Phi_j)$, $S(\Psi_\Lambda)$, respectively. This suggests replacing $\langle f, \theta_\lambda\rangle$ by $\langle Q_{|\lambda|+p} f, \theta_\lambda\rangle$. In fact, since

$$|\langle f, \theta_\lambda\rangle - \langle Q_j f, \theta_\lambda\rangle| \le \|f\|_{L_2} \|(I - Q_j^*)\theta_\lambda\|_{L_2},$$

the precision depends on the approximation properties of the spaces $S(\tilde{\Phi}_j)$ and on the regularity of the θ_λ and hence of the ψ_λ. Now suppose $|\lambda| = j$ and write

$$Q_{j+p} f = Q_{j+1} f + \sum_{l=1}^{p-1} (Q_{j+l+1} - Q_{j+l})\, f = \mathbf{c}_{j+1}^{\mathrm{T}} \Phi_{j+1} + \sum_{l=1}^{p-1} \mathbf{d}_{j+l}^{\mathrm{T}} \Psi_{j+l},$$

where $\mathbf{c}_{j+1}^T = \langle f, \tilde{\Phi}_{j+1}\rangle$, $\mathbf{d}_{j+l}^T = \langle f, \tilde{\Psi}_{j+l}\rangle$, so that

$$\langle Q_{j+p} f, \theta_\lambda\rangle = \mathbf{c}_{j+1}^{\mathrm{T}} \langle \Phi_{j+1}, \theta_\lambda\rangle + \sum_{l=1}^{p-1} \mathbf{d}_{j+l}^{\mathrm{T}} \langle \Psi_{j+l}, \theta_\lambda\rangle. \tag{7.20}$$

This amounts to discrete convolution of the wavelet coefficients of f with filters that depend only on Ψ and $\mathcal{L}$. To compute these filters once and for all, one can again resort to Fourier transforms.

Moreover, if the right-hand side f is smooth except at isolated points, then only a small number of the coefficients $d_\lambda = \langle f, \tilde{\psi}_\lambda\rangle$ exceed a given threshold in magnitude. The sequences $\langle \Psi_{j+l}, \theta_\lambda\rangle$, $l = 1, \dots, p-1$, describe how the wavelet coefficients d_λ, $|\lambda| = j + l$, are *smeared* by the application of $\mathcal{L}^{-*}$. Thus (7.20) is to be applied to the *compressed* arrays of wavelet coefficients, which result from thresholding.

Let us first add a few comments on the structure of the sequences $\langle \Psi_{j+l}, \theta_\lambda\rangle$ which can be viewed as one column of the matrix $\langle \Psi_{j+l}, \Theta_j\rangle$ where $\Theta_j := \{\theta_\lambda : |\lambda| = j\}$. Recalling the two-scale relations (3.4) and (3.8), $\Psi_j^{\mathrm{T}} = \Phi_{j+1}^{\mathrm{T}} \mathbf{M}_{j,1}$, (which is here stationary in j but would sizewise depend on j in the periodic case), one obtains

$$\begin{aligned} \langle \Psi_{j+l}, \Theta_j\rangle &= \langle \Psi_{j+l}, \mathcal{L}^{-*} \tilde{\Psi}_j\rangle = \mathbf{M}_{j+l,1}^{\mathrm{T}} \langle \Phi_{j+l+1}, \mathcal{L}^{-*} \tilde{\Phi}_{j+1}\rangle \tilde{\mathbf{M}}_{j,1} \\ &= \mathbf{M}_{j+l,1}^{\mathrm{T}} \langle \Phi_{j+l+1}, \mathcal{L}^{-*} \tilde{\Phi}_{j+1+l}\rangle \tilde{\mathbf{M}}_{j+l,0} \cdots \tilde{\mathbf{M}}_{j+1,0} \tilde{\mathbf{M}}_{j,1}. \end{aligned} \tag{7.21}$$

Thus, once the arrays $\mathbf{F}_q := \langle \Phi_q, \mathcal{L}^{-*} \tilde{\Phi}_q\rangle$ are known, the matrices $\langle \Psi_{j+l}, \Theta_j\rangle$ are obtained with the aid of pyramid-type schemes like (3.26). Moreover, a typical entry of $\mathbf{F}_q$ has the form

$$\begin{aligned} \langle \phi_{q,k}, \mathcal{L}^{-*} \tilde{\phi}_{q,l}\rangle &= (2\pi)^{-n} \langle \hat{\phi}_{q,k}, \bar{\sigma}^{-1} \hat{\tilde{\phi}}_{q,l}\rangle \\ &= (2\pi)^{-n} \int_{\mathbb{R}^n} \hat{\phi}(y) \overline{\hat{\tilde{\phi}}(y)} \sigma(2^q y)^{-1} e^{-iy\cdot(k-l)}\, \mathrm{d}y. \end{aligned} \tag{7.22}$$

As for the cost of these operations, let us consider the periodic case (see Section 4.3), where J denotes the highest discretization level and p as above is fixed. Suppose that the arrays $\mathbf{F}_q$ have been (approximately) computed once and for all in an *initialization step*. Note that the fast decay of $\sigma(2^q y)^{-1}$ for large q implies that $\mathbf{F}_q$ can be ever better approximated by a bounded matrix with small bandwidth. Hence the computation of all vaguelette coefficients on level r boils down to

$$\langle Q_{r+p} f, \Theta_r \rangle = \mathbf{c}_{r+1}^{\mathrm{T}} \mathbf{F}_{r+1} \tilde{\mathbf{M}}_{r,1} + \sum_{l=1}^{p-1} \mathbf{d}_{r+l}^{\mathrm{T}} \mathbf{M}_{r+l,1}^{\mathrm{T}} \mathbf{F}_{r+l+1} \tilde{\mathbf{M}}_{r+l,0} \cdots \tilde{\mathbf{M}}_{r+1,0} \tilde{\mathbf{M}}_{r,1}. \tag{7.23}$$

Thus, when the wavelets have compact support and each $\mathbf{F}_{r+l+1}$ is replaced by a sparse matrix, this requires the order of $2^{(r+p)n}$ (n being the spatial dimension) operations, where the constant depends on $\Psi, \tilde{\Psi}$ and the accuracy of $\mathbf{F}_{r+l+1}$. Consequently, the computation of u_Λ is of the order $2^{(p+|\Lambda|)n} \sim N_\Lambda$, $N_\Lambda := \dim S(\Phi_{|\Lambda|})$, where $|\Lambda| := \max\{|\lambda| + 1 : \lambda \in \Lambda\}$, with a constant depending on p. Note that when rapidly decaying wavelets with global support (but very good localization in Fourier space – see Section 4.2) are used, the matrices $\tilde{\mathbf{M}}_{j,0}$, $\tilde{\mathbf{M}}_{j,1}$ are no longer sparse but, in the periodic case, are circulants, so that FFT can be employed to limit the order of operations to $N_\Lambda \log N_\Lambda$.

On the other hand, the above work estimate has been very crude. In fact, when the right-hand side f is smooth except at isolated points, only very few of its wavelet coefficients d_λ are expected to exceed a given tolerance $\varepsilon > 0$. Due to the localization and cancellation properties of the vaguelettes θ_λ, the coefficients $\langle f, \theta_\lambda \rangle$ are expected to exhibit similar behaviour. In fact, since $\mathcal{L}^{-*}$ is a CZO the decay of the entries in $\langle \Psi_{j+l}, \Theta_j \rangle$ is governed by estimates of the form (7.11), and the *spread* of the wavelet coefficients of f due to $\mathcal{L}^{-*}$ can be seen from (7.20). This suggests computing $\langle f, \theta_\lambda \rangle$ or, better, $\langle Q_{|\lambda|+p} f, \theta_\lambda \rangle$ only for those λ in a certain neighbourhood of the significant coefficients of f. The number of these coefficients may, of course, be much smaller than $\dim S(\Phi_{|\Lambda|})$. A more formal treatment of this issue in Liandrat and Tchamitchian (1997) is based on the notion of (ε, s)-*adapted spaces*.

Nevertheless, it does not appear to be completely obvious how to carry out all computations without requiring the full complexity of the highest discretization level at some point. In fact, while thresholding the arrays $\mathbf{d}_{r+l}$ on the right-hand side of (7.23) facilitates the successive multiplication with possibly very short vectors, the first summand involving $\mathbf{c}_{r+1}$ does not seem to be compressible in this form.

Remark 7.5 Another point concerns the various tolerances in the above procedure. Uniformity of the work estimates in $\#\Lambda$ are ultimately of limited value when the involved tolerances and thresholds are kept *fixed*. In

fact, increasing $\#\Lambda$ should produce better overall accuracy. Correspondingly tighter thresholds, in turn, are expected to require a larger p, and hence a higher computational cost that may no longer stay proportional to $\#\Lambda$. Questions of this form will be encountered in similar contexts several times.

A hybrid scheme

There exist several variants of the above scheme (see, for instance, Ponenti (1994)) among which I would like to mention the *hybrid* scheme proposed by Fröhlich and Schneider (1995), which differs from the above procedure in an essential way. The main point in Fröhlich and Schneider (1995) is to economize the evaluation of the vaguelette coefficients $\langle f, \theta_\lambda \rangle$ by incorporating *interpolation techniques.* Again it is designed for the periodic case. For simplicity, we consider bases on $\mathbb{R}$ and refer the reader to Section 4.3 for standard periodization (see also Fröhlich and Schneider (1995)) and tensor product versions for the bivariate situation.

Let $\mathcal{L} = \sum_{m=0}^{s} a_m \left(\frac{d}{dx}\right)^m$ be an elliptic operator; that is, as before, its symbol $\sigma(y) := \sum_{m=0}^{s} a_m (iy)^m$ is strictly positive on $\mathbb{R}$. By construction, the family $\tilde{\Theta} := \mathcal{L}\Psi$ is biorthogonal to $\Theta = (\mathcal{L}^{-1})^* \tilde{\Psi}$. Hence one has

$$f = \langle f, \Theta \rangle \tilde{\Theta}. \tag{7.24}$$

Thus, instead of approximating f first by projecting into the spaces $S(\Phi_j)$, as in the previous approach, one could try to expand f approximately with respect to $\tilde{\Theta}$. Thus, consider the spaces

$$S_{\mathcal{L},J} := S(\{\tilde{\theta}_\lambda \in \tilde{\Theta} : |\lambda| < J\}) = S(\mathcal{L}\Phi_J). \tag{7.25}$$

The idea is now to employ *Lagrange interpolation* to efficiently obtain an approximation to f in $S_{\mathcal{L},J}$, say. Therefore one is interested in finding the *fundamental Lagrange functions*

$$L_j(x) = \sum_{k \in \mathbb{Z}} g_k \left(\mathcal{L}\phi_{j,k}\right)(x) = \mathbf{g}^{\mathrm{T}} \mathcal{L}\Phi_j(x), \tag{7.26}$$

such that

$$L_j(2^{-j}k) = \delta_{0,k}, \quad k \in \mathbb{Z}. \tag{7.27}$$

This is equivalent to saying that

$$1 = \sum_{k \in \mathbb{Z}} \langle L_j(2^{-j}\cdot), \delta(\cdot - k) \rangle e^{iky} = 2^j \sum_{k \in \mathbb{Z}} \hat{L}_j(2^j(y + 2\pi k)).$$

Standard arguments (Dahmen et al. 1994*c*, Fröhlich and Schneider 1995) yield

$$\hat{L}_j(y) = \frac{\sigma(y)\hat{\phi}(2^{-j}y)}{\sum_{k \in \mathbb{Z}} \sigma(2^j y + 2^j 2\pi k)\hat{\phi}(y + 2\pi k)}, \tag{7.28}$$

which, of course, requires the sum in the denominator to be nonzero.

Recall that, by (7.14), the solution u of $\mathcal{L}u = f$ is given by

$$u = \langle f, \mathcal{L}^{-*}\tilde{\Phi}_0\rangle\Phi_0 + \langle f, \Theta\rangle\Psi. \tag{7.29}$$

In order to project the right-hand side f first into $S_{\mathcal{L},J}$, one can use the samples of f at $2^{-J}k$. In view of (7.27),

$$f_J(x) := \sum_{k\in\mathbb{Z}} f(2^{-J}k)L_J(x - 2^{-J}k) \tag{7.30}$$

interpolates f in $S_{\mathcal{L},J}$. Thus, to obtain an approximation to the coefficients in $\langle f, \Theta^J\rangle$, say, one has to rewrite f_J in (7.30), in view of (7.24), in terms of $\tilde{\Theta}^J = \mathcal{L}\Psi^J$. Since this is the central point, we describe this change of bases in a little more detail. Let $\mathbf{L}_j = (L_j(\cdot - 2^{-j}k) : k \in \mathbb{Z})$ and define

$$\begin{aligned} \mathbf{D}_j &:= \langle \mathbf{L}_j, \Theta_{j-1}\rangle, \quad j = 1, \dots, J, \quad \mathbf{D}_{-1} := \langle \mathbf{L}_0, (\mathcal{L}^{-1})^*\tilde{\Phi}_0\rangle, \\ \mathbf{f}_j &:= (f_j(2^{-j}k) : k \in \mathbb{Z}). \end{aligned} \tag{7.31}$$

Since $f_J \in S_{\mathcal{L},J}$, it can be expanded as

$$f_J = \mathbf{f}_J^{\mathrm{T}}\mathbf{L}_J = \langle \mathbf{f}_J^{\mathrm{T}}\,\mathbf{L}_J, \mathcal{L}^{-*}\tilde{\Phi}_J\rangle\mathcal{L}\Phi_J = \mathbf{f}_J^{\mathrm{T}}\,\langle \mathbf{L}_J, \mathcal{L}^{-*}\tilde{\Phi}_J\rangle\mathcal{L}\Phi_J.$$

Now, combining (3.11) and (3.13) with Proposition 3.8, one obtains $\tilde{\Phi}_J^T = \tilde{\Phi}_{J-1}^T\mathbf{M}_{J-1,0}^* + \tilde{\Psi}_{J-1}^T\mathbf{M}_{J-1,1}^*$. Substituting this into the above relation and using (3.12) yields

$$f_J = \mathbf{f}_J^{\mathrm{T}}\,\langle \mathbf{L}_J, \mathcal{L}^{-*}\Phi_{J-1}\rangle\mathcal{L}\tilde{\Phi}_{J-1} + \mathbf{f}^{\mathrm{T}}\,\langle \mathbf{L}_J, \Theta_{J-1}\rangle\tilde{\Theta}_{J-1}, \tag{7.32}$$

where

$$f_{J-1} := \mathbf{f}_J^{\mathrm{T}}\langle \mathbf{L}_J, \mathcal{L}^{-*}\tilde{\Phi}_{J-1}\rangle\mathcal{L}\Phi_{J-1} \in S_{\mathcal{L},J-1}. \tag{7.33}$$

Thus we have determined the vaguelette coefficients $\langle f, \Theta_{J-1}\rangle \approx \langle f_J, \Theta_{J-1}\rangle$ of f relative to $\tilde{\Theta}$ as

$$\mathbf{d}_{J-1} = \mathbf{f}_J^{\mathrm{T}}\,\mathbf{D}_J, \tag{7.34}$$

where $\mathbf{D}_J$ is given by (7.31). To continue this process, one only has to determine the samples of f_{J-1} defined in (7.33) on the coarse grid, that is,

$$f_{J-1}(2^{-J+1}k) = f_J(2^{-J}2k) - \mathbf{d}_{J-1}^{\mathrm{T}}\tilde{\Theta}_{J-1}(2^{-J}k), \tag{7.35}$$

once one has computed $\tilde{\Theta}_{J-1} = \mathcal{L}\Psi_{J-1}$. Instead of performing (7.35) exactly, one can discard entries in $\mathbf{d}_{J-1}$ that stay below a certain threshold, to generate step-by-step compressed vectors $\mathbf{d}_j$, $j < J$, such that

$$u_J := \mathbf{c}_0^{\mathrm{T}}\Phi_0 + \sum_{0\le j<J} \mathbf{d}_j^{\mathrm{T}}\Psi_j$$

approximates u in (7.29). The following algorithm, from Fröhlich and Schneider (1995), does exactly that.

ALGORITHM 2 (FS)
Initialization: Compute the filters $\mathbf{D}_j$ and $\mathcal{L}\phi_{0,0}, \mathcal{L}\psi_{j,0}$, $j = 0, \ldots, J-1$.

(1) Set $j = J$ and determine $f_J(2^{-J}k) = f(2^{-J}k)$, $k \in \mathbb{Z}$.
(2) For $j = J-1, J-2, \ldots, 0$,

$$\begin{aligned} \mathbf{d}_j &= \mathbf{f}_{j+1}^{\mathrm{T}} \mathbf{D}_{j+1}, \\ \mathbf{f}_j &= (f_j(2^{-j}k) : k \in \mathbb{Z}), \quad f_j(2^{-j}k) = f_{j+1}(2^{-j}k) - \mathbf{d}_j^{\mathrm{T}} \tilde{\Theta}_j(2^{-j-1}k). \end{aligned}$$

(3) Compute $\mathbf{c}_0^{\mathrm{T}}$ with the aid of the filter in $\langle \mathbf{L}_0, \mathcal{L}^{-*}\tilde{\Phi}_0 \rangle$.

A few comments on this scheme are in order.

- When the change of bases is done exactly and no thresholding is applied, the above scheme is a collocation scheme.

- Instead of starting with a set Φ_j of orthonormal scaling functions as in Fröhlich and Schneider (1995), we have kept the flexibility of using biorthogonal pairs $\Phi_j, \tilde{\Phi}_j$. In Fröhlich and Schneider (1995), orthogonality was paid for by infinite masks which require additional truncation. Starting with biorthogonal spline wavelets (4.25), the collection $\tilde{\Theta}$ still consists of compactly supported functions. Likewise the representation of u_J involves only the compactly supported functions in Ψ and Φ_0. This might favour embedding techniques for more general domains. It is clear that the L_j have typically global support but decay exponentially. Here the actual computation requires a truncation. Of course, the $\mathbf{D}_j$ are obtained by computing only *one* mask, which involves truncation of the vaguelettes too. The matrix formulation for the periodic case is identical once $\mathbb{Z}$ is replaced by $\mathbb{Z}/2\mathbb{Z}$.

- In Fröhlich and Schneider (1995), it is assumed that a reduced set $\Lambda \subset \{\lambda : |\lambda| < J\}$ is given from the start. The above algorithm is formulated there in a way that takes advantage of this data reduction. This requires *a priori* knowledge about the solution u. Such information is often available when dealing with *time-dependent problems* and an initial guess of Λ can be obtained from the approximation on the previous time level. In this case the samples of f are not required on the full grid of level J. This can be incorporated above as well by requiring samples only at places determined by significant vaguelette coefficients.

The scheme is applied by Fröhlich and Schneider (1996) to Helmholz-type problems as well as to nonlinear parabolic PDEs and to the computation of flame fronts. The experiments indicate dramatic savings if the computation can be fully confined to the significant wavelet contributions.

7.5. Freezing coefficients

The numerical feasibility of the above vaguelette schemes hinges in an essential way on the constant coefficient model problem. Let us now sketch some ideas from Lazaar, Liandrat and Tchamitchian (1994) about how to extend these techniques to the case of non-constant coefficients. Roughly speaking, an exact solver on a coarse scale is employed in conjunction with a *freezing coefficient/vaguelette* scheme on higher scales. For simplicity, we consider the univariate case $n = 1$ only, that is, $\mathcal{L} = \mathcal{I} - \frac{\partial}{\partial x}\left(\nu(x)\frac{\partial}{\partial x}\right)$ where $\nu(x) \geq \nu > 0$ is Lipschitz continuous. Here we have $\mathcal{L} = \mathcal{L}^*$.

The objective is to evaluate a projection of the inverse $\mathcal{L}^{-1}$. Consider the Galerkin projection of the low-frequency part

$$\mathcal{A}_q := Q_q(Q_q^*\mathcal{L}Q_q)^{-1}Q_q^* \tag{7.36}$$

of the inverse, where $Q_q f = \langle f, \tilde{\Phi}_q\rangle\Phi_q$. Due to the variable coefficient $\nu(x)$, the evaluation of $\mathcal{L}^{-1}\tilde{\psi}_\lambda$ is not feasible where, as before, $\lambda = 2^{-j}\left(k + \frac{1}{2}\right)$. Instead one defines functions θ_λ by

$$-\nu(\lambda)\frac{d^2}{dx^2}\theta_\lambda = \tilde{\psi}_\lambda, \quad \lambda \in \nabla, \tag{7.37}$$

which, according to the preceding discussion, are vaguelettes (see Tchamitchian (1997)), so that the operator $\mathcal{P}_q$ defined by

$$\mathcal{P}_q(\tilde{\psi}_\lambda) = \begin{cases} \theta_\lambda, & |\lambda| \geq q, \\ 0, & |\lambda| < q, \end{cases} \tag{7.38}$$

is a bounded mapping from $L_2(\mathbb{R})$ to $H^2(\mathbb{R})$. Here, $\mathcal{P}_q$ is often termed a *parametrix* of $\mathcal{L}$, that is, the *exact* inverse of an approximation to $\mathcal{L}$ at high frequencies. In fact, by definition, one has, for $|\lambda| \geq q$,

$$\begin{aligned} \mathcal{L}\mathcal{P}_q(\tilde{\psi}_\lambda) &= \Theta_\lambda - \frac{d}{dx}(\nu(\cdot) - \nu(\lambda))\frac{d}{dx}\Theta_\lambda - \nu(\lambda)\frac{d^2}{dx^2}\Theta_\lambda \\ &=: \mathcal{R}_q(\tilde{\psi}_\lambda) + \tilde{\psi}_\lambda, \end{aligned}$$

while $\mathcal{L}\mathcal{P}_q(\tilde{\psi}_\lambda) = 0$ for $|\lambda| < q$. Hence one obtains

$$\mathcal{L}\mathcal{P}_q = (\mathcal{I} - Q_q) + \mathcal{R}_q, \tag{7.39}$$

and one can show that (Lazaar et al. 1994)

$$\|\mathcal{R}_q g\|_{L_2} \lesssim 2^{-q}\|g\|_{L_2}. \tag{7.40}$$

Now $\mathcal{A}_q + \mathcal{P}_g$ is expected to approximate $\mathcal{L}^{-1}$ well. In fact, a von Neumann series argument yields the following theorem.

Theorem 7.6 (Lazaar et al. 1994) Let $\mathcal{U}_q := \mathcal{I} - \mathcal{L}(\mathcal{A}_q + \mathcal{P}_q)$. Then $\|\mathcal{U}_q^2\|_{L_2} \lesssim 2^{-q}$ so that, for q sufficiently large (depending on $\Psi, \tilde{\Psi}$ and ν),

$$\mathcal{L}^{-1} = (\mathcal{A}_q + \mathcal{P}_q) \sum_{l=0}^{\infty} \mathcal{U}^l. \tag{7.41}$$

As for the numerical realization, choose some $J > q$. The idea is to replace the role of L_2 in the above scheme by $S_J = S(\Phi_J)$, that is, let $\mathcal{L}_J := Q_J^* \mathcal{L} Q_J$, and denote by $\mathcal{A}_{J,q}$ the Galerkin projection obtained by replacing $\mathcal{L}$ in (7.36) by $\mathcal{L}_J$.

The next step is to approximate $\mathcal{L}_J^{-1}$ in the neighbourhood of each wavelet on high scales by $\mathcal{P}_{J,q}$ defined by

$$\mathcal{P}_{J,q} g := \sum_{q \le |\lambda| < J} \langle g, \psi_\lambda \rangle \tau_\lambda,$$

where the τ_λ are here defined by

$$-\nu(\lambda) Q_J^* \partial_x^2 Q_J \tau_\lambda = \tilde{\psi}_\lambda, \quad |\lambda| \ge q,$$

As above, $\mathcal{A}_{J,q} + \mathcal{P}_{J,q}$ approximates $\mathcal{L}_J^{-1}$ on S_J. One can now formulate an analogue to Theorem 7.6, setting $\mathcal{U}_{J,q} = \mathcal{I} - \mathcal{L}_p(\mathcal{A}_{J,q} + \mathcal{P}_{J,q})$, so that

$$\mathcal{L}_J^{-1} = (\mathcal{A}_{J,q} + \mathcal{P}_{J,q}) \sum_{l \ge 0} \mathcal{U}_{J,q}^l.$$

Thus the solution of $\mathcal{L}u = f$ in S_J is given by

$$u_J = \sum_{k \ge 1} f_k, \tag{7.42}$$

where

$$f_k := (\mathcal{I} - (\mathcal{A}_{J,q} + \mathcal{P}_{J,q})\mathcal{L}) f_{k-1}, \quad k \ge 1,$$

and $f_1 := (\mathcal{A}_{J,q} + \mathcal{P}_{J,q}) f$. An approximate solution in S_J is obtained by truncating the series (7.42).

Note that, in view of (5.16), $\mathcal{A}_{J,q}$ is defined by

$$Q_q^* (Q_J^* \mathcal{L} Q_J) Q_q \mathcal{A}_{J,q} = Q_q^* \mathcal{L} Q_q \mathcal{A}_{J,q} = \mathcal{I},$$

so that the application of $\mathcal{A}_{J,q}$ requires solving the (small) linear system in $S_q \subset S_J$. However, the discretization of $\mathcal{L}$ that involves inner products with non-constant coefficients has to have the accuracy of the highest discretization level J in order not to spoil the overall accuracy. The application of $\mathcal{P}_{J,q}$ again requires a sufficiently accurate evaluation of the vaguelettes τ_λ. Here the remarks of the preceding discussion apply. Again, in the periodic setting wavelets with global support are usually admitted at the expense of an additional log term introduced by FFT. Employing compactly supported wavelets and approximating the vaguelettes as indicated before, one

may still hope to keep the computational work proportional to $\dim S_J$. Of course, the practical realization involves several approximation and truncation steps depending on the choice of J, q and $\Psi, \tilde{\Psi}$, and these have to be carefully balanced. One has to keep in mind that the general philosophy is to spend quite some effort on initialization and precomputation in order to reduce the solution to rapid evaluation schemes. It should be interesting to compare the scheme with conventional preconditioning schemes. Numerical tests for a periodic model problem show rapid convergence of the scheme (Lazaar et al. 1994). Details of the numerical schemes, their analysis and numerical experiments are presented in Lazaar (1995).

7.6. Energy pre-wavelets

One drawback of the above vaguelette schemes is that even when biorthogonal pairs $\Psi, \tilde{\Psi}$ of compactly supported wavelets are used, the collections Θ generally involve globally supported functions. This can be remedied in certain cases at the expense of exact diagonalization. In fact, consider again a constant coefficient elliptic operator $\mathcal{L} = \sum_{|\alpha|,|\beta|\le s} a_{\alpha\beta}\partial^\alpha\partial^\beta$ with strictly positive symbol σ. Suppose that $\phi \in L_2(\mathbb{R}^n)$, $n \le 3$, is a stable generator (see (4.9)), which is smooth enough to satisfy

$$\sum_{k\in\mathbb{Z}^n} \left|\phi\left(\frac{y+4\pi k}{2}\right)\right|^2 \sigma(y+4\pi k) \sim 1, \tag{7.43}$$

and $\int_{[0,1]^n} \left(\sum_{k\in\mathbb{Z}^n} |\phi(x-k)|\right)^{-2} \mathrm{d}x < \infty$. Moreover, assume that ϕ is skew-symmetric about some point $\alpha \in \mathbb{R}^n$, that is, $\phi(\alpha + x) = \overline{\phi(\alpha - x)}$, $x \in \mathbb{R}^n$. It was shown by Dahlke and Weinreich (1994) (see also Dahlke (1996) and Dahlke and Weinreich (1993)) that there exist $\psi_e \in S(\Phi_1)$, $e \in E_* = \{0,1\}^n \setminus \{0\}$, such that

$$S(\Phi_1) = S(\Phi_0) \oplus S\Big(\{\psi_e(\cdot - k) : k \in \mathbb{Z}^n, e \in E_*\}\Big),$$

and

$$\langle \mathcal{L}\Phi_0, \Psi_e\rangle = 0, \quad e \in E_*. \tag{7.44}$$

Thus the ψ_e generate complement spaces that are *orthogonal* relative to the energy inner product $a(u,v) = \langle \mathcal{L}u, v\rangle$ (when $\mathcal{L}$ is symmetric, recall Section 1.4). One should note that this also covers *genuinely* multivariate generators ϕ not obtained by tensor products of univariate ones. The restriction to spatial dimensions $n \le 3$ arises from the fact that in these cases the masks for the wavelets can be retrieved from the mask of the generators in an explicit way, which plays a central role in the construction.

The above result (7.44) concerns the decomposition for *one* level. Due to the appearance of the symbol σ, the adapted wavelets are (as in the vaguelette case) *scale-dependent*. To obtain a complete wavelet basis, one has to demand that (7.43) holds for all symbols $\sigma_j := \sigma(2^j \cdot)$. Let $\{\psi_{j,e}\}_{e \in E_*}$ be the wavelet family constructed above relative to σ_j. Then $\{\phi(\cdot - k) : k \in \mathbb{Z}^n\} \cup \{\psi_{j,e}(2^j \cdot - k) : k \in \mathbb{Z}^n,\ e \in E_*,\ j = 0, 1, 2, \ldots\}$ forms a wavelet basis satisfying

$$\langle \mathcal{L}\psi_{j,e}(\cdot - k), \psi_{j',e'}(\cdot - k') \rangle = 0, \quad e, e' \in E_*, \quad k, k' \in \mathbb{Z}^n, \quad j \neq j', \tag{7.45}$$

(Dahlke and Weinreich 1994, Dahlke 1996).

Returning to the periodic case, the stiffness matrices relative to this basis is therefore *block-diagonal*. In particular, the case $\mathcal{L} = -\Delta + a$, $a > 0$ is covered. In this case the wavelets can be chosen to have *compact support*. Therefore the diagonal blocks are sparse. Properly scaled, each block is *well conditioned*. Thus such matrices are easily inverted, which suggests using them for preconditioning purposes, when $\mathcal{L}$ has a more complicated form.

7.7. Evolution equations

The next step is to consider problems of the form (2.14) described in Section 2.2. A common approach to such problems is to fix a time discretization that is *implicit* in the leading second-order term $\mathcal{L}u$ and *explicit* in $\mathcal{G}(u)$. The simplest example is the Euler scheme

$$\frac{u^{(l+1)} - u^{(l)}}{\Delta t} + \mathcal{L}u^{(l+1)} + \mathcal{G}(u^{(l)}) = 0, \tag{7.46}$$

where the upper index l denotes the time level, that is, $u^{(l)}$ is an approximation to $u(\cdot, t_l)$, $t_l = t_{l-1} + \Delta t$. Thus for each time step one has to solve an elliptic problem

$$(I + \Delta t \mathcal{L})u^{(l+1)} = u^{(l)} - \Delta t \mathcal{G}(u^{(l)}), \tag{7.47}$$

of the form discussed in previous sections. Of course, any elliptic solver such as the robust FE-based wavelet preconditioners discussed in Section 6 can be used for that purpose too. Once space discretizations $\mathcal{L}_j$, $\mathcal{G}_j$ for $\mathcal{L}$ and $\mathcal{G}$ relative to the spaces $S(\Phi_j) =: S_j$, say, have been chosen, one has to solve *linear* problems

$$(I + \Delta t \mathcal{L}_j)u_j^{(l+1)} = u_j^{(l)} - \Delta t \mathcal{G}_j(u_j^{(l)}).$$

In particular, when $\mathcal{G} \equiv 0$ one formally obtains

$$u_j^{(l+1)} = (I + \Delta t \mathcal{L}_j)^{-l} u_j^{(0)}. \tag{7.48}$$

Note that the projection scheme from the previous section, with respect to orthonormal wavelet bases, gives rise to a conceptually somewhat different

scheme of the form

$$u_j^{(l+1)} = (Q_j(I + \Delta t\mathcal{L})^{-1})^l u_j^{(0)}, \tag{7.49}$$

where one can use the fact that

$$(Q_j(I + \Delta t\mathcal{L})^{-1})^l = (Q_j(I + \Delta t\mathcal{L})^{-1}Q_j)^{l-1}(Q_j(I + \Delta t\mathcal{L})^{-1}). \tag{7.50}$$

Liandrat and Tchamitchian (1997) have pointed out that there exists *no* discretization $\mathcal{L}_j$ of $\mathcal{L}$, independent of Δt, such that $(I + \Delta t\mathcal{L}_j)^{-1} = Q_j(I + \Delta t\mathcal{L})^{-1}Q_j$.

An algorithm based on (7.49) is proposed and analysed in Liandrat and Tchamitchian (1997). This *time-dependent vaguelette scheme* looks schematically as follows.

ALGORITHM 3 (TIME-DEPENDENT V-SCHEME)

(1) Choice of Ψ, $\tilde{\Psi}$, Δt.
(2) Initialization:
   - When globally supported wavelets are used, fix truncated versions of the filter matrices $\mathbf{M}_{j,0}$, $\mathbf{M}_{j,1}$ from (3.4), (3.8).
   - Approximate the filters

$$\mathbf{F}_q := \langle \Phi_q, (I + \Delta t\mathcal{L})^{-1}\tilde{\Phi}_q \rangle.$$

(3) Compute the scalar products

$$\langle u_j^{(l)} - \Delta t\mathcal{G}(u_j^{(l)}), \theta_\lambda \rangle$$

according to (7.23) (or the hybrid evaluation scheme in Section 7.4).
(4) The representation of $u_j^{(l)}$ in terms of Φ_j is obtained with the aid of (3.26).

To ensure that this scheme is competitive with finite element schemes, an efficient evaluation of the nonlinear terms $\mathcal{G}(u_j^{(l)})$ is needed. This is a nontrivial task when working in the wavelet representation. Some proposals on how to deal with this task can be found in Liandrat and Tchamitchian (1997). We will address this issue again later.

Wavelet representation of evolution operators

We will now briefly describe an alternative approach pursued by a number of researchers; see, for example, Beylkin and Keiser (1997), Dorobantu (1995), Enquist, Osher and Zhong (1994) and Perrier (1996). The order of time and space discretization is now reversed. The basic ideas will be explained again for the model case of univariate evolution equations on $[0, 1]$ with periodic boundary conditions of the form

$$\frac{\partial u}{\partial t} = \mathcal{L}u + \mathcal{G}(u), \quad u(\cdot, t_0) = u_0, \quad u(x, t) = u(x + 1, t), \tag{7.51}$$

where, as before, $\mathcal{G}$ is a possibly nonlinear operator and $\mathcal{L}$ is a constant coefficient second-order operator $\mathcal{L}u = \nu\frac{\partial^2}{\partial x^2}$, $\nu > 0$. Again the examples (2.15) and (2.16) are covered.

A key role is played by the classical *semi-group* approach. In fact, by *Duhamel's principle* the solution $u(x,t)$ to (7.51) is given by

$$u(x,t) = e^{(t-t_0)\mathcal{L}}u_0(x) + \int_{t_0}^{t} e^{(t-\tau)\mathcal{L}}\mathcal{G}(u(x,\tau))\,\mathrm{d}\tau. \tag{7.52}$$

In particular, for the heat equation where $\mathcal{G} = 0$ and $\nu = 1$, the solution operator $e^{t\mathcal{L}}$ has the form

$$(e^{t\mathcal{L}}v)(x) = \frac{1}{\sqrt{2\pi t}}\int_{\mathbb{R}} e^{\frac{(x-y)^2}{4t}}v(y)\,\mathrm{d}y. \tag{7.53}$$

The reason why the use of this representation has been mainly confined to theoretical purposes is that conventional discretizations of the involved operators are *not sparse*. The main thrust of the above mentioned papers is that this is different when employing wavelet-based discretizations (recall Section 1.5 (c)).

An example of this type is the following proposal from Enquist et al. (1994) concerning *long time solutions*. It begins with a conventional discretization by the method of lines

$$\frac{d}{dt}\mathbf{U} = \mathcal{L}_j\mathbf{U} + \mathbf{F} \tag{7.54}$$

of $\frac{\partial u}{\partial t} = \mathcal{L}u + f$, where $\mathbf{U} = (U_k(t) \approx u(k2^{-j},t))_{k=0}^{2^j-1}$, $\mathbf{F} = (F_k(t))_{k=0}^{2^j-1}$, $F_k(t) = f(k2^{-j},t)$ and

$$(\mathcal{L}_j\mathbf{U})_k = \nu 2^{2j}(U_{k-1}(t) - 2U_k(t) + U_{k+1}(t))$$

is the classical second divided difference operator. Now Duhamel's principle applied to (7.54) yields

$$\mathbf{U}(t+\Delta t) = e^{\Delta t\mathcal{L}_j}\mathbf{U}(t) + \int_{t}^{t+\Delta t} e^{(t+\Delta t-s)\mathcal{L}_j}\mathbf{F}(s)\,\mathrm{d}s. \tag{7.55}$$

Conventional numerical schemes are now obtained by expanding the evolution operator $e^{\Delta t\mathcal{L}_j}$. For instance, Taylor expansion and truncation yields explicit schemes with the usual stability constraints on the time step Δt relative to spatial mesh size $\Delta x = 2^{-j}$ in the present case. Any such approximation $\mathcal{E}$ to $e^{\Delta t\mathcal{L}_j}$ provides

$$\mathbf{U}(n\Delta t) \approx \mathbf{U}^n = \mathcal{E}^n\mathbf{U}^0 + \sum_{i=1}^{n}\mathcal{E}^{n-i}\mathbf{F} \tag{7.56}$$

as a discrete counterpart of (7.55). The simplest examples are $\mathcal{E} = (I + \Delta t \mathcal{L}_j)$ or $\mathcal{E} = (I - \Delta t \mathcal{L}_j)^{-1}$ for the *explicit* and *implicit Euler* schemes, respectively. Alternatively, $\mathcal{E} = (I - \frac{\Delta t}{2}\mathcal{L}_j)^{-1}(I + \frac{\Delta t}{2}\mathcal{L}_j)$ corresponds to the *Crank–Nicholson scheme.*

Now, suppose one is interested in long time solutions of the heat equation. This requires high powers of $\mathcal{E}$. In particular, the powers $\mathcal{E}^{2^l}$ can be obtained by repeated squaring. Setting $\mathcal{S}_m := \mathcal{E}^{2^m}$, $\mathcal{C}_m := \sum_{i=0}^{2^m-1} \mathcal{E}^i \mathbf{F}$, and noting that

$$\begin{aligned}\sum_{i=0}^{2^m-1} \mathcal{E}^i &= I + \mathcal{E} + \mathcal{E}^2(I + \mathcal{E}) + \mathcal{E}^4(I + \mathcal{E} + \mathcal{E}^2 + \mathcal{E}^3) + \cdots \\ &\quad + \mathcal{E}^{2^{m-1}}(I + \mathcal{E} + \cdots + \mathcal{E}^{2^{m-1}-1}),\end{aligned} \tag{7.57}$$

the following algorithm approximates the solution at time $t = 2^m \Delta t$ after m steps.

ALGORITHM 4
Set $\mathcal{S}_0 := \mathcal{E}$, $\mathcal{C}_0 = \mathbf{F}$.

(1) For $i = 1, 2, \ldots, m$:

$$\begin{aligned}\mathcal{S}_i &= \mathcal{S}_{i-1}^2, \\ \mathcal{C}_i &= (I + \mathcal{S}_{i-1})\mathcal{C}_{i-1}.\end{aligned}$$

(2) Then $\mathbf{U}^{(2^m)} = \mathcal{S}_m \mathbf{U}^{(0)} + \mathcal{C}_m$ is the approximate solution of (7.54) at time $2^m \Delta t$.

The conceptual advantage is that time is rapidly advanced by a few applications of powers of $\mathcal{E}$. However, in this form Algorithm 4 cannot be applied in practice since the corresponding matrices fill up after a few squarings, so that each step becomes too costly. The basic idea of Enquist et al. (1994) is to transform Algorithm 4 in such a way that the $\mathcal{S}_i$ become sparse (within some tolerance). One exploits the fact that wavelet representations of CZO (and their powers) are nearly sparse (recall Section 7.2). Similar ideas are used by Perrier (1996). Consider again the 1-periodic case and a corresponding dual pair of periodized generator bases Φ_j, $\tilde{\Phi}_j$. Let $N = 2^j$. Each $\mathbf{c} \in \mathbb{R}^N$ can then be indentified with $\mathbf{c}^T \Phi_j \in S(\Phi_j)$ and the transform $\mathbf{T}_j^{-1}$ defined by (3.28) transforms $\mathbf{c}$ into the corresponding wavelet coefficient vector $\mathbf{d} = \mathbf{T}_j^{-1}\mathbf{c}$. If Φ_j consists of pairwise orthogonal functions, for instance periodized Daubechies scaling functions, so that $\Phi_j = \tilde{\Phi}_j$, the transformation $\mathbf{T}_j$ is orthogonal and $\mathbf{T}_j^{-1} = \mathbf{T}_j^T$. Hence the application of $\mathcal{S}_i^2$ in wavelet representation becomes

$$\mathbf{T}_j^T \mathcal{S}_i^2 \mathbf{c} = \mathbf{T}_j^T \mathcal{S}_i^2 \mathbf{T}_j \mathbf{d} = (\mathbf{T}_j^T \mathcal{S}_i \mathbf{T}_j)^2 \mathbf{d}.$$

Replacing $\mathcal{S}_0$ in Algorithm 4 by $\mathbf{T}_j^T \mathcal{E} \mathbf{T}_j$ produces an equivalent scheme.

The gain lies in the fact that the iterates $\mathcal{S}_i$ now become sparse (*cf.* (7.11), (6.24)). To increase the efficiency of the above scheme, one can introduce the operation $\mathrm{trunc}(\mathcal{S}_i, \varepsilon)$, which sets all entries to zero whose absolute value stays below the threshold $\varepsilon > 0$. This leads to the following.

ALGORITHM 5
Set $\mathcal{S}_o = \mathrm{trunc}(\mathbf{T}_j^{\mathrm{T}} \mathcal{E} \mathbf{T}_j, \varepsilon)$ and $\mathcal{C}_0 = \mathbf{T}_j^{\mathrm{T}} \mathbf{F}$.

(1) For $i = 1, 2, \ldots, m$:

$$\begin{aligned} \mathcal{S}_i &= \mathrm{trunc}(\mathcal{S}_{i-1}^2, \varepsilon), \\ \mathcal{C}_i &= (I + \mathcal{S}_{i-1})\mathcal{C}_{i-1}, \end{aligned}$$

$$\mathbf{U}^{(2^m)} = \mathbf{T}_j(\mathcal{S}_m \mathbf{T}_j^{\mathrm{T}} \mathbf{U}^{(0)} + \mathcal{C}_m).$$

Of course, the threshold ε has to be chosen appropriately. Also, modifications and additional assumptions are necessary when f depends on t explicitly. The error analysis carried out by Dorobantu (1995) indicates that $\varepsilon \leq \Delta t$ is a reasonable choice. Although an explicit scheme constrains Δt relative to $\Delta x = 2^{-j}$, the experiments in Dorobantu (1995) suggest that a simple explicit Euler scheme $\mathcal{E} = (I + \Delta t \mathcal{L}_j)$ is in this context superior to an implicit scheme, although the solution of corresponding systems benefits from the preconditioning effects of the wavelet bases. One should also note that the choice of the wavelet basis is not necessarily related to the space discretization, which above was just finite differences. On one hand, this increases flexibility and reminds us of algebraic multigrid. On the other hand, the scope for rigorous analysis of the scheme certainly decreases.

The non-standard form
The key idea of the above scheme is that, as soon as sparse representations of evolution operators are available, discretizations of the integral representation (7.52) reduce to matrix-vector multiplications with (nearly) sparse matrices. Therefore, the efficiency of this operation is crucial (just as in the context of iterative solvers). So far, we have primarily exploited the (near) sparseness of the matrices $\langle \mathcal{T}\Psi, \Psi \rangle^{\mathrm{T}}$, which are often referred to as the standard form of the operator $\mathcal{T}$. In particular, in the context of periodic problems the following alternate representation has been propagated by several researches. It is called *non-standard (NS) form*; see, for example, Beylkin, Coifman and Rokhlin (1991), Beylkin and Keiser (1997) and Dorobantu (1995). While $\langle \mathcal{T}\Psi, \Psi \rangle^{\mathrm{T}}$ arises from the formal expansion (see (5.37))

$$\mathcal{T} = \Sigma_0^* \mathcal{T} \Sigma_0 = \sum_{j,l=0}^{\infty} (Q_j^* - Q_{j-1}^*) \mathcal{T} (Q_l - Q_{l-1}) \tag{7.58}$$

setting $Q_{-1} = 0$, the alternative telescoping expansion gives

$$\mathcal{T} = \sum_{j=0}^{\infty} (Q_{j+1}^* \mathcal{T} Q_{j+1} - Q_j^* \mathcal{T} Q_j) + Q_0^* \mathcal{T} Q_0,$$

where as before $Q_j f = \langle f, \tilde{\Phi}_j \rangle \Phi_j = \langle f, \tilde{\Phi}_0 \rangle \Phi_0 + \sum_{l=0}^{j-1} \langle f, \tilde{\Psi}_l \rangle \Psi_l$. One readily checks that

$$\mathcal{T} = \sum_{j=0}^{\infty} \left\{ (Q_{j+1}^* - Q_j^*) \mathcal{T} (Q_{j+1} - Q_j) + Q_j^* \mathcal{T} (Q_{j+1} - Q_j) + (Q_{j+1}^* - Q_j^*) \mathcal{T} Q_j \right\} + Q_0^* \mathcal{T} Q_0. \tag{7.59}$$

Of course, one can start the expansion at any other fixed coarsest level j_0 instead of $j_0 = 0$. Another way of looking at the NS form is to expand the kernel of $\mathcal{T}$ relative to the dilates and translates of the bivariate wavelets

$$\psi(x)\psi(y), \quad \phi(x)\psi(y), \quad \psi(x)\phi(y).$$

Since

$$\begin{aligned} (Q_{j+1}^* - Q_j^*) \mathcal{T} (Q_{j+1} - Q_j) v &= \langle v, \tilde{\Psi}_j \rangle \langle \mathcal{T} \Psi_j, \Psi_j \rangle \tilde{\Psi}_j, \\ (Q_{j+1}^* - Q_j^*) \mathcal{T} Q_j v &= \langle v, \tilde{\Phi}_j \rangle \langle \mathcal{T} \Phi_j, \Psi_j \rangle \tilde{\Psi}_j, \\ Q_j^* \mathcal{T} (Q_{j+1} - Q_j) v &= \langle v, \tilde{\Psi}_j \rangle \langle \mathcal{T} \Psi_j, \Phi_j \rangle \tilde{\Phi}_j, \end{aligned}$$

the matrix representations of the block operators are

$$\begin{aligned} \mathbf{A}_j \ ;= \ \langle \mathcal{T} \Psi_j, \Psi_j \rangle &= \mathbf{M}_{j,1}^{\mathrm{T}} \langle \mathcal{T} \Phi_{j+1}, \Phi_{j+1} \rangle \mathbf{M}_{j,1}, \\ \mathbf{B}_j \ := \ \langle \mathcal{T} \Phi_j, \Psi_j \rangle &= \mathbf{M}_{j,0}^{\mathrm{T}} \langle \mathcal{T} \Phi_{j+1}, \Phi_{j+1} \rangle \mathbf{M}_{j,1}, \\ \mathbf{C}_j \ := \ \langle \mathcal{T} \Psi_j, \Phi_j \rangle &= \mathbf{M}_{j,1}^{\mathrm{T}} \langle \mathcal{T} \Phi_{j+1}, \Phi_{j+1} \rangle \mathbf{M}_{j,0}, \end{aligned} \tag{7.60}$$

and $\mathbf{H}_0 := \langle \mathcal{T} \Phi_0, \Phi_0 \rangle$ for the coarse level contribution. Thus these blocks involve the three types of scalar products

$$\alpha_{k,l}^j = \langle \mathcal{T} \psi_{j,k}, \psi_{j,l} \rangle, \quad \beta_{k,l}^j = \langle \mathcal{T} \phi_{j,k}, \psi_{j,l} \rangle, \quad \gamma_{k,l}^j = \langle \mathcal{T} \psi_{j,k}, \phi_{j,l} \rangle, \tag{7.61}$$

which, in contrast to the standard form, involve *only* functions on the *same* level j in each block.

As a consequence, several practical advantages can be attributed to the NS form. In contrast to the standard form, the NS form maintains the convolution structure of an operator. Thus FFT can be used to enhance further the efficiency of matrix vector multiplication in NS form. Moreover, since the scalar products only involve functions on the same level, the methods described in Section 4.2 can be used to calculate them efficiently. Only finitely different coefficients are needed to represent a constant coefficient

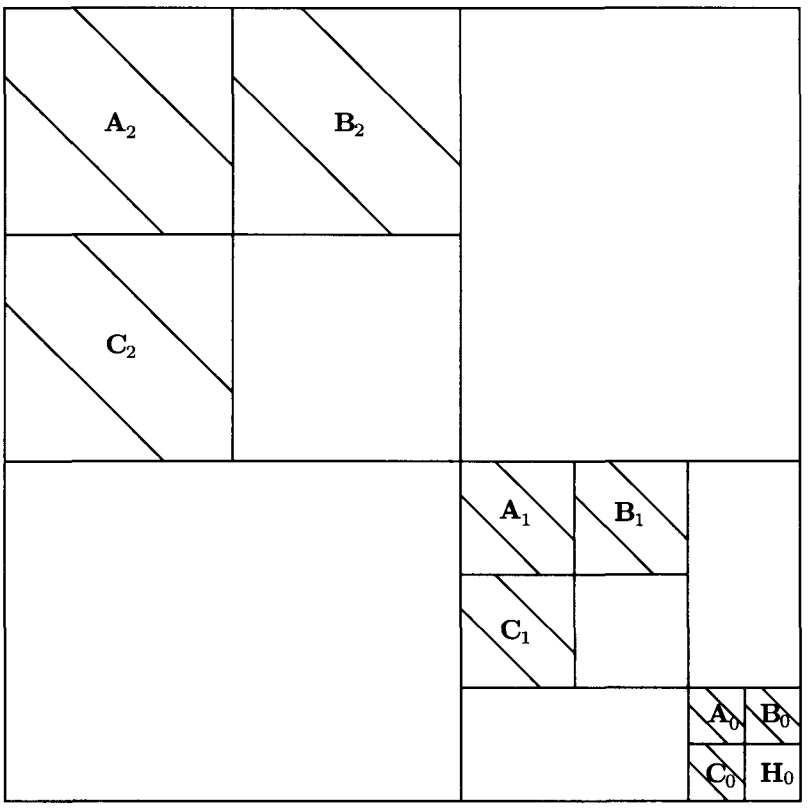

Fig. 3. Schematic view of the NS form

differential operator in NS form. For an extensive discussion of the representation of operators in wavelet bases, the reader is referred to Beylkin (1992). However, one has to stress that one consequence of uncoupling levels in the NS form is that the vectors it applies to are *not* representations of the original vector with respect to *any* basis. Instead they could be viewed as coefficient vectors relative to a redundant spanning set. Accordingly the size of the NS form is up to almost twice the size of the corresponding standard form; see, for example, Beylkin (1992) and Beylkin and Keiser (1997).

More precisely, the action of the truncated operator $\mathcal{T}_J := Q_J^* \mathcal{T} Q_J$ can be described as

$$\begin{aligned} \mathcal{T}_J v &= \mathcal{T}_0 v + \sum_{j=0}^{J-1} \left\{ \langle v, \tilde{\Psi}_j \rangle \mathbf{A}_j \tilde{\Psi}_j + \langle v, \tilde{\Phi}_j \rangle \mathbf{B}_j \tilde{\Psi}_j + \langle v, \tilde{\Psi}_j \rangle \mathbf{C}_j \tilde{\Phi}_j \right\} \\ &=: \sum_{j=0}^{J-1} \{ \tilde{\mathbf{d}}_j^{\mathrm{T}} \tilde{\Psi}_j + \tilde{\mathbf{c}}_j^{\mathrm{T}} \tilde{\Phi}_j \}, \end{aligned} \tag{7.62}$$

where

$$\begin{aligned}\tilde{\mathbf{d}}_j^{\mathrm{T}} &:= \langle v, \tilde{\Psi}_j\rangle \mathbf{A}_j + \langle v, \tilde{\Phi}_j\rangle \mathbf{B}_j, \quad j = 0, \dots, J-1,\\ \tilde{\mathbf{c}}_j^{\mathrm{T}} &:= \langle v, \tilde{\Psi}_j\rangle \mathbf{C}_j, \quad j = 1, \dots, J-1,\\ \tilde{\mathbf{c}}_0^{\mathrm{T}} &:= \langle v, \tilde{\Phi}_0\rangle \langle \mathcal{T}\Phi_0, \Phi_0\rangle + \langle v, \tilde{\Psi}_0\rangle \mathbf{B}_0.\end{aligned} \tag{7.63}$$

Thus the application of $\mathcal{T}_J$ in the NS form requires, in addition to the wavelet coefficients $\langle v, \tilde{\Psi}_j\rangle$ of v in $S(\Psi_j)$, the scaling function coefficients $\langle v, \tilde{\Phi}_j\rangle$ on level j. Hence, when $v \in S(\Phi_J)$ is given in single-scale representation $v = \langle v, \tilde{\Phi}_J\rangle \Phi_J$, the array $\tilde{\mathbf{d}}_J^{\mathrm{NS}} := (\tilde{\mathbf{c}}_0, \tilde{\mathbf{d}}_0, \dots, \tilde{\mathbf{c}}_{J-1}, \tilde{\mathbf{d}}_{J-1})$ representing the application of $\mathcal{T}_j$ to v is obtained, according to (7.63), by applying the blocks $\mathbf{A}_j, \mathbf{B}_j, \mathbf{C}_j$ to the result of the pyramid scheme (3.28) for the corresponding level.

Conversely, to transform the output $\tilde{\mathbf{d}}_J^{\mathrm{NS}}$ back into a coefficient vector relative to a basis of $S(\tilde{\Phi}_J)$, one can proceed as follows. Since by (3.4), (3.8),

$$\tilde{\mathbf{c}}_j^{\mathrm{T}} \tilde{\Phi}_j = \tilde{\mathbf{c}}_j^{\mathrm{T}} \tilde{\mathbf{M}}_{j,0}^{\mathrm{T}} \tilde{\Phi}_{j+1}, \quad \tilde{\mathbf{d}}_j^{\mathrm{T}} \tilde{\Psi}_j = \tilde{\mathbf{d}}_j^{\mathrm{T}} \tilde{\mathbf{M}}_{j,1}^{\mathrm{T}} \tilde{\Phi}_{j+1},$$

the pyramid scheme

$$\begin{array}{ccccccccccc} \tilde{\mathbf{c}}_0 & \to & \hat{\mathbf{c}}_1 & \to & \hat{\mathbf{c}}_2 & \to & \cdots & \to & \hat{\mathbf{c}}_{J-1} & \to & \hat{\mathbf{c}}_J \\ \tilde{\mathbf{d}}_0, \tilde{\mathbf{c}}_1 & \nearrow & \tilde{\mathbf{d}}_1, \tilde{\mathbf{c}}_2 & \nearrow & \tilde{\mathbf{d}}_2, \tilde{\mathbf{c}}_3 & \nearrow & \cdots & \nearrow & \tilde{\mathbf{d}}_{J-1}, \tilde{\mathbf{c}}_J & \nearrow & \end{array}, \tag{7.64}$$

where

$$\hat{\mathbf{c}}_0 := \tilde{\mathbf{c}}_0, \quad \hat{\mathbf{c}}_j := \tilde{\mathbf{M}}_{j-1,0} \hat{\mathbf{c}}_{j-1} + \tilde{\mathbf{M}}_{j-1,1} \tilde{\mathbf{d}}_{j-1} + \tilde{\mathbf{c}}_j, \quad j = 1, \dots, J,$$

similarly to (3.26) produces, in view of (7.62), the single-scale representation $\mathcal{T}_J v = \hat{\mathbf{c}}_J^{\mathrm{T}} \tilde{\Phi}_J$. Likewise, in view of

$$\tilde{\mathbf{c}}_j^{\mathrm{T}} \tilde{\Phi}_j = \tilde{\mathbf{c}}_j^{\mathrm{T}} (\overline{\mathbf{M}}_{j-1,0} \tilde{\Phi}_{j-1} + \overline{\mathbf{M}}_{j-1,1} \tilde{\Psi}_{j-1}),$$

(recall (3.13))

$$\begin{array}{ccccccccc} \tilde{\mathbf{c}}_{J-1}, \tilde{\mathbf{d}}_{J-2} & \to & \hat{\mathbf{c}}_{J-2}, \tilde{\mathbf{d}}_{J-3} & \to & \hat{\mathbf{c}}_{J-3}, \tilde{\mathbf{d}}_{J-4} & \to & \cdots & \to & \hat{\mathbf{c}}_0 \\ \tilde{\mathbf{d}}_{J-1} & \searrow & \hat{\mathbf{d}}_{J-2} & \searrow & \hat{\mathbf{d}}_{J-3} & \searrow & & \searrow & \hat{\mathbf{d}}_0 \end{array}, \tag{7.65}$$

where

$$\hat{\mathbf{c}}_j := \tilde{\mathbf{c}}_j + \mathbf{M}_{j,0}^* \tilde{\mathbf{c}}_{j+1}, \quad \hat{\mathbf{d}}_j := \tilde{\mathbf{d}}_j + \mathbf{M}_{j,1}^* \tilde{\mathbf{c}}_{j+1}, \quad j = J-2, \dots, 0,$$

generates the wavelet representation

$$\mathcal{T}_J v = \hat{\mathbf{c}}_0^{\mathrm{T}} \tilde{\Phi}_0 + \hat{\mathbf{d}}_0^{\mathrm{T}} \tilde{\Psi}_0 + \ldots + \hat{\mathbf{d}}_{J-2}^{\mathrm{T}} \tilde{\Psi}_{J-2} + \tilde{\mathbf{d}}_{J-1}^{\mathrm{T}} \tilde{\Psi}_{J-1}. \tag{7.66}$$

Computation of the blocks $\mathbf{A}_j, \mathbf{B}_j, \mathbf{C}_j$

When $\mathcal{T}$ is a convolution operator one only has to determine the filter coefficients $\alpha_l^j, \beta_l^j, \gamma_l^j$. Moreover, in view of (7.60), it suffices to determine the

coefficients of $\langle \mathcal{T}\Phi_j, \Phi_j\rangle$ and then apply portions of the pyramid schemes (3.26). When $\mathcal{T}$ is a homogeneous operator of order p such as $\left(\frac{d}{dx}\right)^p$ this is not even necessary, and one puts

$$\langle \mathcal{T}\Phi_j, \Phi_j\rangle = 2^{-(J-j)p}\langle \mathcal{T}\Phi_J, \Phi_J\rangle. \tag{7.67}$$

Specifically, when $\mathcal{T} = \left(\frac{d}{dx}\right)^p$, the coefficients in $\langle \mathcal{T}\Phi_J, \Phi_J\rangle$ form finite difference approximations of $\mathcal{T}$ in $S(\tilde{\Phi}_J)$ of order $d + \tilde{d} - 1$, when $d, \tilde{d}$ are the respective orders of exactness of the $\Phi_j, \tilde{\Phi}_j$ and $\tilde{d} \geq d$. In fact, let $\square_{j,k} := 2^{-j}(k + [0,1])$ and $v \in C^\infty(\mathbb{R})$. Hence there exists a polynomial P of degree $d-1$ and a smooth remainder R so that

$$v|_{\square_{j,k}} = (P + 2^{-jd}R)|_{\square_{j,k}}.$$

Thus

$$\begin{aligned} Q_j^*\mathcal{T}Q_j v - \mathcal{T}v &= Q_j^*\mathcal{T}P + 2^{-jd}Q_j^*\mathcal{T}R - \mathcal{T}P - 2^{-jd}\mathcal{T}R \\ &= (Q_j^* - I)\mathcal{T}P + 2^{-jd}(Q_j^* - I)\mathcal{T}R. \end{aligned}$$

Since for $\mathcal{T} = \left(\frac{d}{dx}\right)^p$, $\mathcal{T}P$ is a polynomial of degree at most $d-1-p < \tilde{d}$, the first summand on the right-hand side vanishes. Moreover, locally $(Q_j^* - I)\mathcal{T}R$ behaves for smooth R in a neighbourhood of $\square_{j,k}$ like $2^{-\tilde{d}j}$ in the L_∞-norm, say (see Proposition 5.1).

Moment conditions of the $\mathbf{B}$*-blocks*

By (7.62), the blocks $\mathbf{A}_j$ and $\mathbf{C}_j$ are multiplied by vectors that contain wavelet coefficients. Since possibly only a few of these coefficients exceed a given threshold, one expects that these multiplications can be carried out efficiently within a desired accuracy. Instead, the vectors multiplying the blocks $\mathbf{B}_j$ consist of scaling function coefficients representing averages. Therefore these arrays are generally dense. However, it is important to note that the matrices $\mathbf{B}_j = \langle \mathcal{T}\Phi_j, \Psi_j\rangle$ have *vanishing moments* when

$$\mathcal{T} = \mathcal{H} \quad \text{or } \mathcal{T} = f\left(\frac{\partial}{\partial x}\right), \tag{7.68}$$

where $\mathcal{H}$ is the Hilbert transform (see Section 1.3) and f is analytic. More precisely, for any $\mathbf{p} := (P(l))_{l\in\mathbb{Z}}$, P a polynomial of degree $\leq d-1$, one has

$$\mathbf{p}^{\mathrm{T}}\mathbf{B}_j = 0 \tag{7.69}$$

for any $\mathcal{T}$ from (7.68); see Beylkin and Keiser (1997). In fact, by (7.60), one has

$$\mathbf{p}^{\mathrm{T}}\mathbf{B}_j = \langle \mathcal{T}\mathbf{p}^{\mathrm{T}}\Phi_j, \Psi_j\rangle. \tag{7.70}$$

By (4.30), $\mathbf{p}^{\mathrm{T}}\Phi_j$ is a polynomial of degree $d-1$. Expanding $f(\frac{d}{dx})$ in powers

of $\frac{d}{dx}$, it is clear that $f(\frac{d}{dx})\mathbf{p}^{\mathrm{T}}\Phi_j$ is a polynomial of degree $\leq d-1$. Since, by assumption $\tilde{d} \geq d$, (7.69) follows in this case from (4.33). When $\mathcal{T} = \mathcal{H}$ an argument similar to the one used in Section 1.3 also confirms (7.69); see Beylkin and Keiser (1997).

Some theoretical remarks

Due to the appearance of at least one wavelet in the scalar products (7.61) there is still a compression effect. In fact, if the kernel K of $\mathcal{T}$ satisfies

$$\left|\frac{\partial^r}{\partial x^r}K(x,y)\right| + \left|\frac{\partial^r}{\partial y^r}K(x,y)\right| \lesssim |x-y|^{-(r+1)} \tag{7.71}$$

for $x \neq y \bmod 1$, one can show that (Tchamitchian 1996)

$$\left|\alpha_{k,l}^j\right| + \left|\beta_{k,l}^j\right| + \left|\gamma_{k,l}^j\right| \lesssim |k-l|^{r+1}, \tag{7.72}$$

provided that the corresponding functions with indices k and l have disjoint supports. For the remaining cases the additional assumption

$$\left|\alpha_{k,l}^j\right| + \left|\beta_{k,l}^j\right| + \left|\gamma_{k,l}^j\right| \lesssim 1, \quad k,l,j \in \mathbb{Z}, \tag{7.73}$$

is needed, which is called the *weak boundedness property*. This condition is weaker than L_2-boundedness of the operator. It plays an important role in the following celebrated theorem due to G. David and J. L. Journé; see, for instance, Tchamitchian (1996).

Theorem 7.7 Suppose that the kernel of $\mathcal{T}$ satisfies (7.9), (7.10). Then $\mathcal{T}$ is continuous on L_2 if and only if it has the weak boundedness property (7.73), $\mathcal{T}(1) \in \mathrm{BMO}$ and $\mathcal{T}^*(1) \in \mathrm{BMO}$.

7.8. A pseudo-wavelet approach

The previous section contains major ingredients of an approach to solving periodic nonlinear equations of the form (7.51) proposed by Beylkin and Keiser (1997). There it is termed the *pseudo-wavelet approach*. It is a systematic attempt to compute an approximate solution to (7.51) at the expense of a number of arithmetic operations proportional to the number of wavelet coefficients required for representing the approximate solution to the desired accuracy. The central idea is to employ appropriate discretizations of (7.52) which ultimately reduce to the *adaptive* application of certain operators in the NS form to corresponding coefficient vectors.

One basic tool is a class of time discretization schemes presented by Beylkin, Keiser and Vozovoi (1996). For instance, in the case of Burgers' equation (2.16) the term

$$I(t,t_0) := \int_{t_0}^{t} e^{(t-\tau)\mathcal{L}} u(\cdot,\tau)\frac{\partial}{\partial x}u(\cdot,\tau)\,\mathrm{d}\tau \tag{7.74}$$

is approximated by

$$I(t+\Delta t,t) \tag{7.75}$$
$$= \frac{1}{2}\mathcal{O}_{\mathcal{L},1}\left(u(\cdot,t_0)\frac{\partial}{\partial x}u(\cdot,t+\Delta t)+u(\cdot,t+\Delta t)\frac{\partial}{\partial x}u(\cdot,t)\right)+\mathcal{O}((\Delta t)^2),$$

where

$$\mathcal{O}_{\mathcal{L},m} := \left(e^{m\Delta t\mathcal{L}}-\mathcal{I}\right)\mathcal{L}^{-1}. \tag{7.76}$$

So the idea is to discretize $\mathcal{G}(u(\cdot,\tau))$ in the time variable τ so that the *exact* application of $\int_t^{t+\Delta t} e^{(t+\Delta t-\tau)\mathcal{L}}\,\mathrm{d}\tau$ reduces to the application of $\mathcal{O}_{\mathcal{L},1}$. This is essentially different from the procedure in Section 7.7 where the space discretization was fixed before. For the derivation of higher-order schemes and a corresponding stability analysis see Beylkin et al. (1996). Here it is important that the operators $\mathcal{L}^{-1}$ or $e^{m\Delta t\mathcal{L}}$ can be evaluated *exactly* within any chosen accuracy. Again in the case of Burgers' equation one has to evaluate

$$u(x,t+\Delta t) = e^{\Delta t\mathcal{L}}u(x,t) - I(t+\Delta t,t), \tag{7.77}$$

where $\mathcal{L}=\nu\left(\frac{\partial}{\partial x}\right)^2$ and $\mathcal{O}_{\mathcal{L},1}$ is given by (7.76).

In order to apply the operator functions $e^{\Delta t\mathcal{L}}$ and $\mathcal{O}_{\mathcal{L},1}$ efficiently one is interested in computing their NS form. Therefore it is important to determine the NS form of $f\left(\frac{\partial}{\partial x}\right)$ when f is analytic. Beylkin and Keiser (1997) propose two approaches, namely to compute

$$Q_J^* f\left(\frac{\partial}{\partial x}\right) Q_J, \tag{7.78}$$

or

$$f\left(Q_J^*\frac{\partial}{\partial x}Q_J\right). \tag{7.79}$$

To compute the NS form of $f\left(\frac{\partial}{\partial x}\right)$ via (7.79) one can diagonalize $\frac{\partial}{\partial x}$ with the aid of the discrete Fourier transform and apply the spectral theorem (Beylkin and Keiser 1997).

Using (7.78), according to the discussion in the previous section (see (7.60)), one can first determine the arrays $\mathbf{c}^j := \langle f\left(\frac{\partial}{\partial x}\right)\Phi_j,\Phi_j\rangle$, consisting of the coefficients

$$c^j_{k,k'} = 2^j\int_{\mathbb{R}}\phi(2^jx-k)f\left(\frac{\partial}{\partial x}\right)\phi(2^jx-k')\,\mathrm{d}x = \mathbf{C}^j_{k-k'}.$$

Using Fourier transforms, one can show that (Beylkin and Keiser 1997)

$$c_l^j = \int_0^{2\pi} g_j(\xi) e^{i\xi l} \, \mathrm{d}\xi, \tag{7.80}$$

where

$$g_j(\xi) = \sum_{k \in \mathbb{Z}} f(-i2^j(\xi + 2\pi k)) |\hat{\phi}(\xi + 2\pi k)|^2,$$

and exploit the fact that $|\hat{\phi}(\xi)|^2$ acts like a *cut-off function*. Thus $g_j(\xi)$ can be approximated arbitrarily well by a finite sum $\tilde{g}_j(\xi)$ which can be used to discretize (7.80). Recall the similar reasoning in the vaguelette approach.

Adaptive application of operators in NS form
According to (7.63), the application of an operator in NS form requires evaluating

$$\tilde{\mathbf{d}}^j = \langle v, \tilde{\Psi}_j \rangle \mathbf{A}_j + \langle v, \tilde{\Phi}_j \rangle \mathbf{B}_j, \quad \tilde{\mathbf{c}}_j = \langle v, \tilde{\Psi}_j \rangle \mathbf{C}_j.$$

To accomplish the goal of realizing an overall solution complexity, which is proportional to the number of significant wavelet coefficients of the solution relative to a given accuracy, each calculation of $\mathcal{T}_J v$ has to be realized within this order of complexity. A heuristic reasoning towards this goal can be summarized as follows. The solution to the differential equations under consideration are typically smooth except at isolated locations where singularities such as shocks can build up. Consequently, many wavelet coefficients of the solution can be expected to stay below a given threshold. Hence the arrays $\langle v, \tilde{\Psi}_j \rangle$ are typically *short*. However, the arrays $\langle v, \tilde{\Phi}_j \rangle$ consist of averages and may be dense in spite of the smoothness of v. At this point the vanishing moment property of the **B**-blocks established in the previous section is crucial. Exploiting this property, Beylkin and Keiser (1997) argue that, when a smooth vector is applied to $\mathbf{B}_j$, the result will be sparse. In fact, Beylkin and Keiser (1997) indicate how to use the wavelet coefficients of v to replace the dense array $\langle v, \tilde{\Phi}_j \rangle$ by a sparse vector $\mathbf{s}^j$ so as to realize an efficient application of the $\mathbf{B}_j$ block within a desired tolerance of accuracy. For a more detailed discussion of the components of such a scheme we refer to Beylkin and Keiser (1997) and the literature cited there.

In addition, some interesting numerical experiments are discussed by Beylkin and Keiser (1997). First a classical Crank–Nicholson scheme for the heat equation $\frac{\partial u}{\partial t} = \nu \frac{\partial^2}{\partial x^2} u$ is compared to the wavelet-based scheme, which consists in this case of a repeated application of the NS form of $e^{\Delta t \mathcal{L}}$ via

$$u(\cdot, t_{j+1}) = e^{\Delta t \mathcal{L}} u(\cdot, t_j), \quad u(\cdot, t_0) = u_0.$$

This is an explicit procedure, yet unconditionally stable, once the evalu-

ation of $e^{\Delta t \mathcal{L}}$ is *accurate enough* to cope with the higher-order oscillations introduced by the nonlinear terms. In particular, the advantage of *higher-order schemes* is stressed. In fact, due to the higher number of vanishing moments, they result in better sparseness of the NS form of the operators. Subsequently, the scheme is tested on several versions of Burgers' equation and its generalizations. These experiments apparently confirm that the number of operations needed to update the solution in each time step remains proportional to the number of significant wavelet coefficients.

7.9. Wavelet packets and best bases

There is yet another technique for discretizing evolution equations of the form (2.14), which has been proposed by Joly, Maday and Perrier (1997), for instance. It aims at realizing best possible *compression* of the approximate solution by employing the concept of *wavelet packets* and *best bases* developed by Coifman, Meyer, Quake and Wickerhauser (1993) and Coifman, Meyer and Wickerhauser (1992). This technique is also used by Farge, Goirand, Meyer, Pascal and Wickerhauser (1992). Therefore we will briefly indicate some of the ideas in Joly et al. (1997), where further details and relevant references can be found.

To describe the concept of wavelet packets, we confine the discussion to scaling functions $\phi \in L_2(\mathbb{R})$, whose translates $\phi(\cdot - k)$, $k \in \mathbb{Z}$, are *orthonormal*, that is, $\phi = \tilde{\phi}$. Let us denote by $\mathbf{a}, \mathbf{b}$ the masks of ϕ and the wavelet ψ, that is, $b_k = (-1)^k a_{1-k}$, $k \in \mathbb{Z}$ (see (4.20)). One can use these masks to recursively generate further basis functions, defined with $\psi_0 = \phi$, $\psi_1 = \psi$, for $n \geq 1$ by

$$\psi_{2n}(x) = \sum_{k \in \mathbb{Z}} a_k \psi_n(2x - k), \quad \psi_{2n+1}(x) = \sum_{k \in \mathbb{Z}} b_k \psi_n(2x - k). \tag{7.81}$$

One can then show that

$$S(\Phi_j) = S\Big(\{\psi_n(\cdot - k) : 0 \leq n < 2^j, k \in \mathbb{Z}\}\Big),$$

so that a variety of orthonormal bases are available. Let $\mathcal{E}_J$, respectively $\mathcal{E}$, denote any subset of $\mathbb{N} \times (\mathbb{Z} \cap (-\infty, J])$, respectively $\mathbb{N} \times \mathbb{Z}$. With each $(n, j) \in \mathbb{N} \times \mathbb{Z}$ associate the interval $I_{n,j} := [2^j n, 2^j(n+1))$. Let

$$\psi_{n,j,k}(x) := 2^{j/2} \psi_n(2^j x - k).$$

Theorem 7.8 (Coifman et al. 1992) Any collection $\{\psi_{n,j,k} : (n, j, k) \in \mathcal{E}_J \times \mathbb{Z} \text{ resp. } \mathcal{E} \times \mathbb{Z}\}$ is an orthonormal basis of $S(\Phi_J)$, respectively $L_2(\mathbb{R})$, if and only if

(i) $\bigcup_{(n,j) \in \mathcal{E}_J} I_{n,j} = [0, 2^J)$, resp. $\overline{\bigcup_{(n,j) \in \mathcal{E}_J} I_{n,j}} = \mathbb{R}^+$

(ii) for all $(n,j),(n',j') \in \mathcal{E}_J$, resp. $\mathcal{E}$, one has $I_{n,j} \cap I_{n',j'} = \emptyset$ if $(n,j) \neq (n',j')$.

It is again convenient to abbreviate $\lambda = (n,j,k) \in \mathcal{E} \times \mathbb{Z}$. For a given problem, the point is now to select that basis which is *best* in a certain sense, rather in the spirit of signal analysis. The decision is based on an appropriate notion of *entropy*. For a given $v \in L_2(\mathbb{R})$, let

$$v = \sum_{\lambda \in \mathcal{E} \times \mathbb{Z}} c_\lambda \psi_\lambda.$$

For $\mathcal{D} \subset \mathcal{E}$ and any $\varepsilon > 0$, the quantity

$$H^{\varepsilon,\mathcal{D}}(v) := \#\{c_\lambda : |c_\lambda| \geq \varepsilon, \lambda \in \mathcal{D} \times \mathbb{Z}\} \tag{7.82}$$

is called the *cardinal entropy* of v. Note the difference from the following more familiar measure for the content of information,

$$H^{\mathcal{D}}(v) := -\sum_{\lambda \in \mathcal{D} \times \mathbb{Z}} |c_\lambda|^2 \ln |c_\lambda|, \tag{7.83}$$

which is called *Shannon entropy*. Minimizing the cardinal entropy over $\mathcal{D}$ corresponds to selecting a basis with respect to which the representation of v has possibly few coefficients above the threshold ε. The adaptation of these notions to the periodic case is again standard. Corresponding decomposition and reconstruction algorithms as well as the recursive determination of best bases are described in Joly et al. (1997). A comparison between the different notions of entropy favours the cardinal entropy for present purposes. Let us denote by $B_\varepsilon(v) \subset \mathcal{E}_J$ the index set for a best basis for v relative to (7.82) and set $\Lambda_\varepsilon(v) := \{\lambda = (n,j,k) : (n,j) \in B_\varepsilon(v), k \in \mathbb{Z}/2^j\mathbb{Z}\}$.

To make use of these concepts here, one associates with each λ the *centre* x_λ of the basis function ψ_λ. Moreover, one assigns to $\lambda = (n,j,k)$ an *influence rectangle* centred at $(x_\lambda, n2^j)$ in the position–frequency diagram, which symbolizes the time–frequency support of ψ_λ. Once the position of ψ_λ is determined, one can define the *neighbours* of each ψ_λ; see Joly et al. (1997) for the precise definition. Given the best basis, the reduced representation of v is given by

$$Q_\varepsilon v := \sum_{\lambda \in \tilde{\Lambda}_\varepsilon(v)} \langle v, \psi_\lambda \rangle \psi_\lambda, \tag{7.84}$$

where

$$\tilde{\Lambda}_\varepsilon(v) = \{\lambda \in \Lambda_\varepsilon(v) : |\langle v, \psi_\lambda \rangle| \geq \varepsilon\}. \tag{7.85}$$

The central step of the adaptive procedure proposed by Joly et al. (1997) is to add to $\tilde{\Lambda}_\varepsilon(v)$ the neighbours of its indices, which typically results in a set that is not much larger than $\tilde{\Lambda}_\varepsilon(v)$. In fact, for most elements in $\tilde{\Lambda}_\varepsilon(v)$, one expects that its neighbours already belong to $\tilde{\Lambda}_\varepsilon(v)$.

Of course, for non-stationary problems the solution will change in time. Small variations of a function may actually cause significant changes of the best basis. However, Joly et al. (1997) observe that the entropy does not change much when dealing with evolution equations like Burgers' equation. Thus the best basis does not have to be updated after each time step.

We now outline the algorithm from Joly et al. (1997) for Burgers' equation (2.16). We wish to approximate the solution $u(m\Delta t, \cdot)$ at time $m\Delta t$ by $u^m = u_J^m \in S(\Phi_J)$ as follows.

ALGORITHM 6 (JMP)
Let u_J^0 be the approximation of the initial value u^0 and let $B_\varepsilon(u_J^0)$ denote its best basis. Let

$$\Lambda^0 := \tilde{\Lambda}_\varepsilon(u_J^0)$$

be the reduced index set after thresholding.

$(m+1)$st step: Given

$$u^m = \sum_{\lambda \in \Lambda^m} c_\lambda(m)\psi_\lambda,$$

where $\Lambda^m \subseteq \Lambda(u_J^0)$.

(i) Form $\tilde{\Lambda}^m$ as in (7.85) and $\tilde{u}^m$ as in (7.84) (relative to $\tilde{\Lambda}^m$).
(ii) Form Λ^{m+1} by adding to $\tilde{\Lambda}^m$ those elements in $\Lambda_\varepsilon(u_J^0)$ that are neighbours of elements from $\tilde{\Lambda}^m$.
(iii) Determine

$$u^{m+1} = \sum_{\lambda \in \Lambda^{m+1}} c_\lambda(m+1)\psi_\lambda$$

by requiring that the following Galerkin conditions hold:

$$\left\langle \frac{1}{\Delta t}(u^{m+1} - \tilde{u}^m) + \frac{1}{2}\frac{\partial}{\partial x}(\tilde{u}_*^m)^2, \psi_\lambda \right\rangle = -\nu \left\langle \frac{\partial}{\partial x}\tilde{u}_*^m, \frac{\partial}{\partial x}\psi_\lambda \right\rangle, \tag{7.86}$$

for $\lambda \in \Lambda^{m+1}$, where $\tilde{u}_*^m := \frac{3}{2}\tilde{u}^m - \frac{1}{2}\tilde{u}^{m-1}$.

We conclude with a brief discussion of the implementation.

Since the sets Λ^m change, the stiffness matrices needed in (7.86) change as well. Therefore Joly et al. (1997) propose to precompute the *whole* stiffness matrix relative to the entire (periodized) basis Φ_J. To generate the possible wavelet packets, one then has to use the corresponding multiscale transformations, providing a matrix of size $2^J \times J2^J$.

Using orthonormal spline wavelets, the masks are no longer short, so that one has to resort to FFT in the multiscale transformations. This introduces additional log factors in the operation count. The decomposition of the columns in the extended stiffness matrix according to the chosen best basis

requires the order of $J2^{2J}$ operations. It seems that this strategy at some stage requires computational work which is quadratic in the dimension of the uncompressed problem size corresponding to $S(\Phi_J)$. This may be a serious drawback when dealing with several spatial variables.

The evaluation of nonlinear terms requires special care, which will be discussed in the following section.

7.10. Evaluation of nonlinear terms

A critical role in all the above developments is played by the evaluation of nonlinear terms. It is perhaps worthwhile to comment briefly on the principal problems arising in this context. As before, let $\Psi_\Lambda = \{\psi_\lambda : \lambda \in \Lambda\}$. So far the discussion has stressed important advantages of multiscale representations of the form $u_\Lambda = \mathbf{d}_\Lambda^T \Psi_\Lambda$, where Λ selects only those wavelets that are needed to represent a function u to some given tolerance. However, note that at any point in the domain, wavelets from all levels appearing in Λ may contribute. Thus the cost of evaluating a function in multiscale representation at a single point could be proportional to the highest level J appearing in Λ. When frequent evaluations are necessary, this could of course significantly diminish efficiency. In contrast, evaluating a function in single-scale representation requires only a finite number of operations *independent* of the level J. On the other hand, $\#\Lambda$ could be very small compared to $\dim S(\Phi_J)$. Thus the transformation of u_Λ into single-scale representation in $S(\Phi_J)$ would produce a much larger array of coefficients which, due to their nature of representing averages, may all be significant. This would waste the significant reduction of complexity gained by the sparse representation of u_Λ in wavelet coordinates.

This problem is exacerbated when, instead of point evaluations, one has to compute *nonlinear* functionals of a function u_Λ given in multiscale representation. A typical example arises in connection with the elliptic problem (7.47). Suppose that the approximate solution $u^{(l)}$ from the previous time step l is given as $u^{(l)} = \mathbf{d}_\Lambda^T \Psi_\Lambda$, where Λ is a possibly small lacunary subset of ∇. If one uses a collocation scheme for solving (7.47), one has to evaluate the nonlinear term $\mathcal{G}(u^{(l)})$ on some grid. This requires the evaluation of $u^{(l)}$ on that grid, which is the task discussed above, followed by the application of $\mathcal{G}$. If the application of $\mathcal{G}$ is expensive, an alternative is to *approximate* $\mathcal{G}(u^{(l)})$ first and then evaluate this approximation. When J is the highest scale in Λ, depending on the nature of $\mathcal{G}$, one expects that $\mathcal{G}(\mathbf{d}_\Lambda^T \Psi_\Lambda)$ can be accurately resolved on a level $\hat{J} > J$. But, again, if the approximation were given in a single-scale form, its evaluation would be inexpensive, but the representation itself would possibly involve far more coefficients than those in the array $\mathbf{d}_\Lambda$. This suggests also seeking some lacunary multiscale representation $\hat{\mathbf{d}}_{\hat{\Lambda}}^T \Xi_{\hat{\Lambda}} \approx \mathcal{G}(\mathbf{d}_\Lambda^T \Psi_\Lambda)$ with respect to a suitable basis Ξ (not

necessarily equal to Ψ). In fact, when (7.47) is to be solved by a Galerkin scheme, one would have to approximate the quantities $\langle \mathcal{G}(\mathbf{d}_\Lambda^T \Psi_\Lambda), \psi_\lambda \rangle$. Thus, for $\Xi = \tilde{\Psi}$, the array $\hat{\mathbf{d}}_{\hat{\Lambda}}^T$ would readily provide these quantities.

Overall, since in many cases one expects that the somewhat higher cost of evaluating a function in multiscale representation is by far offset by the sparseness of the representation, the central objective can be summarized as follows. If, for $\epsilon > 0$, the set $\Lambda \subset \nabla$ is needed to approximate the solution by $u_\Lambda = \mathbf{d}_\Lambda^T \Psi_\Lambda$ within a tolerance ϵ, find a possibly small set $\hat{\Lambda}$ and an approximation $\hat{\mathbf{d}}_{\hat{\Lambda}}^T \Xi_{\hat{\Lambda}}$ to $\mathcal{G}(\mathbf{d}_\Lambda^T \Psi_\Lambda)$ that is sufficiently accurate to preserve the overall precision of the solution scheme. Moreover, when ϵ tends to zero, so that the cardinality of $\Lambda = \Lambda_\epsilon$ increases, the *ideal* situation would be that the corresponding size of $\hat{\Lambda} = \hat{\Lambda}_\epsilon$ stays *proportional* to $\#\Lambda_\epsilon$ *uniformly* in ϵ. Likewise, the computational work needed to determine the approximation $\hat{\mathbf{d}}_{\hat{\Lambda}_\epsilon}^T \Xi_{\hat{\Lambda}_\epsilon}$ should also be of the order of $\#\Lambda_\epsilon$ (perhaps times a logarithmic term).

It seems that we are at present far from this goal, at least in the above strict asymptotic sense. Since this is currently a subject of intense research, the state of the art will probably change quickly in the near future. Giving a detailed account of the various existing approaches would certainly go beyond the scope of this paper. Nevertheless, sketching some ideas, at least, should be worthwhile.

A typical nonlinear term arising, for instance, in (2.16) and (2.17) is $u \frac{\partial}{\partial x} u$. Since products of functions can be obtained as differences of squares, it suffices to consider $f(u) = u^2$. The approach pursued by Beylkin and Keiser (1997) starts with the expansion

$$\begin{aligned} (Q_J v)^2 - (Q_0 v)^2 &= \sum_{j=0}^{J-1} \left((Q_{j+1} v)^2 - (Q_j v)^2 \right) \\ &= \sum_{j=0}^{J-1} \left(2(Q_j v)(R_j v) + (R_j v)^2 \right), \end{aligned} \tag{7.87}$$

where we abbreviate $R_j := Q_{j+1} - Q_j$.

In fact, within a given tolerance, one has $v \approx Q_J v$ for J sufficiently large. This gives

$$v^2 \approx (Q_0 v)^2 + \sum_{j=0}^{J-1} \left(2(Q_j v)(R_j v) + (R_j v)^2 \right). \tag{7.88}$$

The evaluation of $(Q_0 v)^2$ is inexpensive. The problem is that products in the summation will generally not belong to the same space as the factors. Since products correspond to convolutions in the Fourier domain, one can estimate the extent to which higher oscillations are introduced. To resolve

them accurately enough one needs a higher level of resolution. However, estimating the spread in the Fourier domain, one can argue that, again within some tolerance, $(Q_j v)(R_j v)$ and $(R_j v)^2$ belong to $S(\Phi_{j+j_0})$ for some positive j_0. For any given tolerance a positive j_0 does indeed exist independent of j (Beylkin and Keiser 1997). Thus, by repeated application of the refinement matrices, one can determine the representation of $Q_j v$ and $R_j v$ in $S(\Phi_{j+j_0})$, for instance,

$$R_j v = (\mathbf{d}^{j+j_0})^{\mathrm{T}} \Phi_{j+j_0}.$$

If, in addition, the functions in Φ_j were *interpolatory*, that is, $\phi(\cdot - k) = \delta_{0,k}$, the assumption $v, f(v) \in S(\Phi_j)$ would give

$$f(v) = \sum_k f(c_k)\phi(\cdot - k), \quad v = \sum_k c_k \phi(\cdot - k). \tag{7.89}$$

In this case one would have $(R_j v) = \left((\mathbf{d}^{j+j_0})^2\right)^{\mathrm{T}} \Phi_{j+j_0}$, where the square is to be understood componentwise. Thus the coefficients $\mathbf{c}^{j+j_0}(v^2)$ in $S(\Phi_{j+j_0})$ are computed (approximately) as

$$\mathbf{c}^{j+j_0}(v^2) = 2(\mathbf{c}^{j+j_0}(Q_j v))(\mathbf{d}^{j+j_0}(R_j v)) + (\mathbf{d}^{j+j_0}(R_j v))^2. \tag{7.90}$$

The justification for taking componentwise products of the coefficient sequences assumes the use of scaling functions whose shifts are orthogonal and which are *almost interpolatory* in the sense of (4.17). Recall that this is the case when the scaling function has sufficiently many vanishing moments (4.16); see Beylkin et al. (1991).

Note that at least one factor in each product on the right-hand side of (7.90) involves wavelet coefficients. These arrays are usually sparse, so that only significant products need be calculated. Accordingly, one should only compute those scaling function coefficients in $\mathbf{c}^{j+j_0}(Q_j v)$ affected by large wavelet coefficients. This requires suitably localized multiscale transformations. Depending on the context, the resulting (local) single-scale arrays $\mathbf{c}^{j+j_0}(v^2)$ can be used for point evaluations, or have to be transformed into wavelet representations. Therefore, the development of appropriate data structures is certainly an important issue.

A promising alternative is offered by an adaptation of Algorithm 2 from Section 7.4, by which an interpolating approximation is transformed into a wavelet representation. A different strategy is pursued in Joly et al. (1997); see also related work in Danchin (1997), Maday, Perrier and Ravel (1991). Joly et al. (1997) propose interpolating the reduced approximation $\tilde{u}^m$ at *all* the points x_λ corresponding to the entire basis. The values at these points are computed through a fast evaluation scheme (setting those coefficients to zero whose indices do not belong to Λ^m: see the algorithm in the previous section). The values of $(\tilde{u}^m)^2$ are then computed at each point and $(\tilde{u}^m)^2$ is interpolated with respect to the best basis. The overall cost is $\mathcal{O}(J2^J)$,

which unfortunately exceeds the number of significant coefficients in Λ^m. If the best bases need not be changed, the collocation can be based on the centres x_λ, $\lambda \in \Lambda^{m+1}$ (Danchin 1997).

7.11. Stokes and Navier–Stokes equations

Of course, the above evolution equations can be viewed as simplified test cases for the next higher mathematical model, namely the Navier–Stokes equations for *incompressible* fluids, which, properly normalized, read

$$\begin{aligned} \frac{\partial u}{\partial t} - \nu \Delta u + (u \cdot \nabla) u + \nabla p &= f, \\ \nabla \cdot u &= 0; \end{aligned} \tag{7.91}$$

see, for example, Girault and Raviart (1986). Here $u : \mathbb{R}^n \times \mathbb{R}^+ \to \mathbb{R}^n$ represents the velocity of the fluid, and $p : \mathbb{R}^n \times \mathbb{R}^+ \to \mathbb{R}$ the pressure, and ν is a positive number called the *kinematic viscosity*. So when ν gets small the formally parabolic first system becomes hyperbolic. One usually looks for (u, p) satisfying (7.91) in some domain Ω subject to initial and boundary conditions

$$u(\cdot, 0) = u_0 \quad \text{in } \Omega, \quad u = 0 \text{ on } \partial\Omega \times [0, T], \tag{7.92}$$

and

$$\int_\Omega p(x, t)\, \mathrm{d}x = 0, \quad t \in (0, T), \tag{7.93}$$

since p is only determined up to a constant. In this section we outline some recent attempts that tackle the numerical solution of this kind of problem by means of wavelet discretizations.

Amongst the difficulties in treating (7.91) numerically is the constraint $\nabla \cdot u = 0$. One way to avoid this is to write (7.91) in the so-called *vorticity stream function formulation*

$$\begin{aligned} \frac{\partial u}{\partial t} + u \cdot \nabla \omega - \nu \Delta \omega &= 0, \\ \nabla \times u &= \omega, \end{aligned} \tag{7.94}$$

which is valid in this form for $n = 2$; see, for example, Quartapelle (1993).

For $\Omega = \mathbb{R}^2/\mathbb{Z}^2$ and periodic boundary conditions, Algorithm 2 (see Section 7.4) is applied to (7.94) by Fröhlich and Schneider (1996). Wavelet schemes based on the methodologies described in Section 7.7 are applied to (7.94), among other model problems, in Charton and Perrier (1996), accompanied by a complexity analysis which indicates that the complexity of the scheme is proportional to the dimension of the highest resolution. Although working with possibly highly lacunary sets Λ of wavelet indices, one has to employ at some stage the transformation between single and

multiscale representation. So part of the principal efficiency is diminished again by this typical bottleneck. On the other hand, if this is the only place where the complexity of the full spaces enters, the constants appearing in the complexity estimates can be expected to be moderate.

To become competitive with the existing numerical methods, it is conceptually important to overcome the restriction to the vorticity stream function formulation, which is convenient only in the two-dimensional case. When working with the primitive variables u, p in (7.91), it is known that the constraint $\nabla \cdot u = 0$ imposes certain compatibility conditions on the trial spaces for velocity and pressure that are necessary for an asymptotically stable solution procedure Suitable families of such finite element spaces are known (Girault and Raviart 1986). However, for $n = 3$, they become quite involved when trying to raise the order of exactness. We will therefore briefly discuss what potential contributions of wavelet concepts in this regard can be expected.

Saddle point formulation
Suppose we fix trial spaces $V_h \subset (H_0^1(\Omega))^n$ and

$$M_h \subset L_{2,0}(\Omega) = \{g \in L_2(\Omega) : \int_\Omega g(x)\,dx = 0\}.$$

A semi-implicit discretization of (7.91) in conjunction with a weak formulation of the corresponding linear problem yields

$$\begin{aligned} \langle u_h^{m+1}, v_h\rangle + \Delta t\nu\langle \nabla u_h^{m+1}, \nabla v_h\rangle + \langle v_h, \nabla p\rangle &= \langle f - u_h^m \cdot \nabla u_h^m, v_h\rangle, \quad v_h \in V_h, \\ \langle \operatorname{div} u_h^{m+1}, \mu_h\rangle &= 0, \quad \mu_h \in M_h. \end{aligned} \tag{7.95}$$

One may question for which time steps and under which circumstances it is reasonable to use an explicit discretization of the transport term $u_h^m \cdot \nabla u_h^m$ and put it on the right-hand side. But for the time being we ignore this point and remark that (7.95) corresponds to the linear system of equations

$$\begin{pmatrix} \mathbf{A}_{h,\alpha} & \mathbf{B}_h^{\mathrm{T}} \\ \mathbf{B}_h & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{u}_h \\ \mathbf{p}_h \end{pmatrix} = \begin{pmatrix} \mathbf{F}_h \\ \mathbf{0} \end{pmatrix}. \tag{7.96}$$

Here $\mathbf{A}_{h,\alpha}$ is the stiffness matrix of the operator $\mathcal{L} = I - \alpha\Delta$, $\alpha = \Delta t\nu$, and $\mathbf{B}_h^{\mathrm{T}}$ is the discretization of the gradient ∇ relative to the chosen bases in V_h, M_h. Recall from Section 2.2 (b) that the stationary Stokes problem (2.7) leads to an analogous system where $\mathbf{A}_{h,\alpha}$ is replaced by the discretization of $\mathcal{L} = \nu\Delta$. Both share the same operator $\mathbf{B}_h$, however.

To ensure that the discretizations are stable, that is, that the inverses of the discretized operators $\mathcal{L}_h$ are uniformly bounded, the pairs (V_h, M_h) have to satisfy the *Ladyšenskaya–Babuška–Brezzi condition* (LBB), which means

that the *inf–sup condition* (2.13) holds uniformly in h

$$\inf_{\mu_h \in M_h} \sup_{v_h \in V_h} \frac{b(v_h, \mu_h)}{\|v_h\|_V \|\mu_h\|_M} \geq \beta > 0, \tag{7.97}$$

where as before $b(v, \mu) = \langle \operatorname{div} v, \mu \rangle$. The following well-known fact characterizes the validity of (7.97); see Fortin (1977).

Proposition 7.9 Suppose (2.13) holds for the pairs (V, M) and some $\tilde{\beta} > 0$. Then the subspaces V_h, M_h satisfy (7.97) uniformly in h, if and only if there exist linear operators $Q_h : V \to V_h$ satisfying

$$\|Q_h v\|_V \lesssim \|v\|_V, \quad v \in V, \tag{7.98}$$

and

$$b(v - Q_h v, \mu_h) = 0, \quad v \in V, \quad \mu_h \in M_h. \tag{7.99}$$

While in the finite element context this observation is primarily of theoretical use, it does offer a constructive angle in the wavelet setting. We briefly sketch the approach of Dahmen et al. (1996*c*). In fact, (7.99) may be viewed as a biorthogonality condition with respect to $b(\cdot, \cdot)$. Again this is most conveniently explained first for the case $\Omega = \mathbb{R}^n$.

To describe this consider any dual pair of biorthogonal compactly supported generators $\phi, \tilde{\phi} \in L_2(\mathbb{R}^n)$ (see (4.19)). Assuming that $\phi \in H^{1+\varepsilon}(\mathbb{R})$, the procedure mentioned in Section 4.2 (see (4.34)) yields another dual pair $(\phi^-, \tilde{\phi}^+)$ of biorthogonal compactly supported generators. More generally, let us set

$$\psi_0^- := \phi^-, \quad \psi_1^- := \psi^-, \quad \tilde{\psi}_0^+ := \tilde{\phi}^+, \quad \tilde{\psi}_1^+ := \tilde{\psi}^+,$$

and likewise $\psi_i, \tilde{\psi}_i$, $i = 0, 1$. Here $\psi^-, \tilde{\psi}^+$ are the corresponding compactly supported new biorthogonal mother wavelets, which, by (4.36), arise from ψ and $\tilde{\psi}$ essentially by *differentiation* and *integration*, respectively. The trial spaces on $\mathbb{R}^n$ are again obtained by taking tensor products. In particular, the multivariate scaling functions and mother wavelets $\psi_e^{-,i}(x), \tilde{\psi}_e^{+,i}(x)$ are obtained for $e \in E = \{0, 1\}^n$, $i = 1, \ldots, n$, by replacing in (4.38) the ith factor by $\psi_{e_i}^-(x_i), \tilde{\psi}_{e_i}^+(x_i)$, respectively. Specifically, we set $\phi^{-,i} =: \psi_0^{-,i}$, $\tilde{\phi}^{+,i} =: \tilde{\psi}_0^{+,i}$.

Now let

$$V_j := \{v \in V : v_i \in S(\tilde{\Phi}_j^{+,i}), i = 1, \ldots, n\} \tag{7.100}$$

and

$$M_j := S(\Phi_j). \tag{7.101}$$

Thus we can also write

$$V_j = S(\tilde{\Phi}_{j,\otimes}^+), \quad \tilde{\Phi}_{j,\otimes}^+ := \tilde{\Phi}_j^{+,1} \times \cdots \times \tilde{\Phi}_j^{+,n}, \tag{7.102}$$

and the pairs $\tilde{\Phi}^+_{j,\otimes}, \Phi^-_{j,\otimes}$ are biorthogonal, where $\Phi^-_{j,\otimes}$ is defined analogously. Hence

$$Q_j v := \langle v, \Phi^-_{j,\otimes}\rangle \tilde{\Phi}^+_{j,\otimes} \tag{7.103}$$

are projectors from V onto V_j. Since

$$v - Q_j v = \sum_{l=j+1}^{\infty} (Q_{l+1} - Q_l)v \tag{7.104}$$

and

$$((Q_{l+1} - Q_l)v)_i = \langle v_i, \Psi_l^{-,i}\rangle \tilde{\Psi}_l^{+,i},$$

we conclude from (4.36) that

$$\frac{\partial}{\partial x_i}((Q_{l+1} - Q_l)v)_i = -4\langle v_i, \Psi_l^{-,i}\rangle \tilde{\Psi}_l. \tag{7.105}$$

But since biorthogonality of Ψ and $\tilde{\Psi}$ ensures that $\langle \Phi_j, \tilde{\Psi}_l\rangle = 0$, $l \geq j$, one immediately infers from (7.101), (7.104) and (7.105) that

$$\langle \text{div } (v - Q_j v), \mu_j\rangle = 0, \quad \text{for all } \mu_j \in M_j = S(\Phi_j). \tag{7.106}$$

Moreover, for $\tilde{\phi} \in H^{1+\varepsilon}$, Theorem 5.8 implies that the Q_j are uniformly bounded on V. Thus Proposition 7.9 applies and confirms that the spaces V_j, M_j defined by (7.100) and (7.101) do satisfy the LBB condition (7.97).

There is no difficulty in adapting this construction to the periodic case. It is perhaps more interesting to note that it can also be extended to $\Omega = \square = (0,1)^n$ and homogeneous boundary conditions (7.92). This is done in Dahmen et al. (1996*c*) by starting with a dual pair $\phi, \tilde{\phi}$ as above and constructing biorthogonal refinable bases $\Phi_j, \tilde{\Phi}_j$ adapted to $[0,1]$ as indicated in Section 4.4. The key is that the modified dual pair $\phi^-, \tilde{\phi}^+$ again gives rise to pairs of refinable biorthogonal bases $\Phi_j^-, \tilde{\Phi}_j^+$ where, however, we now have the inclusion $S(\tilde{\Phi}_j^+) \subset H_0^1(\square)$. Defining the collections $\Psi_j^{-,i}, \tilde{\Psi}_j^{+,i}$ in analogy to the previous construction, one can prove that one still has

$$S\left(\frac{\partial}{\partial x_i}\tilde{\Psi}_j^{+,i}\right) = S(\tilde{\Psi}_j). \tag{7.107}$$

Thus the same reasoning as before shows that for analogously defined projectors Q_j (7.106) is still valid. The validity of the LBB condition (7.97) also follows in this case from Proposition 7.9.

Solving the linear systems

Since the matrix in (7.96) is indefinite, the solution of (7.96) by iterative methods requires a bit more care; see Bramble and Pasciak (1988). The upshot of all the options is that, whenever a good preconditioner for the (positive definite) matrix $\mathbf{A}_{j,\alpha}$, as well as for the *Schur complement*

$\mathbf{K}_{j,\alpha} := \mathbf{B}_j \mathbf{A}_{j,\alpha}^{-1} \mathbf{B}_j^{\mathrm{T}}$, is available, one can combine both so as to obtain a correspondingly efficient iterative scheme for the treatment of (7.96). An example is mentioned below in Section 8.3 in a different context. In the case of the stationary Stokes problem, the block $\mathbf{A}_{j,\alpha}$ corresponds to a stiffness matrix for the Laplace operator Δ. Asymptotically optimal preconditioners for this component were discussed in Section 6. In this case the Schur complement is an operator of order zero and hence does not require any further preconditioning. In the time-dependent case, the roles are reversed: $\mathbf{A}_{j,\alpha}$ is a discretization of the Helmholtz operator $I - \alpha\Delta$, which, for small α, resembles the identity. Thus $\mathbf{A}_{j,\alpha}$ is already moderately well conditioned. Nevertheless, the robust wavelet-based preconditioners discussed in Section 6.6 would here cover the full range of possible values of α. Now, for small α the Schur complement tends more and more to a second-order operator, so that preconditioning becomes necessary; see Bramble and Pasciak (1994). Again, an asymptotically optimal preconditioner-based on the above wavelet bases, namely Algorithm 1 in Section 6.2, is proposed by Dahmen et al. (1996*c*). The concrete examples considered there are based on dual pairs $\phi, \tilde{\phi}$ where $\tilde{\phi}$ is chosen as a B-spline (see Section 4.4). All basis functions and wavelets for V_j and M_j have compact support. The construction allows one to realize any desired order of exactness for any spatial dimension. The numerical experiments in Dahmen et al. (1996*c*) for the linearized problem cover two- and three-dimensional examples and confirm the predicted asymptotic optimality, that is, iteration numbers are independent of the size of the problem.

Divergence-free wavelets

Instead of seeking pairs of trial spaces V_j, M_j satisfying the LBB conditions, one could try to find trial spaces V_j which satisfy the constraint div $v = 0$, $v \in V_j$ *weakly*, that is, $b(v, \mu) = 0$ for all $\mu \in M_j$, $v \in V_j$. This has been realized in the finite element context but corresponding constructions are rather involved, in particular for the 3D case. One could even go one step further and try to construct spaces $V_j^0 \subset V^0 := \{v \in (H_0^1(\Omega))^n : \text{div } v = 0\}$. Orthogonal *divergence-free* wavelets have been constructed by Battle and Federbusch. These wavelets have necessarily global support (Lemarié-Rieusset 1994) although they decay exponentially. A somewhat different line based on Section 4.2 has been pursued by Jouini (1992) and Lemarié-Rieusset (1992). Dispensing with orthogonality, one can construct *divergence-free biorthogonal wavelets* with *compact* support. This point was taken up by Urban (1995*a*, 1995*b*), where, in addition to tensor products, *genuinely multivariate* divergence-free wavelets are constructed.

Using such trial spaces, the weak formulation (7.95) reduces to

$$\langle u_j^{m+1}, v_j \rangle + \Delta t \nu \langle \nabla u_j^{m+1}, \nabla v_j \rangle = \langle F, v_j \rangle, \quad v_j \in V_j^0, \tag{7.108}$$

that is, to a Helmholtz problem on V^0. Here F collects the terms on the right-hand side of (7.95).

First, numerical experiences for the 3D case are reported by Urban (1995*c*, 1996). These experiments concern classical Galerkin schemes. On the other hand, the above reduction to the Helmholtz problem suggests the following interesting alternative. Because of the constraint $\operatorname{div} u = 0$, the fast vaguelette evaluation schemes based on orthonormal wavelets have been confined to the vorticity stream function formulation and thus to the bivariate case. Recalling that the vaguelette approach can be extended to biorthogonal wavelet bases, one can combine it with the above divergence-free wavelets, which also work in the three-dimensional case.

We conclude this section with some brief remarks on the construction. The key observation (Lemarié-Rieusset 1992, Urban 1995*a*, Urban 1995*b*) is the commutation property

$$\frac{\partial}{\partial x_i}\langle v, \tilde{\Phi}_j\rangle \Phi_j = \left\langle \frac{\partial}{\partial x_i} v, \tilde{\Phi}_j^{+,i} \right\rangle \Phi_j^{-,i}. \tag{7.109}$$

To make use of this fact for the construction of divergence-free wavelets, one has to iterate the modifications from Section 4.2 by setting

$$\phi^{-,(i,l)} = (\phi^{-,i})^{-,l},$$

and analogously for $\tilde{\phi}, \psi$ and $\tilde{\psi}$. For $i \in \{1, \ldots, n\}$ let $\mathcal{N}_i = \{1, \ldots, n\} \setminus \{i\}$ and define functions $\psi_{e,\nu}^{\nabla}$, $\nu \in \{1, \ldots, n\}$, by

$$\left(\psi_{e,\nu}^{\nabla}\right)_i = \begin{cases} 0, & i \notin \{\nu, i_e\}, \\ \psi_e^{-,\mathcal{N}_\nu}, & i = \nu, \\ -\frac{1}{4}\frac{\partial}{\partial x_\nu}\psi_e^{-,\mathcal{N}_\nu \setminus \{i_e\}}, & i = i_e, \end{cases}$$

where $i_e \in \{1, \ldots, n\}$ is any index such that $e_i = 1$. The collections

$$\Psi^{\nabla} := \{\psi_{e,\nu,j,k}^{\nabla} : e \in E_*, \nu \neq i_e, j \in \mathbb{Z}, k \in \mathbb{Z}^n\}$$

can then be shown to be a divergence-free wavelet basis (Urban 1995*a*, 1995*b*). Again, further analysis, implementations and numerical experiments can be found in Urban (1995*b*), (1995*c*) and (1996).

8. Extension to more general domains

Except for the extensions to wavelets on cubes (see Section 4.4 and the comments in the previous section) all approaches described so far rely in an essential way on the underlying stationary shift-invariant structure of the discretization. It has long been known in numerical analysis that, beyond mere asymptotic estimates, regular discretizations often support efficiency in many ways, reflected by superconvergence effects, for instance. Therefore it may pay in the end in many situations to exploit such advantages for

the bulk of computation and treat boundary effects separately. The matrix capacitance method is an example of such a strategy which has been extensively studied in connection with finite difference schemes. Moreover, when dealing with problems where the geometry changes in time, such a concept may even be a necessity rather than an option. One possibility that may come to mind first is to enforce essential boundary conditions by means of penalty terms; see, for example, Glowinski, Pan, Wells and Zhou (1996) and Glowinski, Rieder, Wells and Zhou (1993). Aside from accuracy issues, a conceptual difficulty with this approach seems to be that it stiffens the problem significantly and thereby wastes previously gained advantages on the preconditioning side. A relatively simple alternative is to refine the spaces near the boundary. Incorporating additional basis functions on higher discretization levels whose supports are still inside the domain can compensate the loss of accuracy encountered otherwise. From a complexity point of view this works in the bivariate case but no longer for domains in $\mathbb{R}^3$ (Jaffard 1992, Oswald 1997). Therefore we will now concentrate on three alternative possibilities.

8.1. An extension technique

Throughout this subsection assume that $\mathcal{L}$ is a selfadjoint elliptic operator so that, for $a(\cdot,\cdot)$ defined by (2.5), with respect to natural boundary conditions, the problem in variational form is to find u in $H = H^t(\Omega)$ such that

$$a(u,v) = \langle f, v\rangle, \quad v \in H \tag{8.1}$$

for some $f \in H^*$ (*cf.* (6.9)). We briefly sketch some ideas from Oswald (1997) that fit into the multilevel Schwarz concepts described in Section 6.5.

The starting point is a nested sequence of finite element or spline spaces S_j, $j \in \mathbb{N}_0$, defined on regular meshes of types 1 or 2 (see Section 6.6). Now $\Omega \subset \mathbb{R}^n$ is supposed to be an arbitrary bounded domain with sufficiently regular boundary to admit the existence of extension operators E, which are bounded in an appropriate Sobolev scale. For instance, the validity of a *uniform cone condition* or Lipschitz boundaries would do (Johnen and Scherer 1977).

The first step is to construct collections $\Phi_{j,\Omega}$ consisting mainly of functions $\phi(2^j \cdot -k)$, $k \in \mathbb{Z}^n$, whose support does not intersect $\partial\Omega$, where ϕ is in this case a tensor product B-spline, say. In addition, one needs functions that are adapted to the boundary. Their restriction to Ω is supported in a margin of width $\sim 2^{-j}$ along the boundary. They consist of fixed linear combinations of $\phi(2^j \cdot -k)$ designed in such a way that the span of the entire collection Φ_j contains all polynomials up to some degree $d-1$ on Ω. This is similar to the ideas presented in Section 4.4 and to the recent developments in Cohen, Dahmen and DeVore (1995). However, in order to keep these boundary-near

cluster functions as simple as possible, they are, in contrast to Cohen et al. (1995), *not* required to be refinable. Thus one generally has

$$S(\Phi_{j,\Omega}) \not\subset S(\Phi_{j+1,\Omega}). \tag{8.2}$$

However, along with $\Phi_{j,\Omega}$, a biorthogonal collection $\Xi_{j,\Omega}$ is constructed in such a way that the projectors

$$Q_j v := \langle v, \Xi_{j,\Omega}\rangle \Phi_{j,\Omega} \tag{8.3}$$

satisfy the direct estimates (5.7). Essential hypotheses are that the $\phi(\cdot - k)$ are *locally* linearly independent (that is, the vanishing of a linear combination on any neighbourhood implies that the coefficients of the overlapping translates $\phi(\cdot - k)$ are zero), as well as the availability of extension operators.

Due to the lack of nestedness (8.2), the techniques from Section 6.5 cannot yet be applied directly. To remedy this, Oswald (1997) shows how to construct another sequence of nested spaces $\hat{S}_j \subset H^t(\mathbb{R})$ spanned by suitably chosen B-splines (on all levels $l \leq j$) which overlap Ω. In addition, appropriate restriction and extension operators

$$R_j : \hat{S}_j \to S(\Phi_{j,\Omega}), \quad E_j : S(\Phi_{j,\Omega}) \to \hat{S}_j,$$

are identified. In fact, $R_j v = \langle v, \Xi_{j,\Omega}\rangle \Phi_{j,\Omega}$. The E_j have the form

$$E_J v_J := \sum_{j=0}^{J} (P_j - P_{j-1}) v_J \in \hat{S}_J,$$

where the P_j are similar quasi-interpolant type operators as the Q_j in (8.3) above. In fact, $P_j v = \langle v_1 \Xi_j\rangle \Phi_j$, where the elements $\xi_{j,k} \in \Xi_j$ are either supported in Ω when $\operatorname{supp} \phi(2^j \cdot - k) \cap \Omega \neq \emptyset$, or zero otherwise. To establish suitable norm estimates for these operators, one needs certain additional requirements on the domain which, for instance, ensure that the margin of boundary affected basis functions has width $\sim 2^{-j}$ on level j. One can then prove that (Oswald 1997)

$$R_j E_j v_j = v_j, \quad v_j \in S(\Phi_{j,\Omega}),$$

and

$$\begin{aligned} a(R_j \hat{v}_j, R_j \hat{v}_j) &\lesssim \|\hat{v}_j\|^2_{H^t(\mathbb{R}^n)}, \quad \hat{v}_j \in \hat{S}_j, \\ \|E_j v_j\|^2_{H^t(\mathbb{R}^n)} &\lesssim a(v_j, v_j), \quad v_j \in S(\Phi_{j,\Omega}). \end{aligned} \tag{8.4}$$

One can then proceed as follows. Fix any $H^t(\mathbb{R}^n)$-elliptic form $\hat{a}(\cdot,\cdot)$ and determine a preconditioner $\hat{C}_j$ on $\hat{S}_j$ for the operator $\hat{\mathcal{L}}_j$ defined by $\hat{a}(\hat{u}_j, \hat{v}_j) = \langle \hat{\mathcal{L}}_j \hat{u}_j, \hat{v}_j\rangle$, $\hat{u}_j, \hat{v}_j \in \hat{S}_j$, by the methods from Section 6.5. Then, defining

$$\mathcal{C}_j := R_j \hat{C}_j R_j^* \tag{8.5}$$

and $\mathcal{L}_j$ by $\langle \mathcal{L}u_j, v_j\rangle_\Omega = a(u_j, v_j)$, $u_j, v_j \in S(\Phi_{j,\Omega})$, the operators $\mathcal{C}_j\mathcal{L}_j$ satisfy

$$\kappa_2(\mathcal{C}_j\mathcal{L}_j) \sim 1 \quad \text{if} \quad \kappa_2(\hat{\mathcal{C}}_j\hat{\mathcal{L}}_j) \sim 1. \tag{8.6}$$

The proof relies on the fictitious space Theorem 6.9 (Nepomnyaschikh 1990); see Oswald (1997) for details. The treatment of essential boundary conditions and further extensions are also discussed in Oswald (1997).

In this form the scheme does *not* make explicit use of any wavelet basis or a corresponding exact representation of complement components. Hence it is tailored to the selfadjoint case but otherwise very flexible in connection with many standard discretizations.

8.2. Boundary value correction

Consider operators of the form (2.4), that is, $\mathcal{L} = -\mathrm{div}(A(x)\nabla) + a(x)\mathcal{I}$. Suppose $\Omega \subset \mathbb{R}^n$ is a bounded domain. The following strategy for solving

$$\mathcal{L}u = f \quad \text{on } \Omega, \quad \mathcal{B}u|_{\partial\Omega} = g, \tag{8.7}$$

where $\mathcal{B}$ is some boundary value operator, has been proposed in Averbuch, Beylkin, Coifman and Israeli (1995). Without loss of generality one may assume that $\Omega \subset \square := (0,1)^n$.

(1) Determine a smooth extension f_{ext} and an operator $\mathcal{L}_{\mathrm{ext}}$ of f and $\mathcal{L}$ respectively, from Ω to $\square$.

(2) Solve the problem

$$\mathcal{L}_{\mathrm{ext}}u = f_{\mathrm{ext}} \quad \text{on} \quad \square \tag{8.8}$$

with periodic boundary conditions.

(3) Given the solution u_{ext} of (8.8), solve the *homogeneous* problem

$$\mathcal{L}u = 0 \quad \text{on} \quad \Omega \tag{8.9}$$

subject to the boundary conditions

$$\mathcal{B}u|_{\partial\Omega} = g - \mathcal{B}u_{\mathrm{ext}}|_{\partial\Omega}, \tag{8.10}$$

with the aid of a boundary integral method (see Section 2.2 (d)).

Averbuch et al. (1995) only address (8.8), arguing that efficient methods for (8.9), (8.10) are available. The rationale is that fast wavelet methods such as those described in Section 7 do a particularly efficient job on the bulk of the problem. In fact, in the periodic setting, the significant wavelet coefficients are indeed determined by the significant wavelet coefficients of the right-hand side in the following sense. Suppose that $\Lambda_{f,\varepsilon}$ is the subset of wavelet coefficients needed to represent f_{ext} on $\square$ with accuracy ε. Then the set $\Lambda_{u,\varepsilon}$ of coefficients needed to represent the solution u_{ext} with accuracy ε is contained in a certain *'neighbourhood'* of $\Lambda_{f,\varepsilon}$, that is, a somewhat

larger set containing $\Lambda_{f,\varepsilon}$ where $\#\Lambda_{u,\varepsilon}$ is claimed to be proportional to $\#\Lambda_{f,\varepsilon}$. However, it is not apparent how this proportionality depends on ε and on the norm with respect to which accuracy is measured. Nevertheless, according to Theorem 6.2, a diagonally preconditioned conjugate gradient scheme constrained to the space $S(\Psi_{\Lambda_{u,\varepsilon}})$ would produce (perhaps combined with nested iteration) an approximate solution of accuracy ε at the expense of $\mathcal{O}(\#\Lambda_{u,\varepsilon})$ operations. For higher dimensions, in particular, this seems very tempting, since a high degree of adaptivity can be obtained without worrying about the substantial complications caused by mesh refinement strategies in conventional finite-difference or finite-element schemes.

On the other hand, there are still many points that need to be carefully addressed.

(1) If the boundary $\partial\Omega$ is fairly regular, a standard multilevel finite element scheme, at least in the 2D case, combined with the existing adaptive refinement schemes (see, for instance, Bornemann, Erdmann and Kornhuber (1996), Bank and Weiser (1985)) applied directly to the problem on Ω would realize at least the same favourable complexity.

(2) If the boundary has very little regularity, it is not clear how to properly balance the regularity of the extension to avoid introducing artificial singularities, and how to realize the extension numerically.

(3) For problems in $\mathbb{R}^n$ with $n > 2$, and nonconstant diffusion matrix $A(x)$ in (2.4), the treatment of the boundary integral equation arising from (8.9) and (8.10) may no longer be so trivial, let alone the extension problem.

Nevertheless, this approach offers a methodology for separating the bulk of computation in the highest spatial dimension from the boundary treatment.

8.3. Lagrange multipliers

The following alternative is in principle by no means new, but has been to some extent revived by the development of wavelet schemes; see, for example, Babuška (1973) and Brezzi and Fortin (1991). The idea of appending essential boundary conditions by means of *Lagrange multipliers* has been taken up again and analysed from the point of view of multilevel schemes, in Kunoth (1994) and Kunoth (1995). Suppose that $\hat{\Omega}$ is a cube containing Ω, let $a(u,v) = \langle \mathcal{L}_{\text{ext}} u, v\rangle_{\hat{\Omega}}$ and $M = \left(H^{s-\beta}(\partial\Omega)\right)^*$, when $\mathcal{B}$ maps $H^s(\Omega)$ onto $H^{s-\beta}(\partial\Omega)$. Choosing $H = H^s(\hat{\Omega})$ or $H = H_0^s(\hat{\Omega})$ or the subspace $H_p^s(\hat{\Omega})$ consisting of periodic functions in $H^s(\hat{\Omega})$ and defining

$$b(v,\mu) := \langle v, \mu\rangle_{\partial\Omega}, \tag{8.11}$$

the corresponding weak formulation of (8.7) requires finding $(u,p) \in H \times M$ such that

$$\begin{aligned} a(u,v) + b(v,p) &= \langle f, v\rangle_{\hat{\Omega}}, \quad v \in H, \\ b(u,\mu) &= g, \qquad\quad \mu \in M. \end{aligned} \tag{8.12}$$

The solution (u,p) of (8.12) solves the *saddle-point problem*

$$\inf_{v\in H}\sup_{\mu\in M}\left\{\frac{1}{2}a(v,v) + b(v,\mu) - \langle f, v\rangle_{\hat{\Omega}} - b(g,\mu)\right\} \tag{8.13}$$

(recall (7.91) and (7.95)). For general conditions under which (8.12) and (8.13) are equivalent see Brezzi and Fortin (1991). It is also known that, for instance for $\mathcal{L} = -\Delta + a$ and $\mathcal{B} = \mathcal{I}$, the Lagrange multiplier p in the solution of (8.12) agrees with $\frac{\partial u}{\partial \nu}$ on $\partial\Omega$ where $\partial\nu$ denotes the derivative in the direction of the outward normal of $\partial\Omega$.

Let us again denote by

$$\begin{pmatrix} \mathcal{A}_h & \mathcal{B}_h^* \\ \mathcal{B}_h & 0 \end{pmatrix}\begin{pmatrix} u \\ p \end{pmatrix} = \begin{pmatrix} f \\ g \end{pmatrix} \tag{8.14}$$

the operator equation corresponding to (8.12) projected on $S_h \times M_h \subseteq H \times M$. Recall that solving (8.14) (and hence solving (8.7) approximately for the above choice of $a(\cdot,\cdot)$ and $b(\cdot,\cdot)$), requires addressing the following two issues.

- Ensure that (S_h, M_h) satisfies the corresponding LBB condition.
- Find an efficient iteration scheme coping with the fact that the matrices in (8.14) are indefinite.

The first issue depends on the particular situation at hand; see Bramble (1981) and Glowinski et al. (1996). Following Bramble and Pasciak (1988), Kunoth (1994) and Kunoth (1995), the second task can be tackled for instance as follows.

Suppose that the selfadjoint positive definite operator $\mathcal{C}_h$ is a preconditioner for $\mathcal{A}_h$ satisfying

$$\langle \mathcal{C}_h^{-1} v, v\rangle \sim \langle \mathcal{A}_h v, v\rangle, \quad \langle (\mathcal{A}_h - \mathcal{C}_h)v, v\rangle \le \eta \langle \mathcal{A}_h v, v\rangle, \quad v \in S_h. \tag{8.15}$$

Moreover, assume that $\mathcal{K}_h$ is a preconditioner for the Schur complement, that is,

$$\langle \mathcal{K}_h^{-1}\mu, \mu\rangle_{\partial\Omega} \sim \langle \mathcal{B}_h \mathcal{A}_h^{-1}\mathcal{B}_h^*\mu, \mu\rangle_{\partial\Omega}, \quad \mu \in M_h. \tag{8.16}$$

According to Bramble and Pasciak (1988) (see also Kunoth (1995)), one can use the fact that (8.14) is equivalent to

$$\mathcal{M}_h\begin{pmatrix} u \\ p \end{pmatrix} := \begin{pmatrix} \mathcal{C}_h\mathcal{A}_h & \mathcal{C}_h\mathcal{B}_h^* \\ \mathcal{B}_h(\mathcal{C}_h\mathcal{A}_h - \mathcal{I}) & \mathcal{B}_h\mathcal{C}_h\mathcal{B}_h^* \end{pmatrix}\begin{pmatrix} u \\ p \end{pmatrix} = \begin{pmatrix} \mathcal{C}_h f \\ \mathcal{B}_h\mathcal{C}_h f - g \end{pmatrix}, \tag{8.17}$$

and then show that, under the assumption (8.15), $\mathcal{M}_h$ is positive definite relative to the inner product

$$\left[\begin{pmatrix} v \\ \mu \end{pmatrix}, \begin{pmatrix} w \\ \nu \end{pmatrix}\right] := \left\langle (\mathcal{A}_h - \mathcal{C}_h^{-1})v, w \right\rangle_{\hat{\Omega}} + \langle \mu, \nu \rangle_{\partial\Omega}.$$

Moreover, when in addition (8.16) holds, one can verify that

$$\kappa_2 \left(\mathcal{G}_h^{1/2} \mathcal{M}_h \mathcal{G}_h^{1/2} \right) \sim 1, \quad h \to 0, \tag{8.18}$$

where

$$\mathcal{G}_h := \begin{pmatrix} \mathcal{I} & 0 \\ 0 & \mathcal{K}_h \end{pmatrix}.$$

Kunoth (1994, 1995) has shown how to construct preconditioners $\mathcal{C}_h$ and $\mathcal{K}_h$ based on multilevel decompositions of appropriate trial spaces $S_j = S_{h_j}$, $M_j = M_{h_j}$, $j \in \mathbb{N}_0$. For $\mathcal{C}_h$ one could use a multilevel Schwarz scheme or a wavelet-based preconditioner, as detailed in Sections 6.2 (Algorithm 1), 6.4 and 6.5. The operator $\mathcal{B}_h \mathcal{A}_h^{-1} \mathcal{B}_h^*$ takes $\left(H^{s-\beta}(\partial\Omega)\right)^* = H^{\beta-s}(\partial\Omega)$ into $H^{s-\beta}(\partial\Omega)$ and thus has typically *negative* order and BPX or Schwarz schemes do not apply directly. The following strategy is suggested by the results in Kunoth (1995). Let $\Gamma = \partial\Omega$ and suppose that Ψ^Γ, $\tilde{\Psi}^\Gamma$ are biorthogonal Riesz bases for $L_2(\Gamma)$ with corresponding single-scale bases $\Phi_j^\Gamma, \tilde{\Phi}_j^\Gamma$. Let

$$Q_j^\Gamma \mu = \langle \mu, \tilde{\Phi}_j^\Gamma \rangle_\Gamma \Phi_j^\Gamma,$$

and define $\mathcal{B}_j$ by

$$\langle \mathcal{B}_j v, \mu \rangle_\Gamma = \langle Q_j^\Gamma \mathcal{B} v, \mu \rangle_\Gamma, \quad v \in S_{j+j_0}, \quad \mu \in S(\tilde{\Phi}_j^\Gamma),$$

that is, $M_j = S(\tilde{\Phi}_j^\Gamma)$. Here the choice of $j_0 \in \mathbb{Z}$ leaves some flexibility for satisfying the LBB condition. One can then realize an asymptotically optimal preconditioner $\mathcal{K}_j = \mathcal{K}_{h_j}$ for the Schur complement $\mathcal{B}_j \mathcal{A}_j^{-1} \mathcal{B}_j$ with the aid of the change of bases scheme (Algorithm 1) from Section 6.2. For further details see Kunoth (1995).

The tempting aspect of this strategy is that it has the potential to be extended to a wider class of problems. For instance, using divergence-free wavelets for discretizing the Stokes problem on a cube or torus and appending boundary conditions by Lagrange multipliers leads to the type of saddle-point problem considered above with $\mathcal{L} = -\Delta$. Moreover, in view of (8.17), one need not deal with the exact Schur complement but retain sparse representations of the zero-order operator on Γ.

On the other hand, one needs suitable multiscale bases Ψ^Γ, $\tilde{\Psi}^\Gamma$ on Γ. When $n = 2$, Γ is a curve and one can readily resort, at least for sufficiently smooth curves, to periodic univariate wavelets, or to composite wavelet bases of the

type considered in Section 4.4. When $n > 2$, things are more complicated. However, due to the typically low order of the Schur complement, the bases $\Psi^\Gamma, \tilde{\Psi}^\Gamma$ generally need not be very regular. A more detailed discussion of constructing wavelet bases on manifolds such as closed surfaces, particularly in the context of boundary integral equations, is given in Sections 9 and 10.

9. Pseudo-differential and boundary integral equations

So far the discussion has been essentially confined to differential operators. Of course, the appearance of integral operators is also implicit in the vaguelette concept. Moreover, they occur explicitly in Section 8.2 as a solution component for treating partial differential equations (see also Section 2.2 (d)). In addition to the issue of preconditioning, the numerical treatment of integral operators or, more generally, of operators with global Schwartz kernel faces a further serious obstruction: conventional discretizations lead to *dense* matrices so that both assembling these matrices and solving the linear systems quickly become prohibitively expensive for realistic problems. In fact, direct solvers require the order of N^3 operations when N denotes the problem size, and each matrix vector multiplication in an iterative method is of the order N^2. A conceptual remedy is to perform the matrix vector multiplications only *approximately* within some tolerance. In many cases this indeed allows one to reduce the computational complexity to *almost linear* growth, if the analytical background of the problem is properly exploited. Examples of this type are *panel clustering* (Hackbusch and Nowak 1984, Hackbusch and Sauter 1993, Sauter 1992) or the closely related multipole expansions (Carrier, Greengard and Rokhlin 1988, Greengard and Rokhlin 1987, Rokhlin 1985). A similar finite difference-based approach is presented in Brandt and Venner (preprint) and Brandt and Lubrecht (1990).

Yet another direction has been initiated by the startling paper by Beylkin et al. (1991). As announced in Section 1.5 (c), the representation of certain integral operators in wavelet coordinates is nearly sparse (see Section 1.3). Roughly speaking, the idea is to replace the exact stiffness matrix $\mathbf{A}_J := \langle \mathcal{L}\Psi_J, \Psi_J \rangle^{\mathrm{T}}$ by a *compressed* matrix $\mathbf{A}_J^c$ arising from $\mathbf{A}_J$ by setting all entries below a given threshold to zero. Beylkin et al. (1991) have shown that the product $\mathbf{A}_J^c \mathbf{d}$, $\mathbf{d} \in \mathbb{R}^{N_J}$, $N_J = \dim S(\Phi_J)$, is still within accuracy ε from $\mathbf{A}_J \mathbf{d}$ if only the order of $N_J \log N_J$ entries in $\mathbf{A}_J^c$ are different from zero. This result has since started a number of investigations centred upon the following questions.

(1) How to deal with operators of *nonzero order*?
(2) What can be said about other schemes such as *collocation*?
(3) How to deal with *nonperiodic problems*, specifically with boundary integral equations on closed surfaces?
(4) What can be said about asymptotics, that is, how sparse can $\mathbf{A}_J^c$

be made while still guaranteeing that the solution exhibits the same asymptotic accuracy as the solution of the uncompressed system?

(i) Beylkin et al. (1991) consider a zero-order operator, so that no preconditioning is necessary. Important applications (see Section 2.2 (d)) involve operators of order different from zero, such as the single-layer potential operator. Aside from a possible need of regularization in such cases, preconditioning again becomes necessary. For operators of order minus one, a preconditioner-based on multigrid techniques was developed by Bramble, Leyk and Pasciak (1994), by introducing a suitable discrete norm for H^{-1}. One then obtains a fast method by combining this concept with any of the above mentioned fast matrix-vector multiplication schemes. It seems that, in the context of wavelet-based schemes, the preconditioning of operators of *any* order, as explained in Section 6.2, and its effect on matrix compression were first solved by Dahmen, Pröβdorf and Schneider (1993*b*) and Dahmen et al. (1994*b*).

(ii) While Beylkin et al. (1991) only consider a 'classical Galerkin scheme', in practice collocation is often preferred to Galerkin schemes as a discretization tool for integral operators, because it reduces the dimension of numerical integration. Comparatively little is known about stability criteria for collocation schemes in that context. The class of classical periodic pseudo-differential operators

$$(\mathcal{L}u)(x) = \sum_{k\in\mathbb{Z}^n} \sigma(x,k)\hat{u}(k)e^{2\pi ik\cdot x}, \tag{9.1}$$

where $\hat{u}(k)$ are the Fourier coefficients of u, was chosen by Dahmen et al. (1994*c*) as a model setting for studying the following issues: stability criteria for various types of elliptic pseudo-differential operators and various types of generalized Petrov–Galerkin discretization in a multiresolution context, as well as an asymptotic analysis of fast solution by compression techniques. The schemes considered there are of the type (6.3) with projectors of the form

$$P_j v = \sum_{k\in\mathbb{Z}^n/2^j\mathbb{Z}^n} \eta_{j,k}(v)\phi_{j,k},$$

where Φ_j are periodized refinable single-scale bases and

$$\eta_{j,k}(v) := 2^{-nj/2}\eta\Big(v(2^{-j}(\cdot + k))\Big)$$

arise from some fixed functional η defined on $S(\mathcal{L}\Phi_j)$. Thus $\eta = \delta(\cdot - \alpha)$ corresponds to collocation while $\eta = \phi$ covers the Galerkin scheme. The operators $\mathcal{L}$ under consideration are assumed to be *elliptic* in the sense that the principal part $\sigma_0(x,\xi)$ of their symbol $\sigma(x,\xi)$ is *coercive*, that is, for

some $\delta > 0$ one has

$$\operatorname{Re}\sigma_0(x,\xi) \gtrsim |\xi|^{2|t|}, \text{ for } |\xi| > \delta, \quad x \in \mathbb{R}^n/\mathbb{Z}^n. \tag{9.2}$$

The main result of Dahmen et al. (1994*c*) can be stated as follows. For any fixed $y \in \mathbb{R}^n/\mathbb{Z}^n$, let $\sigma_y(k) := \sigma_0(y,k)$ induce the constant coefficient operator $\mathcal{L}_y$. Let

$$\alpha_\eta(w,y) := \sum_{k\in\mathbb{Z}^n} \sigma_y(w+2\pi k)\hat{\phi}(w+2\pi k)\overline{\hat{\eta}(w+2\pi k)}$$

denote the *numerical symbol* relative to $\mathcal{L}_y$, which under suitable assumptions on ϕ is well defined. Here $\hat{\eta}$ is the Fourier transform of η in the distributional sense. The numerical symbol is called *elliptic* (see Wendland (1987)) if

$$|\alpha_\eta(w,y)| \gtrsim |w|^r, \quad \text{for } w \in \left[-\tfrac{1}{2},\tfrac{1}{2}\right]^n, \quad \text{and } y \in \mathbb{R}^n/\mathbb{Z}^n. \tag{9.3}$$

A freezing coefficient technique based on superconvergence results in connection with the so-called *discrete commutator property* is used by Dahmen et al. (1994*c*) to show that the generalized Petrov–Galerkin scheme (6.3) is (s,r)-*stable*, in the sense of (6.4), if and only if the scheme is (s,r)-stable for $\mathcal{L}_y$ for all $y \in \mathbb{R}^n/\mathbb{Z}^n$. This in turn finally yields that, under the above assumptions on $\mathcal{L}$, the scheme (6.3) is (s,r)-stable if and only if the numerical symbol α_η is elliptic in the sense of (9.3). Condition (9.3) naturally extends the stability condition (4.11), which refers to $\mathcal{L} = \mathcal{I}$.

This criterion is useful for verifying stability of collocation schemes where $\hat{\eta} = 1$ (compare with (7.28)); see, for instance, Dahmen et al. (1996*a*).

(iii) The above-mentioned results on periodic pseudo-differential equations immediately apply to boundary integral equations for two-dimensional domains with smooth boundary, which, via a smooth reparametrization, can be identified with the circle. Univariate periodic wavelets provide all the necessary tools for this case. Important contributions for Galerkin schemes are given by von Petersdorff and Schwab (1997*b*).

However, when the boundary integral equation lives on a surface of higher dimension, being able to treat periodic problems is ultimately *not* sufficient any longer. This puts conceptually new demands on the tools, that is, on the construction of appropriate wavelets. This issue will be addressed later in more detail.

To see how well the analysis of the periodic case predicts the right behaviour in more realistic situations, a multiscale collocation method for the double-layer potential equation on two-dimensional polyhedral surfaces in $\mathbb{R}^3$ was developed and tested by Dahmen et al. (1994*a*). The multiresolution spaces consist of continuous piecewise linear finite elements relative to uniform triangulations of the (triangular) faces of the polyhedron. The functions indicated in Figure 2 were used as wavelets. Since in this construction

the order of vanishing moments decreases near face edges, this approach was still provisional. Although, in contrast to the torus, the surfaces considered by Dahmen et al. (1994*a*) were no longer smooth, by and large the same compression and convergence behaviour could be observed as predicted by the analysis of the idealized situation. However, full practical use of these findings requires computing the compressed matrices $\mathbf{A}_J^c$ at costs which are essentially of the order of nonvanishing entries. Meanwhile substantial progress has been made in this regard, which we sketch later; see Dahmen and Schneider (1997*b*) and von Petersdorff and Schwab (1997*a*).

(iv) While in Beylkin et al. (1991) the compression rate referred to a *fixed* accuracy ε, an asymptotic analysis was carried out in Dahmen et al. (1993*b*), Dahmen et al. (1994*b*), von Petersdorff and Schwab (1997*b*) and von Petersdorff, Schneider and Schwab (1997). In particular, Schneider (1995) has shown that, under certain assumptions on the domain and on the wavelet bases, $\mathbf{A}_J^c$ can be compressed to $\mathcal{O}(N_J)$ nonvanishing entries while still realizing the asymptotic accuracy of the unperturbed scheme. Recently, significant progress on a practicable realization in terms of a nearby asymptotically optimal fully discrete scheme for zero-order operators was accomplished by von Petersdorff and Schwab (1997*a*).

In summary, the practical success of such concepts requires handling the following central tasks.

(a) Construct appropriate wavelet bases $\Psi, \tilde{\Psi}$ defined on a manifold Γ such that the underlying operator can be preconditioned well and efficiently compressed.
(b) Develop a scheme for computing the compressed operator at an expense that stays proportional to the number of nonvanishing entries.
(c) Combine these techniques with adaptive space refinement strategies, that is, with identifying sets $\Lambda \subset \nabla$ adapted to the problem at hand. By Remark 6.3, these together would provide an asymptotically optimal scheme.

In principle, all three goals are in sight. We will first sketch some basic ingredients of several contributions to (a) and (b).

9.1. Geometry considerations

The numerical treatment of realistic boundary integral equations obviously requires more than periodized wavelets. A natural starting point is the representation of the boundary manifold $\Gamma = \partial\Omega$. In the context of boundary integral equations, one is primarily interested in spatial dimensions $n = 1, 2$ of Γ. However, the same ideas also apply in principle to other manifolds, such as bounded domains in $\mathbb{R}^3$, so it is worth keeping n arbitrary at this point.

Concrete *free-form surface* representations are generated by CAD packages. There, a surface Γ is usually *parametrically* defined, that is, Γ is a disjoint union of (open) *patches* Γ_i,

$$\Gamma = \overline{\bigcup_{i=1}^{M} \Gamma_i}, \quad \Gamma_i \cap \Gamma_l = \emptyset, \quad i \neq l. \tag{9.4}$$

The global regularity of Γ is usually described with the aid of an *atlas* $\{(\hat{\Gamma}_i, \kappa_i)\}_{i=1}^{M}$. This consists of a covering $\Gamma = \bigcup_{i=1}^{M} \hat{\Gamma}_i$ and associated *regular* mappings

$$\kappa_i : \hat{\square}_i \to \hat{\Gamma}_i, \quad \hat{\square}_i \subset \mathbb{R}^n, \quad i = 1, \ldots, M,$$

that is, κ_i and κ_i^{-1} are smooth mappings so that, in particular, the corresponding functional determinant $|\partial \kappa_i(x)|$ does not vanish on $\hat{\Gamma}_i$. Moreover, for $\square \subset \bigcap_{i=1}^{M} \hat{\square}_i$, one has

$$\kappa_i|_{\square} = \Gamma_i, \quad i = 1, \ldots, M. \tag{9.5}$$

The set Γ is called a C^m manifold, respectively Lipschitz manifold, if the mappings are C^m, respectively Lipschitz. In practice, one does not work with coverings. Instead the global smoothness requirements are then translated into relations between the control parameters in the mappings κ_i corresponding to adjacent patches. Again, these considerations also apply to domain decompositions of domains in $\mathbb{R}^3$.

9.2. Function spaces on Γ

The discussion in Section 6 has made it very clear that the qualification of a wavelet basis Ψ for a given problem is closely related to the relevant function spaces. Thus one has to understand such function spaces defined on Γ. Denoting by ds the surface measure on Γ, the space $L_2(\Gamma)$ of square integrable functions on Γ is a Hilbert space with respect to the inner product

$$\langle u, v \rangle_{\Gamma} = \int_{\Gamma} u(x) \overline{v(x)} \, ds_x. \tag{9.6}$$

With the aid of the above atlas, one can also define Sobolev spaces $H^s(\Gamma)$ on Γ. On the other hand, it would be extremely useful to relate the function space structure back to the parameter domain $\square$. Locally this is possible. Since the κ_i are smooth, it is easy to see that for $s \geq 0$

$$H^s(\Gamma_i) = \{v \in L_2(\Gamma_i) : v \circ \kappa_i \in H^s(\square)\}. \tag{9.7}$$

Moreover,

$$(u, v) := \sum_{i=1}^{M} (u, v)_i, \tag{9.8}$$

where

$$(u,v)_i = \int_{\square} (u \circ \kappa_i)(x)\overline{(v \circ \kappa_i)(x)}\,\mathrm{d}x, \tag{9.9}$$

defines an inner product on Γ and

$$(v,v) \sim \langle v,v\rangle_\Gamma, \quad v \in L_2(\Gamma). \tag{9.10}$$

Equations (9.7) and (9.8) suggest the norms

$$\|v\|_s^2 := \sum_{i=1}^{M} \|v\|^2_{H^s(\Gamma_i)}, \quad 0 \le s, \tag{9.11}$$

for the space $\prod_{i=1}^M H^s(\Gamma_i)$.

Since the properties of the operator equation are usually specified in terms of a global topology on Γ, such as the one induced by spaces $H^s(\Gamma)$, say, it is important to know how these spaces relate to each other. While $H^s(\Gamma)$ is generally a closed *subspace* of $\prod_{i=1}^M H^s(\Gamma_i)$ with respect to the norm (9.11), one even has

$$H^s(\Gamma) \cong \prod_{i=1}^{M} H^s(\Gamma_i), \quad -\frac{1}{2} < s < \frac{1}{2}, \tag{9.12}$$

that is, both spaces agree as sets and the norms are equivalent. However, there is, of course, the restriction $s < 1/2$, which will be seen later to be an unfortunate obstruction.

9.3. Multi-wavelets

The above geometric setting suggests the following natural concept; see Alpert (1993), Alpert, Beylkin, Coifman and Rokhlin (1993), von Petersdorff and Schwab (1997*a*) and von Petersdorff et al. (1997). Let Π_d be the set of polynomials of total degree less than d on $\mathbb{R}^n$ and let $P := \{P_\nu : |\nu| = \nu_1 + \ldots + \nu_n < d\}$ be an orthonormal basis of Π_d on $\square$, which can be generated by the Gram–Schmidt process from the monomial basis. For simplicity, let us now write $\square = (0,1)^n$. A similar variant of what follows can be developed for the standard simplex (and even more generally for *invariant sets* (Micchelli and Xu 1994)) as well. Let $\square$ be divided into 2^{jn} congruent cubes

$$\square_{j,\eta} := 2^{-j}(\eta + \square), \quad \eta \in \{0,\ldots,2^j-1\}^n =: E_j,$$

and let $\tau_{j,\eta}(x) := 2^j x - \eta$ denote the affine transformation that takes $\square_{j,\eta}$ onto $\square$. Now one easily generates spaces of (discontinuous) piecewise polynomials of degree $< d$ relative to the partition of $\square$ into $\square_{j,\eta}$, $\eta \in E_j$. Transporting these spaces to the patches Γ_i then creates 'piecewise polynomials'

defined on Γ. Formally, this can be described as follows. Let

$$\Delta_j := \{k = (i, \eta, \nu) : \eta \in E_j, \nu \in \mathbb{Z}_+^n, |\nu| < d, i = 1, \dots, M\},$$

and set for $k = (i, \eta, \nu)$

$$\begin{aligned} \phi_{j,k}(x) &:= \begin{cases} 2^{jn/2}(\varphi_{j,\eta,\nu} \circ \kappa_i^{-1})(x), & x \in \bar{\Gamma}_i, \\ 0, & x \notin \Gamma_i, \end{cases} \\ \varphi_{j,\eta,\nu}(y) &= \begin{cases} (p_\nu \circ \tau_{j,\eta})(y), & y \in \square_{j,\eta}, \\ 0, & \text{else.} \end{cases} \end{aligned} \tag{9.13}$$

Obviously the spaces $S(\Phi_j)$ are nested and their union is dense in $L_2(\Gamma)$.

The construction of orthogonal complements between adjacent trial spaces works as follows. Again using Gram–Schmidt, one can construct an orthogonal basis $\{r_l : l = 1, \dots, (2^n - 1)\binom{n+d-1}{d-1}\}$ of the local space $S(P)$ in $S(\{\varphi_{1,\eta,\nu} : \eta \in E_1, |\nu| < d\})$. The complement basis Ψ_j in $S(\Phi_{j+1})$ is then obtained by (9.13) with p_ν replaced by r_l. The collection

$$\begin{aligned} \Psi &= \Phi_0 \cup \bigcup_{j=0}^{\infty} \Psi_j \\ &= \Phi_0 \cup \left\{\psi_\lambda : \lambda = (i, j, l), 1 \le i \le M, j \ge 0, 1 \le l \le (2^n - 1)\tbinom{n+d-1}{d-1}\right\}, \end{aligned}$$

is by construction orthonormal with respect to the inner product $(\cdot, \cdot)$ defined by (9.8), (9.9). Thus every $v \in L_2(\Gamma)$ has a unique expansion

$$v = (v, \Psi)\Psi, \quad \|v\|_{L_2(\Gamma)} \sim \|(v, \Psi)\|_{\ell_2}. \tag{9.14}$$

Moreover, for every $i \in \{1, \dots, M\}$ and any polynomial $p \in \Pi_d$ the *generalized moment conditions* hold

$$(p \circ \kappa_i^{-1}, \psi) = 0, \quad \psi \in \Psi \setminus \Phi_0. \tag{9.15}$$

This relation implies that for any smooth function f on Γ one has for $\lambda = (i, j, l)$, $|\lambda| = j$,

$$|\langle f, \psi_\lambda \rangle_\Gamma| \lesssim 2^{-|\lambda|\left(d + \frac{n}{2}\right)} \|f\|_{W^{\infty,d}(\operatorname{supp} \psi_\lambda)}. \tag{9.16}$$

In fact, setting $w_i(y) = |\partial \kappa_i(y)|$, $g(y) := w_i(y)(f \circ \kappa_i)(y)$, yields

$$\langle f, \psi_\lambda \rangle_\Gamma = \int_{\square} g(y)(\psi_\lambda \circ \kappa_i)(y)\, \mathrm{d}y.$$

Since, by construction, $\int_{\square} p(x)(\psi_\lambda \circ \kappa_i)(x)\, \mathrm{d}x = 0$, $p \in \Pi_d$, (9.16) follows from Taylor expansion of g around any point in $\operatorname{supp} \psi_\lambda \circ \kappa_i$, and the fact that $w_i(y)$ and κ_i are smooth.

Using Proposition 5.1, it is also standard to confirm the direct estimates (Dahmen and Schneider 1996)

$$\|v - (v, \Phi_J)\Phi_J\|_{L_2(\Gamma)} \lesssim 2^{-Jd}\|v\|_d, \tag{9.17}$$

provided that Γ is smooth enough to admit the definition of $H^d(\Gamma)$ in the above sense.

This approach can be extended to other parameter domains exhibiting a certain selfsimilarity; see Micchelli and Xu (1994). The following points should be kept in mind, however.

(i) Note that the order of moment conditions in (9.15) equals the order of accuracy in (9.17). It will be seen later that asymptotic optimality sometimes demands that the order of moment conditions is *higher* than the order of exactness.

(ii) The multi-wavelet basis is very flexible and relatively easy to implement. On the other hand, due to the discontinuous character of the trial functions, $\dim S(\Phi_j) = N2^{nj}\binom{n+d-1}{d-1}$. This effect could be damped by forming *composite wavelet* bases according to the following recipe (Dahmen and Schneider 1996).

- Construct biorthogonal wavelet bases $\Psi^\square, \tilde{\Psi}^\square$ on the parameter domain $\square$ by taking tensor products of the bases discussed in Section 4.4.
- Lift these bases with the aid of the parametric mappings κ_i as above to composite biorthogonal bases $\Psi, \tilde{\Psi}$ with respect to the inner product $(\cdot, \cdot)$ (9.8).

This alternative has the following attractive features.

- Although the same order d of exactness (9.17) is retained, one has $\dim S(\Phi_j) \leq N2^{nj}$, which is the fraction $\binom{n+d-1}{d-1}^{-1}$ of the dimension of the corresponding discontinuous space. Moreover, on each patch the trial functions are still $d-2$ times differentiable.
- The order $\tilde{d}$ of vanishing moments can be chosen as $\tilde{d} \geq d$ independently of the order d of accuracy, which will be seen to support compression.
- The Riesz basis property of the biorthogonal bases is not quite as straightforward as in the orthonormal case. However, it is still straightforward to verify the validity of direct and inverse estimates as in Section 5.1 (see Proposition 5.1), so that Theorem 5.8 applies and confirms, among other things, (9.14) in this case.
- Recall from (6.14) and Theorem 6.1 that optimal preconditioning depends on the validity of norm equivalences (6.1) in a range $(-\tilde{\gamma}, \gamma)$ containing t, where $2t = r$ is the order of $\mathcal{L}$. Thus, by (9.12), bases

of the above type are not optimal for the single-layer potential operator, which requires $-\frac{1}{2} \in (-\tilde{\gamma}, \gamma)$. Since discontinuities are confined to the patch boundaries, this adverse effect is expected to be milder than for a basis with increasingly dense discontinuities. The validity of norm equivalences of the form (6.1) for function spaces on manifolds will be seen later to be closely related to suitable characterizations of the function spaces with respect to *partitions*, not coverings, of the manifold (see Section 10.1 below).

Nevertheless, for operators of order zero, multi-wavelets are admissible. This program has been carried through by von Petersdorff and Schwab (1997*a*), arriving ultimately at a fully discrete scheme which solves the discretized boundary integral equation with matrices of size N_J at a cost of $\mathcal{O}(N_J(\log N_J)^4)$ operations and storage up to nearly asymptotically optimal accuracy. Some ingredients of the schemes in Dahmen et al. (1994*b*), Dahmen and Schneider (1997*b*) and Schneider (1995) will now be sketched, primarily from the point of view taken in Dahmen et al. (1994*b*), namely to identify the precise requirements on a pair of biorthogonal bases $\Psi^\Gamma, \tilde{\Psi}^\Gamma$ for $L_2(\Gamma)$ that gives rise to an symptotically optimal scheme. These findings, in turn, will then guide the construction of suitable bases for the general case. One can then also get rid of logarithmic factors in the work estimates.

9.4. A basic estimate

In the following we will assume that the operator

$$\mathcal{L}v = \int_\Gamma K(\cdot, x)v(x)\,\mathrm{d}s_x \tag{9.18}$$

satisfies the estimate (2.23) and that its Schwartz kernel K is smooth except on the diagonal, such that (2.25) holds (see Section 2.3). Γ is an (at least Lipschitz) manifold of dimension n. To solve the equation

$$\mathcal{L}u = f, \tag{9.19}$$

we wish to employ a pair of biorthogonal wavelet bases $\Psi = \{\psi_\lambda : \lambda \in \nabla\}$, $\tilde{\Psi} = \{\tilde{\psi}_\lambda : \lambda \in \nabla\}$ with $\nabla = \Delta_+ \cup \nabla_-$ as before. We will assume for the moment that this pair is *ideal* in the following sense.

Assumptions. For any order of accuracy d we have $\tilde{d}$th order of vanishing moments

$$(p \circ \kappa^{-1}, \psi_\lambda) = 0, \quad \lambda \in \nabla_-, \quad p \in \Pi_{\tilde{d}}, \tag{9.20}$$

where κ is a regular parametrization as above. Moreover, the pair of biorthogonal bases $\Psi, \tilde{\Psi}$ satisfy the norm equivalence (6.1) (or (5.42)) for the range

$s \in (-\tilde{\gamma}, \gamma)$. The regularity bounds $\gamma, \tilde{\gamma}$ are related to $\mathcal{L}$ by

$$|t| < \gamma, \tilde{\gamma}, \tag{9.21}$$

where again $r = 2t$ is the order of $\mathcal{L}$. Of course, we will also assume that the Galerkin method is stable, that is, (6.7) holds.

The first important step is to verify an estimate of the type (7.11) (see also (1.11)). Denote again by $|\lambda|$ the scale associated with ψ_λ and by Ω_λ the support of ψ_λ. In view of the moment condition (9.20), the argument leading to (9.16) can be applied to each variable consecutively (recall Section 1.3), which provides

$$|\langle \mathcal{L}\psi_{\lambda'}, \psi_\lambda \rangle| \lesssim \frac{2^{-(|\lambda|+|\lambda'|)(n/2+\tilde{d})}}{(\operatorname{dist}(\Omega_\lambda, \Omega_{\lambda'}))^{n+2\tilde{d}+2t}}, \tag{9.22}$$

whenever $\operatorname{dist}(\Omega_\lambda, \Omega_{\lambda'}) \gtrsim 2^{-\min(|\lambda|,|\lambda'|)}$ (Dahmen et al. 1994*b*, von Petersdorff and Schwab 1997*b*, von Petersdorff and Schwab 1997*a*). When the supports of ψ_λ and $\psi_{\lambda'}$ overlap, or more generally, when $\operatorname{dist}(\Omega_\lambda, \Omega_{\lambda'}) \lesssim 2^{-\min(|\lambda|,|\lambda'|)}$, one can use the *norm equivalence* (6.1) as follows (Dahlke et al. 1997*b*). To this end, suppose that $\mathcal{L}$ has the following additional continuity properties. There exists some $\tau > 0$ such that

$$\|Av\|_{H^{-t+s}} \lesssim \|v\|_{H^{t+s}}, \quad v \in H^{t+s}, 0 \le |s| \le \tau. \tag{9.23}$$

Without loss of generality one can assume that $|\lambda| > |\lambda'|$, that is, $\lambda \in \nabla_-, \lambda' \in \nabla$. Using Schwarz's inequality and the continuity of $\mathcal{L}$ (9.23) gives

$$|\langle \mathcal{L}\psi_{\lambda'}, \psi_\lambda \rangle| \le \|\mathcal{L}\psi_{\lambda'}\|_{H^{-t+\sigma}} \|\psi_\lambda\|_{H^{t-\sigma}} \lesssim \|\psi_{\lambda'}\|_{H^{t+\sigma}} \|\psi_\lambda\|_{H^{t-\sigma}}. \tag{9.24}$$

Thus, when

$$\sigma \le \tau, \quad t + \sigma < \gamma, \quad t - \sigma > -\tilde{\gamma},$$

one can apply now the norm equivalence (5.42) to each factor on the right-hand side of (9.24) which, upon using biorthogonality, yields

$$|\langle \mathcal{L}\psi_{\lambda'}, \psi_\lambda \rangle| \le 2^{t(|\lambda|+|\lambda'|)} 2^{\sigma(|\lambda'|-|\lambda|)}. \tag{9.25}$$

Combining (9.22) and (9.25) and assuming that

$$n/2 + \tilde{d} + t \ge \sigma, \tag{9.26}$$

one arrives at the following central estimate

$$2^{-(|\lambda'|+|\lambda|)t} |\langle \mathcal{L}\psi_{\lambda'}, \psi_\lambda \rangle| \lesssim \frac{2^{-||\lambda|-|\lambda'||\sigma}}{(1 + 2^{\min(|\lambda|,|\lambda'|)} \operatorname{dist}(\Omega_\lambda, \Omega_{\lambda'}))^{n+2\tilde{d}+2t}}. \tag{9.27}$$

This is exactly of the form (7.11). Note that the *preconditioning* has already been incorporated so that, in agreement with (7.11), the quantities on the

left-hand side now represent a zero-order operator. Note also that the number of vanishing moments $\tilde{d}$ determines the decay on fixed levels. It was important above in (9.26) to be able to choose $\tilde{d}$ large enough.

As earlier, let $\Psi^J := \Phi_0 \cup_{j=0}^J \Psi_j$. The idea is to replace by zero those entries in the stiffness matrices

$$\mathbf{A}_{\Psi^J} := \langle \mathcal{L}\Psi^J, \Psi^J \rangle_\Gamma^T$$

which, according to the *a priori* estimates (9.27), are guaranteed to stay below a given threshold. However, that would leave the order of $J2^J$ entries for which no decay is predicted by (9.27) (Dahmen et al. 1994*b*, Dahmen and Schneider 1997*b*, von Petersdorff and Schwab 1997*b*, von Petersdorff and Schwab 1997*a*). A further reduction requires more subtle estimates developed by Schneider (1995). To this end, we will assume that the wavelets are, up to parametric transformation, piecewise polynomials, and we will denote the *singular support* of $\psi_{\lambda'}$ (Dahmen, Kunoth and Schneider 1997, Schneider 1995) by

$$\Omega_{\lambda'}^S := \text{sing supp}\, \psi_{\lambda'},$$

which consists of the boundaries of the subdomains in $\Omega_{\lambda'}$ whose parametric preimages in $\square$ are maximal regions where $\psi_{\lambda'} \circ \kappa_i$ is a polynomial (in this case of order d). If $|\lambda'| < |\lambda|$ and $\text{dist}\,(\Omega_\lambda, \Omega_{\lambda'}) \lesssim 2^{-|\lambda'|}$, then it is shown in Schneider (1995) that the estimate

$$|\langle \mathcal{L}\psi_{\lambda'}, \psi_\lambda \rangle_\Gamma| \lesssim \frac{2^{-|\lambda|(1+\tilde{d})} 2^{|\lambda'|}}{(\text{dist}\,(\Omega_\lambda, \Omega_{\lambda'}^S))^{2t+\tilde{d}}} \tag{9.28}$$

holds.

9.5. Matrix compression

With the above estimates at hand, a *level dependent a priori truncation rule* can be designed in such a way that, on zeroing all entries which stay below the corresponding threshold, the resulting compressed matrix $\mathbf{A}_{\Psi^J}^c$ is sparse and contains only $\mathcal{O}(N_J)$ nonvanishing entries. As earlier, $N_J := \dim S(\Phi_J)$ is the dimension of the trial space of highest resolution. In addition to the above constraint (9.26) on $\tilde{d}$ it is important here to have

$$d < \tilde{d} + 2t. \tag{9.29}$$

Thus for operators of nonpositive order the order of vanishing moments should *exceed* the order of accuracy of the underlying scheme.

The compression proceeds in two steps. Fixing some $a > 0$ and $d' \in (d, \tilde{d} + 2t)$, let for $j = |\lambda|, j' = |\lambda'|$

$$b_{j,j'} \sim \max\{a\, 2^{-j}, a\, 2^{-j'}, a 2^{(J(2d'-2t)-j'(\tilde{d}+d')-j(\tilde{d}+d'))/(2\tilde{d}+2t)}\}, \tag{9.30}$$

and set

$$a^1_{\lambda,\lambda'} := \begin{cases} (\mathbf{A}_{\Psi^J})_{\lambda,\lambda'}, & \text{if } \operatorname{dist}(\Omega_\lambda, \Omega_{\lambda'}) \le b_{j,j'}, \\ 0, & \text{otherwise.} \end{cases} \tag{9.31}$$

Hence the bands get narrower when progressing to higher scales. In a second step, one sets

$$(\mathbf{A}^c_{\Psi^J})_{\lambda,\lambda'} := \begin{cases} a^1_{\lambda,\lambda'}, & j' \le j \text{ and } \operatorname{dist}(\Omega_\lambda, \Omega^S_{\lambda'}) \le b^S_{j,j'}, \\ & j \le j' \text{ and } \operatorname{dist}(\Omega^S_\lambda, \Omega_{\lambda'}) \le b^S_{j,j'}, \\ 0, & \text{otherwise.} \end{cases} \tag{9.32}$$

Here the truncation parameters $b^S_{j,j'}$ controlling the distance from the singular support are given by

$$b^S_{j,j'} \sim \max\{a' 2^{-j}, a' 2^{-j'}, a' 2^{(J(2d'-2t)-\max\{j,j'\}\tilde{d}-(j+j')d')/(\tilde{d}+2t)}\}, \tag{9.33}$$

and the parameters a, a' are fixed constants independent of J. For instance, a determines the bandwidth in the block matrices $\mathbf{A}^c_{J,J} = (\mathbf{A}_{\psi^J})_{J,J} = (\langle \mathcal{L}\psi_{\lambda'}, \psi_\lambda\rangle)_{|\lambda'|,|\lambda|=J}$. The choice of a, a' will be further specified later (Dahmen et al. 1994*b*, Schneider 1995).

Theorem 9.1 If the moment conditions (9.20) hold for $\tilde{d}$ satisfying (9.29), then under the above assumptions on $\mathcal{L}$ and $\Psi, \tilde{\Psi}$ the compression strategy (9.31), (9.32) generates matrices $\mathbf{A}^c_{\Psi^J}$ containing only $\mathcal{O}(N_J)$ non-vanishing entries.

9.6. Asymptotic estimates

The basic tool for estimating the effect of the above compression is a suitable version of a *weighted Schur lemma.* Recall that if for some matrix $\mathbf{A} = (a_{i,j})_{i,j\in I}$ there exists a positive constant c and a sequence $\mathbf{b}$ with $b_i > 0$, such that

$$\sum_{i\in I} |a_{i,j}| b_i \le c\, b_j \quad \text{for all} \quad j \in I,$$

and

$$\sum_{j\in I} |a_{i,j}| b_j \le c\, b_i \quad \text{for all} \quad j \in I,$$

then $\|\mathbf{A}\| \le c$, where $\|\cdot\|$ denotes the spectral norm. In the present context the b_j are chosen as 2^{-sj} for suitable choices of $s \ge 0$. Again denoting by $\mathbf{D}^s$ the diagonal matrix with entries $(\mathbf{D}^s)_{\lambda,\lambda'} = 2^{s|\lambda|}\delta_{\lambda,\lambda'}$, the Schur lemma can be used to show that

$$\|\mathbf{D}_J^{-s}(\mathbf{A}_{\Psi^J} - \mathbf{A}^c_{\Psi^J})\mathbf{D}_J^{-\tilde{s}}\| \lesssim J^{-1} a^{-2t-2\tilde{d}} 2^{-J(s+\tilde{s}-2t)}.$$

At this point the *norm equivalences* enter again. In fact, one infers from the above estimate combined with (6.1) *consistency estimates* of the form

$$\|(\mathcal{L}_J - \mathcal{L}_J^c)\, u\|_{H^{s-2t}} \lesssim a^{-2t-2\tilde{d}} 2^{J(s-\tau)} \|u\|_{H^\tau}, \tag{9.34}$$

where $a > 1$ is fixed, $\mathcal{L}_J$, $\mathcal{L}_J^c$ are the finite-dimensional operators corresponding to $\mathbf{A}_{\Psi^J}$ and $\mathbf{A}^c_{\Psi^J}$, respectively, and the range of the parameters s and τ is $-d + 2t \le s < \gamma$, $-\gamma < \tau \le d$. As before, γ and d reflect the regularity and the order of accuracy of the trial functions. In particular, for any $\epsilon > 0$ one can choose $a > 1$ such that

$$\|(\mathcal{L}_J - \mathcal{L}_J^c)\, u\|_{H^{-t}} \lesssim \epsilon \|u\|_{H^t}. \tag{9.35}$$

A perturbation argument combined with these estimates ensures stability of the compressed operator in the energy norm and even for lower norms, we have

$$\|\mathcal{L}_J^c v_J\|_{H^{s-2t}} \gtrsim \|v_J\|_{H^s}, \quad v_J \in S(\Phi_J), \tag{9.36}$$

for $2t - d \le s \le t$; see, for example, Dahmen et al. (1994*b*).

These facts can then be combined to prove the following result (Dahmen et al. 1994*b*, Dahmen et al. 1993*b*, Dahmen and Schneider 1997*b*, Schneider 1995).

Theorem 9.2 Under the above circumstances the compressed system

$$\mathbf{A}^c_{\Psi^J} \mathbf{d}_J = \langle f, \Psi^J \rangle_\Gamma^{\mathrm{T}}$$

possesses a unique solution and $u_J^c := \mathbf{d}_J^T \Psi^J$ has asymptotically optimal accuracy

$$\|u - u_J^c\|_{H^\tau} \lesssim 2^{J(\tau-s)} \|u\|_{H^s}, \tag{9.37}$$

where $-d + 2t \le \tau < \gamma$, $\tau \le s$, $t \le s \le d$ and u is the exact solution of $\mathcal{L}u = f$. Moreover, the matrices $\mathbf{B}_J^c = \mathbf{D}^{-t} \mathbf{A}^c_{\Psi^J} \mathbf{D}^{-t}$ have the order of N_J nonvanishing entries and uniformly bounded condition numbers.

By Remark 6.3, one obtains a scheme that solves (9.19) with asymptotically optimal accuracy in linear time.

We summarize the required conditions on the wavelet basis. To realize an asymptotically optimal balance between accuracy and efficiency, the regularity γ of Ψ, the regularity $\tilde{\gamma}$ of the dual basis $\tilde{\Psi}$, the order of vanishing moments $\tilde{d}$ and the order of exactness d of the trial spaces $S(\Phi_J)$ should be related in the following way.

Regularity	$\gamma > t$ conformity	$\tilde{\gamma} > -t$ preconditioning
Order	d	convergence rate $2^{-J(2d+2-2t)}$
Vanishing moments	$\tilde{d} > d - 2t$	

9.7. Adaptive quadrature

In the above analysis it has been assumed that the matrix entries $\langle \mathcal{L}\psi_\lambda, \psi_{\lambda'}\rangle_\Gamma$ are given exactly. Of course, in general they have no closed analytical representation.

In principle, one can first compute the stiffness matrix $\langle \mathcal{L}\Phi_J, \Phi_J\rangle_\Gamma^{\mathrm{T}}$ relative to the single-scale basis Φ_J (for instance with the aid of the techniques described in Section 4.2) with sufficient accuracy to preserve the overall precision of the above scheme. In fact, the *multiscale transformation* $\mathbf{T}_J$ from (3.25) yields

$$\mathbf{A}_{\Psi^J} = \mathbf{T}_J^T \langle \mathcal{L}\Phi_J, \Phi_J\rangle_\Gamma^{\mathrm{T}} \mathbf{T}_J.$$

However, since $\langle \mathcal{L}\Phi_J, \Phi_J\rangle_\Gamma$ is a dense matrix, this process requires at least the order of N_J^2 operations and storage which would completely destroy the efficiency of the fully discrete scheme.

To find a more economic strategy, one has to bear the following points in mind.

- There is an *a priori* criterion to decide whether a matrix coefficient must be computed or can be neglected.
- Note that dist $(\Omega_\lambda, \Omega_{\lambda'}) > b_{|\lambda|,|\lambda'|}$ implies that dist $(\Omega_\nu, \Omega_{\nu'}) > b_{|\nu|,|\nu'|}$ holds for $\Omega_\nu \subset \Omega_\lambda$ and $\Omega_{\nu'} \subset \Omega_{\lambda'}$, $|\nu| \geq |\lambda|, |\nu'| \geq |\lambda'|$. Thus, one does not have to check condition (9.31) or (9.32) for all pairs λ, λ'. Exploiting the hierarchical structure of multiscale bases, one needs at most $\mathcal{O}(2^{nJ}) = \mathcal{O}(N_J)$ checks to decide whether or not an entry has to be computed.

An accurate computation of the remaining nonzero coefficients by numerical quadrature is a difficult task. Significant coefficients involving low-level wavelets have to be computed with accuracy determined by the discretization error of the scheme. We will see later that, based on the construction outlined in Section 4.4, wavelets with the above ideal properties can be constructed whose pullbacks to the parameter domain are piecewise polynomials. Hence the approximation of $\langle \mathcal{L}\psi_\lambda, \psi_{\lambda'}\rangle_\Gamma$ can be reduced to the evaluation of integrals of the form

$$\int_{\kappa_i(\square_\lambda)} \int_{\kappa_l(\square_{\lambda'})} K(\hat{x}, \hat{y}) \psi_\lambda(\hat{x}) \psi_{\lambda'}(\hat{y}) \, \mathrm{d}s_{\hat{x}} \, \mathrm{d}s_{\hat{y}}, \tag{9.38}$$

where $\square_\lambda \subset \square$ denotes a cube such that $\psi_\lambda \circ \kappa_i \,|_{\square_\lambda}$ is a polynomial of degree $d-1$. Thus one ultimately has to compute expressions of the type

$$\int_{\square_\lambda} \int_{\square_{\lambda'}} H(x, y) p_\lambda(x) p_{\lambda'}(y) \, \mathrm{d}x \, \mathrm{d}y, \tag{9.39}$$

where $p_\lambda, p_{\lambda'}$ are polynomials of degree $d-1$ satisfying

$$\|p_\lambda\|_{W^{s,\infty}(\square_\lambda)} \lesssim 2^{(s+n/2)j}. \tag{9.40}$$

When $\kappa_i(\square_\lambda) \cap \kappa_l(\square_{\lambda'}) = \emptyset$, $H(x,y) = K(\kappa_i(x), \kappa_l(y))|\partial\kappa_i(x)||\partial\kappa_l(y)|$ is arbitrarily smooth. Thus high-order quadrature can be used to compute entries not discarded by the decay estimates. When integrating over pairs of domains that share an edge, a vertex or are identical, then in general the integral is singular. In this case some sort of *regularization* should be applied to reduce the integral to a weakly singular integral (Nedelec 1982, von Petersdorff and Schwab 1997*a*). Then one can use transformation techniques like Duffy's trick proposed by Sauter (1992) to end up with analytical integrals (von Petersdorff and Schwab 1997*a*, Schwab 1994).

The central objective is now to balance the error caused by quadrature with the desired overall accuracy of the scheme, while preserving efficiency. Employing adaptive quadrature in connection with a multi-wavelet discretization for zero-order operators, a fully discrete scheme has recently been developed in von Petersdorff and Schwab (1997*a*), where essential use is also made of the analyticity of the kernel K in a neighbourhood of the two-dimensional surface Γ in $\mathbb{R}^3$. The resulting fully discrete scheme requires $\mathcal{O}(N_J(\log N_J)^4)$ operations. A somewhat different approach is given by Dahmen and Schneider (1997*b*), ending up with a slightly more favourable complexity analysis.

The balancing of errors is guided by the following considerations. The problem of quadrature has to be seen in close connection with compression and the special features of multiscale bases. Basis functions from coarser scales introduce large domains of integration while requiring high accuracy. In particular, on the coarsest scale $\lambda, \lambda' \in \Delta_+$ the full accuracy $2^{-J(2d'-2t)}$ depending on J is required, while on the highest scale $|\lambda|, |\lambda'| = J$ the computation of each entry requires only a fixed number of quadrature points independent of J. In fact, $\operatorname{diam}\operatorname{supp}\psi_\lambda \sim 2^{-J}$ and $|\langle \mathcal{L}\psi_\lambda, \psi_{\lambda'}\rangle_\Gamma| \lesssim 2^{j2t}$ for $|\lambda| = |\lambda'| = j$. Thus, many entries only have to be computed with low accuracy, while high accuracy is merely required for a small portion of the matrix. Using the analysis of matrix compression as a guideline, a careful balancing of the various effects shows that most matrix entries $\langle \mathcal{L}\psi_\lambda, \psi_{\lambda'}\rangle_\Gamma$ must be computed with a precision

$$e_{\lambda,\lambda'} \lesssim 2^{-J(2d'-2t)} 2^{\max\{|\lambda|,\lambda'|\}(d'+1)} 2^{\min\{|\lambda|,|\lambda'|\}(d'+1)} 2^{-2\max\{|\lambda|,|\lambda'|\}}$$

for some $d' > d$ (Dahmen et al. 1997, Dahmen and Schneider 1997*b*).

The fully discretized Galerkin method in Dahmen and Schneider (1997*b*) is based on product-type Gaussian formulae of order D for approximating

inner and outer integrals

$$\int_{\tau}\int_{\tau'} p(x)p'(y)\,\mathrm{d}x\,\mathrm{d}y = Q_x^D \otimes Q_y^D\,(p\cdot p'), \quad \text{for all} \quad p,p' \in \Pi_D, \tag{9.41}$$

where the domains τ and τ' are congruent to $\square$. According to the previous remarks, the error estimate for the quadrature method has much in common with estimating matrix coefficients relative to wavelet bases. The relevant estimates are summarized as follows.

Lemma 9.3 Let $Q_\tau^D \otimes Q_{\tau'}^D$ be a product-type Gaussian quadrature method of order D and $\tau \subset \square_\lambda$, $\tau' \subset \square_{\lambda'}$. Furthermore, suppose that $\mathcal{L}$ is a boundary integral operator with the above properties and Γ is a piecewise analytic boundary surface. In local parametrization let the kernel be denoted by $H(x,y)$ as above and set $G(x,y) := H(x,y)p_\lambda(x)p_{\lambda'}(y)$. If $\tau \cap \tau' = \emptyset$, then there exists a constant c such that the estimate

$$\begin{aligned}&\Big|\int_{\tau}\int_{\tau'} G(x,y)\,\mathrm{d}x\,\mathrm{d}y - Q_\tau^D \otimes Q_{\tau'}^D(G)\Big| \\ &\quad\le\ c\frac{2^{(|\lambda|+|\lambda'|)}(\max\{\operatorname{diam}\tau,\operatorname{diam}\tau'\})^{D-d}(\operatorname{diam}\tau)^2(\operatorname{diam}\tau')^2}{\operatorname{dist}\,(\kappa_i(\tau'),\kappa_l(\tau))^{2+2t+D-d}}\end{aligned}$$

holds, provided that $2+2t+D-d>0$.

The principal strategy is to choose the diameter of the subdomains proportional to the distance from the singularity while the degree D has to be adapted to maintain the desired accuracy taking the decay of the entries into account. Details can be found in Dahmen et al. (1997) and Dahmen and Schneider (1997*b*):

In summary the following result can be proved (Dahmen et al. 1997, Dahmen and Schneider 1997*b*).

Theorem 9.4 Under the above assumptions the fully discretized compressed system $\mathbf{A}_{\Psi^J}^{cq}\mathbf{d}_J^{cq} = \langle f, \Psi^J\rangle^{\mathrm{T}}$ possesses a unique solution and $u_J^{cq} := (\mathbf{d}_J^{cq})^T\Psi^J$ realizes asymptotically optimal accuracy

$$\|u - u_J^{cq}\|_{H^\tau} \ \lesssim\ 2^{j(\tau-s)}\,\|u\|_{H^s} \tag{9.42}$$

where $-d+2t \le \tau < \gamma$, $\tau \le s$, $t \le s \le d$ and u is the exact solution of $\mathcal{L}u = f$. Moreover, the nonzero coefficients of the matrix $\mathbf{A}_{\Psi^J}^{cq}$ can be computed at the expense of $\mathcal{O}(N_J)$ floating point operations and storage.

10. Wavelets on manifolds and domain decomposition

The periodic case is certainly the most convenient setting for constructing wavelets and exploiting their full computational efficiency. On the other hand, the application of embedding techniques as described in Section 8

is certainly limited. For instance, problems defined on closed surfaces, as discussed in Section 9, cannot be treated in this way. Also, the resolution of boundary layer effects may cause difficulties.

This has motivated various attempts to extend wavelet-like tools to more general settings. This is, for instance, reflected by the general framework in Section 3, and concepts like stable completions; see Section 3.2 and Carnicer et al. (1996). The so-called *lifting scheme* (Sweldens 1996, Sweldens 1997) is very similar in spirit. Its applications, for instance in computer graphics, also demonstrate its versatility and efficiency in connection with unstructured grids (Schröder and Sweldens 1995). Unfortunately, the understanding of analytical properties like stability and norm equivalences in a more general setting still appears to be in its infancy. Attempts to develop stability criteria that work in sufficiently flexible settings have been sketched in Section 5; see Dahmen (1994) and (1996). Some recent consequences of these developments will be indicated next.

Many problem formulations suggest in a natural way a decomposition of the underlying domain into subdomains, which in turn are often representable as parametric images of *cubes*. As was indicated in Section 4.4, wavelet bases on cubes are well understood and much of the efficiency of wavelet bases in the ideal setting can be retained. This can readily be combined with the idea described in Section 9.2 to obtain wavelet bases with essentially the same nice properties on any domain Ω, as long as $\Omega = \kappa(\square)$ is a smooth regular parametric image of the unit cube, that is,

$$|\partial\kappa(x)| \neq 0, \quad x \in \square. \tag{10.1}$$

In fact, the canonical inner product $\langle \cdot, \cdot \rangle_\Omega$ can be replaced by the inner product (see (9.9))

$$(u, v) := \int_{\square} (u \circ \kappa)(x)\overline{(v \circ \kappa)(x)}\, \mathrm{d}x, \tag{10.2}$$

which induces an equivalent norm for $L_2(\Omega)$, say. Moreover, when $F \subseteq L_2(\Omega)$ denotes a Besov or Sobolev space on Ω, it can be pulled back to a corresponding space on $\square$ by

$$F(\Omega) = \{g \circ \kappa^{-1} : g \in F(\square)\}, \tag{10.3}$$

with

$$\|v\|_{F(\Omega)} \sim \|v \circ \kappa\|_{F(\square)}. \tag{10.4}$$

Any biorthogonal wavelet bases $\Psi, \tilde{\Psi}$ on $\square$ then induce collections $\Psi^\Omega := \Psi \circ \kappa^{-1}$, $\tilde{\Psi}^\Omega := \tilde{\Psi} \circ \kappa^{-1}$ which are biorthogonal Riesz bases on Ω relative to the inner product (10.2). On account of (10.4), they inherit all the norm equivalences satisfied by $\Psi, \tilde{\Psi}$. In this way, all computations are ultimately carried out on the standard domain $\square$.

Of course, the qualitative properties of the bases $\Psi^\Omega, \tilde{\Psi}^\Omega$ depend on the mapping κ, which in practice confines this approach to rather simple domains Ω. However, the next step, which was made, to some extent, in Section 9, is to consider domains that are *disjoint unions* of such simple domains. Modelling closed surfaces, as considered in Section 9.1, falls exactly into this category. Although the following facts are by no means restricted to this case, we will adopt the same notation and assumptions made in Section 9.1 but keep in mind that Γ may as well denote some bounded domain in Euclidean space.

The whole preceding development shows that the power of wavelet discretizations hinges on its relation to certain function spaces, in particular, on corresponding norm equivalences. However, this is exactly the point where one easily gets stuck. In fact, recall from (9.12) that managing norm equivalences on the individual spaces $F(\Gamma_i)$ with the aid of the transported bases $\Psi^{\Gamma_i}, \tilde{\Psi}^{\Gamma_i}$ does not generally imply corresponding relations with respect to the global space $F(\Gamma)$. The problem is that the norms $\|\cdot\|_{F(\Gamma)}$ and $(\sum_{i=1}^M \|\cdot\|^2_{F(\Gamma_i)})^{1/2}$ do *not* generally determine the same space. Below we indicate several attempts, mostly referring to work in progress, to overcome this difficulty.

10.1. Composite wavelet bases

The following comments are based on Dahmen and Schneider (1996), and related special cases considered in Jouini and Lemarié-Rieusset (1993). The basic idea is to glue the bases defined on each patch together so that the resulting global bases are at least continuous on all of Γ. One way to achieve this is to carefully inspect the construction of biorthogonal spline wavelets on $[0,1]$, described in Section 4.4. One can show that the biorthogonal generator bases Φ_j and $\tilde{\Phi}_j$ on $[0,1]$ can be arranged to have the following property. All but one basis function at each end of the interval vanish at 0 and 1. This fact can then be exploited to construct pairs of refinable biorthogonal generator bases $\Phi_j^\Gamma, \tilde{\Phi}_j^\Gamma$, which belong to $C(\Gamma)$. Unfortunately, the wavelets corresponding to these global generator bases *cannot* be easily obtained by stitching local wavelet bases together. The reason is that not all the wavelets for the local bases can be arranged to vanish at the patch boundaries. Nevertheless, one can employ the concept of stable completions from Section 3.2 to construct compactly supported biorthogonal wavelet bases $\Psi^\Gamma, \tilde{\Psi}^\Gamma$ on Γ, which also belong to $C(\Gamma)$ (Dahmen and Schneider 1996). The disadvantage of this construction is that, since some wavelets have support in more than one patch Γ_i, moment conditions of the form (9.20) no longer hold in full strength near the patch boundaries.

Nevertheless, since all basis functions are local and since the trial spaces $S(\Phi_j^\Gamma)$ retain the same approximation properties as the local spaces transpor-

ted from $\square$, these spaces can be used for conforming Galerkin discretizations for second-order problems, even in connection with *nonoverlapping domain decomposition strategies*, for instance. On the other hand, it is clear that such an approach is limited for principal reasons. For example, it does not provide the *ideal* bases in the sense of Section 9.6 for operators of order minus one. More generally, this approach is not suited for handling duality.

An alternative approach, which is interesting from several points of view, will be outlined next.

10.2. Characterization of function spaces via partitions of domains

In the following we will denote by $F(\Gamma')$ spaces of the form $H^s(\Gamma')$ or Besov spaces, where $\Gamma' \subseteq \Gamma$. The problem with (9.12) is that the spaces $F(\Gamma)$ are usually defined through *overlapping coverings* of Γ, not through a *partition* of Γ. Therefore a fundamental step towards overcoming limitations of the type (9.12) is first to derive a characterization of function spaces on Γ in terms of *partitions*. Such characterizations were developed by Ciesielski and Figiel (1983), in terms of mappings

$$T : F(\Gamma) \to \prod_{i=1}^{M} \chi_{\Gamma_i}(P_i(F(\Gamma))), \quad V : F(\Gamma) \to \prod_{i=1}^{M} \chi_{\Gamma_i}(P_i^*(F(\Gamma))), \tag{10.5}$$

defined by

$$Tv = (\chi_{\Gamma_i} P_i v)_{i=1}^{M}, \quad Vv = (\chi_{\Gamma_i} P_i^* v)_{i=1}^{M}. \tag{10.6}$$

Here χ_{Γ_i} denotes the characteristic function of Γ_i and the P_i are certain projectors on $L_2(\Gamma)$, constructed in such way that T and V are actually *topological isomorphisms* with respect to F, and the factors $\chi_{\Gamma_i}(P_i(F(\Gamma)))$ are closed subspaces of $F(\Gamma_i)$ determined by certain homogeneous *trace conditions*.

The main focus of Ciesielski and Figiel (1983) was the existence of unconditional bases of Sobolev and Besov spaces on compact C^∞-manifolds. The objective of Dahmen and Schneider (1997*a*) is to employ such concepts for the development of practicable schemes. This requires us to identify practically realizable projections P_i needed in (10.5) and to combine this with the recently developed technology of biorthogonal wavelet bases on $\square$. This provides practically feasible wavelet bases for the component spaces $\chi_{\Gamma_i}(P_i(F(\Gamma)))$, and hence through (10.5) also for Γ. The resulting bases can be shown to exhibit all the desired properties listed in Section 9.6. The main ingredients of this program can be outlined as follows.

Ordering of patches

First one *orders* the patches Γ_i in a certain fashion. If $\overline{\Gamma}_i \cap \overline{\Gamma}_l := \epsilon_{i,l}$ is a common face and $i < l$, then $\epsilon_{i,l}$ is called an *outflow* (*inflow*) face for Γ_i (Γ_l).

$\partial\Gamma_i^{\uparrow}$, $\partial\Gamma_i^{\downarrow}$ are called the *outflow* and *inflow* boundary of the patch Γ_i. Let $\Gamma_i^{\uparrow}$ denote an extension of Γ_i in Γ which contains the outflow boundary $\partial\Gamma_i^{\uparrow}$ in its relative interior and whose boundary contains the inflow boundary $\partial\Gamma_i^{\downarrow}$ of Γ_i. Thus $\Gamma_i^{\uparrow}$ could be taken as the union of Γ_i and those patches whose closure intersects the relative interior of the outflow boundary $\partial\Gamma_i^{\uparrow}$. Analogously one defines $\Gamma_i^{\downarrow}$ with respect to the reverse flow.

Extensions
Now suppose that E_i is an extension operator from the domain Γ_i to $\Gamma_i^{\uparrow}$. It turns out that the topological properties of the projectors P_i to be constructed for (10.5) hinge upon the following continuity properties of the extensions E_i. To describe this, the following notation is convenient. Let

$$f^{\uparrow}(x) := \begin{cases} f(x), & x \in \Gamma_i, \\ 0, & x \in \Gamma_i^{\uparrow} \setminus \Gamma_i \end{cases}$$

denote the trivial extension of $f \in F(\Gamma_i)$ to $\Gamma_i^{\uparrow}$ and define

$$F(\Gamma_i)^{\uparrow} := \{f \in F(\Gamma_i) : f^{\uparrow} \in F(\Gamma_i^{\uparrow})\}, \quad \|f\|_{F(\Gamma_i)^{\uparrow}} := \|f^{\uparrow}\|_{F(\Gamma_i^{\uparrow})}.$$

Thus $F(\Gamma_i)^{\uparrow}$ consists of those elements in the local space $F(\Gamma_i)$ whose trace vanishes on the outflow boundary $\partial\Gamma_i^{\uparrow}$. Again the spaces $F(\Gamma_i)^{\downarrow}$ are defined analogously.

Now suppose that the extensions E_i satisfy

$$\|E_i f\|_{F(\Gamma_i^{\uparrow})} \lesssim \|f\|_{F(\Gamma_i)}, \quad \|(E_i^* f)^{\uparrow}\|_{F(\Gamma_i)^{\uparrow}} \lesssim \|f\|_{F(\Gamma_i^{\uparrow})}. \tag{10.7}$$

Due to the simple form of the parameter domain $\square$, such extensions can be constructed explicitly as tensor products of *Hestenes-type extensions* (Ciesielski and Figiel 1983, Dahmen and Schneider 1997*a*). However, some deviations from the construction in Ciesielski and Figiel (1983), which are essential from a practical point of view, will be mentioned later.

Topological isomorphisms
Given E_i as above, one now defines

$$P_1 f := E_1(\chi_{\Gamma_1} f), \quad P_i f := E_i(\chi_{\Gamma_i}(f - \sum_{l<i} P_l f)), \quad i = 2, \ldots, M. \tag{10.8}$$

One can prove the following facts (Dahmen and Schneider 1997*a*).

Theorem 10.1 One has

$$\chi_{\Gamma_i}(P_i(F(\Gamma))) = F(\Gamma_i)^{\downarrow}, \quad \chi_{\Gamma_i}(P_i^*(F(\Gamma))) = F(\Gamma_i)^{\uparrow}. \tag{10.9}$$

The mappings

$$T : f \mapsto \{\chi_{\Gamma_i} P_i f\}_{i=1}^M, \quad V : f \mapsto \{\chi_{\Gamma_i} P_i^* f\}_{i=1}^M \tag{10.10}$$

define topological isomorphisms acting from $F(\Gamma)$ onto the product spaces $\Pi_{i=1}^M F(\Gamma_i)^\downarrow$, $\Pi_{i=1}^M F(\Gamma_i)^\uparrow$, respectively, whose inverses are given for $\mathbf{v} = (v_i)_{i=1}^M \in \prod_{i=1}^M L_2(\Gamma_i)$ by

$$S\mathbf{v} = \sum_{i=1}^M P_i \chi_{\Gamma_i} v_i, \quad U\mathbf{v} = \sum_{i=1}^M P_i^* \chi_{\Gamma_i} v_i, \tag{10.11}$$

respectively. Specifically, one has

$$F(\Gamma) \;\tilde{=}\; \prod_{i=1}^M F(\Gamma_i)^\downarrow \;\tilde{=}\; \prod_{i=1}^M F(\Gamma_i)^\uparrow,$$

and

$$\|v\|_{F(\Gamma)} \sim \left(\sum_{i=1}^M \|P_i v\|^2_{F(\Gamma_i)^\downarrow}\right)^{\frac{1}{2}} \sim \left(\sum_{i=1}^M \|P_i^* v\|^2_{F(\Gamma_i)^\uparrow}\right)^{\frac{1}{2}}, \quad v \in F(\Gamma). \tag{10.12}$$

Moreover, the maps T, V extend to isomorphisms from $F^*(\Gamma)$ onto the spaces $\prod_{i=1}^M F^*(\Gamma_i)^\downarrow$ and $\prod_{i=1}^M F^*(\Gamma_i)^\uparrow$, respectively, and

$$\|v\|_{F^*(\Gamma)} \sim \left(\sum_{j=1}^M \|P_j v\|^2_{F^*(\Gamma_j)^\downarrow}\right)^{\frac{1}{2}}, \quad v \in F^*(\Gamma). \tag{10.13}$$

Note that duality is incorporated in a natural way.

10.3. Biorthogonal wavelets on Γ

With Theorem 10.1 at hand, one can now construct wavelet bases on Γ that give rise to the desired norm equivalences. The basic steps can be roughly sketched as follows.

First, for each i let $\hat{\kappa}_i$ be an extension of κ_i (with as much smoothness as permitted by the regularity of $\Gamma_i^\downarrow$) and $\square_i^\downarrow$ a hyperrectangle such that $\hat{\kappa}_i(\square_i^\downarrow) = \Gamma_i^\downarrow$ and $\hat{\kappa}_i \mid_\square = \kappa_i$. As above, the spaces $F(\square)^{\downarrow,i}$ then consist of those elements in $F(\square)$ whose trivial extension to $\square_i^\downarrow$ by zero belongs to $F(\square_i^\downarrow)$.

- For each pair of *complementary* homogeneous boundary conditions in $F([0,1])$, construct biorthogonal wavelet bases on $[0,1]$ based on the schemes described in Section 4.4. By this we mean, for instance, that when the wavelets and generators on the primal side are to vanish up to some order at zero, there are no boundary constraints at zero for the functions in the dual system (and analogously for all possible combinations).

- Using tensor products, this leads to biorthogonal wavelet bases

$$\Psi^{\square,i} \subset F(\square)^{\downarrow,i}, \quad \tilde{\Psi}^{\square,i} \subset F(\square)^{\uparrow,i}.$$

- The bases

$$\Psi^{\Gamma_i} := \Psi^{\square,i} \circ \kappa_i^{-1} \subset F(\Gamma_i)^{\downarrow}, \quad \tilde{\Psi}^{\Gamma_i} := \tilde{\Psi}^{\square,i} \circ \kappa_i^{-1} \subset F(\Gamma_i)^{\uparrow},$$

are biorthogonal with respect to the inner product $(\cdot,\cdot)_i$ defined by (9.9).

- The collections

$$\Psi^{\Gamma} := S(\{\Psi^{\Gamma_i}\}_{i=1}^M), \quad \tilde{\Psi}^{\Gamma} := U(\{\tilde{\Psi}^{\Gamma_i}\}_{i=1}^M), \tag{10.14}$$

where S, U are defined in (10.11), are biorthogonal wavelet bases on Γ relative to the inner product (9.8). Moreover, from Theorem 10.1 and (10.3), (10.4) one infers that for $F = H^s$

$$\|\mathbf{d}^T\Psi^{\Gamma}\|_{H^s(\Gamma)} \sim \|\mathbf{D}^s\mathbf{d}\|_{\ell_2}. \tag{10.15}$$

The range of $s \in \mathbb{R}$ is constrained here by the regularity bounds $\gamma, \tilde{\gamma}$ of the bases $\Psi^{\square,i}$, $\tilde{\Psi}^{\square,i}$, respectively, and by the regularity of Γ, which restricts the range of Sobolev indices. As before, $\mathbf{D}^s$ denotes the diagonal matrix with entries $(\mathbf{D}^s)_{\lambda,\lambda'} = 2^{s|\lambda|}\delta_{\lambda,\lambda'}$.

10.4. Computational aspects

In practice one would *not* compute Ψ^{Γ} explicitly. To discuss this issue, consider the inner product

$$\langle \mathbf{v}, \mathbf{u}\rangle_{\Pi} := \sum_{i=1}^{M} \langle v_i, u_i\rangle_{\Gamma_i}$$

on $\Pi_{i=1}^M L_2(\Gamma_i)$, which is of course also equivalent to $(\cdot,\cdot)$ defined by (9.8). Formally the stiffness matrix relative to Ψ^{Γ} constructed above is given by

$$\langle \mathcal{L}\Psi^{\Gamma}, \Psi^{\Gamma}\rangle_{\Gamma} = \langle (S^*\mathcal{L}S)\{\Psi^{\Gamma_l}\}_l, \{\Psi^{\Gamma_i}\}_i\rangle_{\Pi},$$

where S is defined in (10.11). When $\mathcal{L}$ is an isomorphism from $F(\Gamma)$ into $F^*(\Gamma)$, Theorem 10.1 assures that $\mathcal{L}_{\Pi} := S^*\mathcal{L}S$ is an isomorphism from $\Pi_{\downarrow} := \prod_{i=1}^M F(\Gamma_i)^{\downarrow}$ into $\Pi_{\uparrow}^* := \prod_{i=1}^M F^*(\Gamma_i)^{\uparrow}$, that is

$$\|\mathcal{L}_{\Pi}\mathbf{v}\|_{\Pi_{\uparrow}^*} \sim \|\mathbf{v}\|_{\Pi_{\downarrow}}, \quad \mathbf{v} \in \Pi_{\downarrow}. \tag{10.16}$$

Thus the problem $\mathcal{L}u = f$ is equivalent to

$$\mathcal{L}_{\Pi}\mathbf{u} = \mathbf{f}, \tag{10.17}$$

where $\mathbf{f} = S^*f = Vf$. Of course, when $\mathbf{u}$ solves (10.17), then $u = S\mathbf{u}$ is the solution to the original problem. Straightforward calculation shows that (10.17), in turn, can be stated as

$$\sum_{l=1}^{M} \mathcal{L}_{i,l} u_l = f_i, , \quad i = 1, \dots, M,$$

where

$$\mathcal{L}_{i,l} = \chi_{\Gamma_i} P_i^* \mathcal{L} P_l \chi_{\Gamma_l}, \quad f_i = \chi_{\Gamma_i} P_i^* f, \quad i, l = 1, \dots M. \tag{10.18}$$

If, in addition, $\mathcal{L}$ is selfadjoint, one infers from (10.16) that

$$\|\mathbf{v}\|_{\Pi_\downarrow}^2 \sim \langle \mathcal{L}_\Pi \mathbf{v}, \mathbf{v} \rangle_\Pi.$$

Thus, choosing $\mathbf{v} := \{v\delta_{i,l}\}_{l=1}^M$, this yields

$$\|\mathcal{L}_{i,i} v\|_{F^*(\Gamma_i)^\uparrow} \sim \|v\|_{F(\Gamma_i)^\downarrow}, \quad i = 1, \dots, M, \tag{10.19}$$

which suggests solving (10.17) by an iteration of the form

$$u_i^{j+1} = u_i^j + \omega \mathcal{L}_{i,i}^{-1} (f_i - \sum_{l=1}^{M} \mathcal{L}_{i,l} u_l^j), \quad i = 1, \dots, M. \tag{10.20}$$

In fact this fits into the framework of Schwarz-type iterations described in Section 6.5. Specifically, on account of Theorem 10.1, one can apply Theorem 6.9, where S, defined by (10.11), plays the role of the mapping $\mathcal{R}$ in Theorem 6.9, so that convergence of the iteration follows from Theorem 6.10.

Hence the solution of (10.17) has been reduced to the *parallel* solution of *local* problems of the form

$$\mathcal{L}_{i,i} u_i = g_i, \quad i = 1, \dots, M, \tag{10.21}$$

which may be viewed as a *domain decomposition* method. On account of the relation $F(\Gamma_i)^\downarrow = \{g \circ \kappa_i^{-1} : g \in F(\square)^{\downarrow,i}\}$ (*cf.* (9.7)) and the definition of the bases Ψ^{Γ_i}, each equation in (10.21) is in effect an elliptic problem defined on the *unit cube*. On the unit cube $\square$, wavelet bases with all the desired properties are available. In addition, full advantage can be taken of highly efficient tensor product grid structures. As will be shown in Section 11, the adaptive potential of wavelet bases for elliptic problems can be fully exploited to facilitate an economic solution of each equation (10.21).

Note that, in principle, the approach works for differential as well as integral operators $\mathcal{L}$. The practical realization of the pullback of $\mathcal{L}_{i,i}$ to $\square$ depends, of course, on the type of $\mathcal{L}$. Let us therefore briefly comment on the practical aspects. First observe that, on Γ_i,

$$\psi_\nu^{\Gamma_i} = \chi_{\Gamma_i} P_i \psi_\nu^{\Gamma_i} = \chi_{\Gamma_i} E_i \psi_\nu^{\Gamma_i}. \tag{10.22}$$

Now, if a wavelet $\psi_\nu^{\Gamma_i}$ is supported inside Γ_i, then its trivial extension $(\psi_\nu^{\Gamma_i})^\uparrow$ to $\Gamma_i^\uparrow$ (by zero) already belongs to $F(\Gamma_i^\uparrow)$. However, the extensions constructed by Ciesielski and Figiel (1983) may still give rise to a nontrivial extension $P_i\psi_\nu^{\Gamma_i} = E_i\psi_\nu^{\Gamma_i}$, which on $\Gamma_i^\uparrow \setminus \Gamma_i$ differs from zero, and hence from $(\psi_\nu^{\Gamma_i})^\uparrow$, even though the wavelet $\psi_\nu^{\Gamma_i}$ is not close to the outflow boundary. To suppress this strong coupling between adjacent patches, Dahmen and Schneider (1997*b*) have shown how to construct extensions with the required continuity properties for which all wavelets in Ψ^{Γ_i} that already belong to $F(\Gamma_i)^\uparrow$ are extended by zero. This is again done by exploiting properties of suitable local multiscale bases on $\square$. The nontrivial extension of the remaining (boundary-near) wavelets represent the (scale-dependent) *coupling conditions* for the domain decomposition method. Thus Lagrange multipliers are *not* necessary for coupling the subproblems so that indefinite sytems are avoided. Note also that the discretizations, particularly their respective order of exactness, can be chosen independently on each patch Γ_i.

Since domain decomposition is comparatively less developed for integral operators, we take a closer look at the case where $\mathcal{L}$ has a global kernel K. One can show (Dahmen and Schneider 1997*b*) that the entries of the stiffness matrices then take the form

$$\left\langle \mathcal{L}_{i,l}\psi_\nu^{\Gamma_l}, \psi_\lambda^{\Gamma_i} \right\rangle_{\Gamma_i} = \int_{\square}\int_{\square} K_{i,l}(x,y)\psi_\nu^{\square,l}(y)\psi_\lambda^{\square,i}(x)\,\mathrm{d}y\,\mathrm{d}x, \tag{10.23}$$

where the kernel $K_{i,l}$ depends on the indices ν, λ of the wavelets in the following way. When both wavelets are supported in the interior of the cube, one has $K_{i,l}(x,y) = |\partial\kappa_i(x)||\partial\kappa_l(y)|K(\kappa_i(x),\kappa_i(y))$, where $|\partial\kappa_i|$ denotes the functional determinant of the mapping κ_i. However, when both wavelets have nontrivial extensions, for instance, one has to set

$$K_{i,l}(x,y) = |\partial\kappa_i(x)||\partial\kappa_l(y)|((E_i^* \otimes E_l^*)K)(\kappa_i(x),\kappa_i(y)).$$

The remaining mixed cases are analogous. Hence, in this case the coupling conditions simply boil down to modifications of the kernel. Note that $(E_i^* \otimes E_l^*)$ are *restriction operators*. In particular, this enforces the appropriate boundary conditions. In fact, one (locally) has

$$K_{i,l}(\cdot,y) \in F(\square)^{\uparrow,i}, \quad H_{i,l}(x,\cdot) \in F(\square)^{\uparrow,l}, \tag{10.24}$$

as long as the parameters y, x stay away from the respective outflow boundaries. This has the following important consequences (Dahmen and Schneider 1997*b*).

Moment conditions

Due to the complementary boundary conditions satisfied by the pairs of bases $\Psi^{\square,i}, \tilde{\Psi}^{\square,i}$ on $\square$, the spaces $S(\tilde{\Phi}_j^{\square,i})$ generally do *not* contain all polynomials of order d on $\square$. Hence the wavelets $\psi^{\square,i}$ near the outflow boundary

do not have vanishing moments of corresponding low orders, but annihilate only those polynomials locally contained in $F(\square)^{\uparrow,i}$. Therefore the wavelets still satisfy the estimate (9.16) for any function $f \in F(\square)^{\uparrow,i}$. In view of (10.24), the kernel $K_{i,l}$ satisfies these boundary conditions. Hence the same argument that led to (9.27) still applies to $K_{i,l}(x,y)$. Therefore the wavelets still give rise to estimates like (9.22), (9.27) and hence to optimal compression determined by the order $\tilde{d}$ of the dual multiresolution. In particular, the kernels $K_{i,l}$ become more and more negligible when Γ_i and Γ_l are far apart.

Norm equivalences
Since the wavelets on $\square$ give rise to norm equivalences of the form (6.1), the individual equations (10.21) are easily preconditioned. Moreover, the analysis of corresponding adaptive schemes described in the next section applies to the situation at hand.

11. Analysis of adaptive schemes for elliptic problems

11.1. Some preliminary remarks

The motivation for the following discussion is twofold. On one hand, the inherent potential of wavelet discretizations for adaptivity has been stated often above. However, as natural as it appears, a closer look reveals that on a rigorous and on a conceptual level a number of questions remain open. The discretizations typically involve several types of truncation that often remain unspecified. It is not always clear how corresponding errors propagate in the global scheme and how the tolerances have to be chosen to guarantee a specified overall accuracy. Moreover, thresholding arguments are often not clearly related to the *norm*, that is to measure global accuracy.

In many studies, some *a priori* assumptions are made about the type of singularity, for instance, in terms of the distribution of significant wavelet coefficients. For periodic problems the singularities of the solution are determined by the right-hand side alone (when the coefficients are smooth). This is no longer the case when essential boundary conditions for more complex geometries are imposed. Finally, what is the preferred strategy? In the spirit of image compression, a *fine-to-coarse* approach would aim at discarding insignificant wavelet coefficients, starting from a discretization for a fixed highest level of resolution. The obvious disadvantage is that such an approach accepts the complexity of a fully refined discretization at some stage. Alternatively, in a coarse-to-fine approach, one would try to track the significant wavelets needed to realize the desired accuracy, starting from a coarse discretization. The risk of missing important information along the way is perhaps even higher in this approach. However, the analysis outlined below indicates ways of dealing with this problem. So the subsequent dis-

cussion can be viewed as an attempt to address such questions on a rigorous level, and thereby complement the intriguing adaptive algorithmic developments discussed before.

On the other hand, adaptive techniques have been extensively studied in the context of finite element discretizations of (primarily) elliptic differential equations; see, for instance, Babuška and Miller (1987), Babuška and Rheinboldt (1978), Bank and Weiser (1985), Bornemann et al. (1996), Eriksson, Estep, Hansbo and Johnson (1995) and Verfürth (1994). These methods are based on *a posteriori error indicators* or *estimators*. In practice they have been proven to be quite successful. However, the analysis and the schemes are rather dependent on the particular problem at hand and on the particular type of finite element discretization. The geometrical problems caused by suitable mesh refinements become nontrivial for 3D problems. From a principal point of view, it is furthermore unsatisfactory that the proof of the overall convergence of such schemes usually requires making an *a priori assumption* on the unknown solution, as explained below in more detail.

The adaptive treatment of integral equations in the context of classical finite element discretizations is comparatively less developed. The global nature of the operator makes a local analysis harder. Typical *a posteriori* strategies therefore constrain the structure of admissible meshes (Carstensen 1996), which certainly interferes with the essence of adaptive methods.

These considerations have motivated recent investigations by Dahlke et al. (1997*b*), which substantiate that the main potential of wavelet discretizations lies in adaptivity. Some of the ingredients of the analysis will be outlined next. As in the context of preconditioning, a wide range of problems, including differential as well as integral operators, can be treated in a unified way. A convergence proof is only based on assumptions on the (accessible) data rather than on the (unknown) solution. Furthermore, there is no restriction on the emerging index sets.

Again, a key role is played by the validity of *norm equivalences* of the form (6.1) in combination with *compression* arguments based on the estimates (9.27) or (11.2) below.

To bring out the essential mechanisms, we will refer to the general problem in Section 2.3. Thus we will assume throughout the rest of this section that $\mathcal{L}$ satisfies (2.23) and (2.25). We consider stationary elliptic problems because they also arise in timestepping schemes. In fact, time-dependent problems are in some sense even easier, because information from the preceding time step can be used. Likewise the present formulation can be viewed as an ingredient of an iteration in nonlinear problems.

Moreover, in view of the developments in preceding sections, it is justified to assume that Ψ and $\tilde{\Psi}$ are biorthogonal wavelet bases satisfying the norm equivalences (6.1). Their range of validity is to satisfy (6.14). Specifically,

there then exist finite positive constants c_3, c_4 such that

$$c_3\|\mathbf{D}^{-t}\mathbf{d}\|_{\ell_2} \leq \|\mathbf{d}^T\tilde{\Psi}\|_{H^{-t}} \leq c_4\|\mathbf{D}^{-t}\mathbf{d}\|_{\ell_2}. \tag{11.1}$$

Moreover, the corresponding spaces $S(\Phi_j)$ are assumed to be exact of order d and the wavelets $\psi_\lambda, \lambda \in \nabla_-$, satisfy a suitable version of moment conditions of order $\tilde{d}$ (see, for instance, (9.15)) when $\mathcal{L}$ is an integral operator, or are regular enough when $\mathcal{L}$ is a differential operator, so that in either case the estimate

$$2^{-(|\lambda'|+|\lambda|)t}|\langle \mathcal{L}\psi_{\lambda'}, \psi_\lambda\rangle| \lesssim \frac{2^{-||\lambda|-|\lambda'||\sigma}}{(1+2^{\min(|\lambda|,|\lambda'|)}\operatorname{dist}(\Omega_\lambda, \Omega_{\lambda'}))^{n+2\tilde{d}+2t}} \tag{11.2}$$

holds (see (9.27)). Finally, we will assume that the Galerkin scheme is stable (6.7) (recall the comments in Section 9).

11.2. The saturation property

Suppose for a moment that $\mathcal{L}$ is selfadjoint, in which case (2.23) means that the bilinear form

$$a(u,v) := \langle \mathcal{L}u, v\rangle \tag{11.3}$$

induces a norm which is equivalent to $\|\cdot\|_{H^t}$

$$\|\cdot\|^2 := a(\cdot,\cdot) \sim \|\cdot\|^2_{H^t}. \tag{11.4}$$

In this case a well-known starting point for finite element-based adaptive schemes is the following observation concerning the equivalence between the validity of two-sided error estimates and the so-called *saturation* property (Bornemann et al. 1996). The basic reasoning can be sketched as follows. Suppose that $S \subset V \subset H^t$ are two trial spaces with respective Galerkin solutions u_S, u_V. By orthogonality one has

$$\|u_V - u_S\| \leq \|u - u_S\|.$$

Moreover, one easily checks that

$$\|u - u_V\| \leq \beta\|u - u_S\| \tag{11.5}$$

holds for some $\beta < 1$, if and only if

$$(1-\beta^2)^{1/2}\|u - u_S\| \leq \|u_V - u_S\|. \tag{11.6}$$

Here and elsewhere u denotes the exact solution to $\mathcal{L}u = f$. Thus, if the refined solution u_V captures a sufficiently large portion of the remainder (11.6) the global energy error is guaranteed to decrease by a factor β when passing to the refined solution u_V. Moreover, one has the bounds

$$\|u_V - u_S\| \leq \|u - u_S\| \leq (1-\beta^2)^{-1/2}\|u_V - u_S\|, \tag{11.7}$$

which are computable. In practice one controls the local behaviour of $u_V - u_S$ and refines the mesh at places where (an estimate for) this difference is largest. This results in balancing the error bounds. Although this has been observed to work well in many cases, the principal problem remains that, to prove convergence of the overall adaptive algorithm, something like (11.6) has to be *assumed* about the unknown solution.

Dahlke et al. (1997*b*) pursue a similar updating strategy. Let some current solution space S_Λ and a Galerkin solution u_Λ be given. The objective is to find for a fixed *decay rate* $\beta < 1$, a possibly small $\tilde{\Lambda} \subset \nabla = \Delta_+ \cup \nabla_-$, $\Lambda \subset \tilde{\Lambda}$ such that

$$\|u - u_{\tilde{\Lambda}}\|_{H^t} \leq \beta \|u - u_\Lambda\|_{H^t},$$

which implies convergence.

11.3. A posteriori *error estimates*

It is well known that for elliptic problems the error in energy norm can be estimated by the residual in a dual norm which, at least in principle, can be evaluated. In fact, since

$$r_\Lambda := \mathcal{L}u_\Lambda - f = \mathcal{L}(u_\Lambda - u),$$

the bounded invertibility of $\mathcal{L}$ (2.23) yields

$$c_1 \|r_\Lambda\|_{H^{-t}} \leq \|u - u_\Lambda\|_{H^t} \leq c_2 \|r_\Lambda\|_{H^{-t}}. \tag{11.8}$$

Expanding the residual r_Λ relative to the dual basis $\tilde{\Psi}$ and taking the Galerkin conditions into account, the norm equivalence (11.1) and (11.8) provide

$$c_1 c_3 \left(\sum_{\lambda \in \nabla \setminus \Lambda} \delta_\lambda(\Lambda)^2 \right)^{1/2} \leq \|u - u_\Lambda\|_{H^t} \leq c_2 c_4 \left(\sum_{\lambda \in \nabla \setminus \Lambda} \delta_\lambda(\Lambda)^2 \right)^{1/2}, \tag{11.9}$$

where the quantities

$$\delta_\lambda = \delta_\lambda(\Lambda) := 2^{-t|\lambda|} \, |\langle r_\Lambda, \psi_\lambda \rangle| \,, \quad \lambda \in \nabla \setminus \Lambda,$$

are, in principle, local quantities bounding the error $\|u - u_\Lambda\|_{H^t}$ from below and above. They indicate which wavelets are significant in the representation of u. However, since these quantities involve infinitely many (unknown) terms, (11.9) is in its present form of no practical use.

The objective of the following considerations is to replace the quantities $\delta_\lambda(\Lambda)$ in (11.9) by finitely many computable ones which, up to a given tolerance depending only on the data, still provide lower and upper bounds.

Denoting by $u_{\lambda'} = \langle u_\Lambda, \tilde{\psi}_{\lambda'} \rangle$, $f_\lambda := \langle f, \psi_\lambda \rangle$ the wavelet coefficients of the current approximation u_Λ and the right-hand side f with respect to Ψ and

$\tilde{\Psi}$, respectively, it is helpful to rewrite

$$\delta_\lambda(\Lambda) = 2^{-t|\lambda|}\left|f_\lambda - \sum_{\lambda'\in\Lambda}\langle \mathcal{L}\psi_{\lambda'}, \psi_\lambda\rangle u_{\lambda'}\right| . \tag{11.10}$$

This shows that the size of $\delta_\lambda(\Lambda)$ is influenced by two quantities. First, if the right-hand side f itself has singularities, this will result in large wavelet coefficients f_λ. Second, the sum $\sum_{\lambda'\in\Lambda}\langle \mathcal{L}\psi_{\lambda'}, \psi_\lambda\rangle u_{\lambda'}$ gives the contribution of the current solution which, for instance, could reflect the influence of the boundary. Thus, to estimate the $\delta_\lambda(\Lambda)$ one needs

(a) estimates on the smearing effect of $\mathcal{L}$
(b) some *a priori* knowledge about f.

So far we have only used the ellipticity (2.23) of $\mathcal{L}$ and the norm equivalence (11.1). To deal with (a) one has to make essential use of the decay estimates (11.2). We now describe their use. Let $\delta < \sigma - n/2$, where $\sigma > n/2$ is the constant in (11.2). Choose for any $\epsilon > 0$, positive numbers ϵ_1, ϵ_2 such that

$$\epsilon_1^{2\tilde{d}+2t} + 2^{-\frac{\delta}{\epsilon_2}} \le \epsilon.$$

For each $\lambda \in \nabla$, define the *influence sets*

$$\nabla_{\lambda,\epsilon} := \{\lambda' \in \nabla : \left||\lambda| - |\lambda'|\right| \le \epsilon_2^{-1} \text{ and } 2^{\min\{|\lambda|,|\lambda'|\}}\operatorname{dist}(\Omega_\lambda, \Omega_{\lambda'}) \le \epsilon_1^{-1}\},$$

where Ω_λ again denotes the support of ψ_λ. The sets $\nabla_{\lambda,\epsilon}$ describe the significant portion of $\langle \mathcal{L}u_\Lambda, \psi_\lambda\rangle$ appearing in the residual weights $\delta_\lambda(\Lambda)$ (11.10). In fact, using the estimate of (9.35), one can show the existence of a constant c_5 independent of f and Λ, such that the remainder

$$e_\lambda := \sum_{\lambda'\in\Lambda\setminus\nabla_{\lambda,\epsilon}}\langle \mathcal{L}\psi_{\lambda'}, \psi_\lambda\rangle u_{\lambda'}$$

can be estimated by

$$\Big(\sum_{\lambda\in\nabla\setminus\Lambda} 2^{-|\lambda|2t}\,|e_\lambda|^2\Big)^{\frac{1}{2}} \le c_5\epsilon\,\|u_\Lambda\|\,; \tag{11.11}$$

see also Dahlke et al. (1997*b*), Dahmen et al. (1993*b*) and Dahmen et al. (1994*b*). Note that, again by (6.1),

$$\|u_\Lambda\| \sim \|u_\Lambda\|_{H^t} \sim \|\mathbf{D}^t\langle u_\Lambda, \tilde{\Psi}_\Lambda\rangle\|_{\ell_2},$$

so that the right-hand side in (11.11) can be evaluated by means of the wavelet coefficients of the current solution u_Λ. Moreover, one can even give an *a priori* bound. In fact, the stability of the Galerkin scheme (6.7) states, on account of the uniform boundedness of the Q_Λ^* in H^{-t} (see Theorem 5.8),

that

$$\|u_\Lambda\| \lesssim \|Q_\Lambda^* f\|_{H^{-t}} \le c_5' \|f\|_{H^{-t}}. \tag{11.12}$$

As for (b) above, by construction, the *significant neighbourhood* of Λ in $\nabla \setminus \Lambda$

$$N_{\Lambda,\epsilon} := \{\lambda \in \nabla \setminus \Lambda : \Lambda \cap \nabla_{\lambda,\epsilon} \neq \emptyset\} \tag{11.13}$$

is finite

$$\#N_{\Lambda,\epsilon} < \infty.$$

Outside $N_{\Lambda,\epsilon}$, the quantities $\delta_\lambda(\Lambda)$ in (11.10) are essentially influenced by wavelet coefficients of f. But this portion is a remainder of f. In fact, by (6.1),

$$\begin{aligned}\Big(\sum_{\lambda\in\nabla\setminus(\Lambda\cup N_{\Lambda,\epsilon})} 2^{-2t|\lambda|}\,|f_\lambda|^2\Big)^{\frac12} &\le c_2\|f - Q^*_{\Lambda\cup N_{\Lambda,\epsilon}}\|_{H^{-t}} \\ &\le c_6 \inf_{v\in\tilde S_{\Lambda\cup N_{\Lambda,\epsilon}}} \|f-v\|_{H^{-t}} \le c_6 \inf_{v\in\tilde S_\Lambda}\|f-v\|_{H^{-t}},\end{aligned}$$

for some $c_6 < \infty$. This suggests defining

$$d_\lambda(\Lambda,\epsilon) := 2^{-t|\lambda|}\Big|\sum_{\lambda'\in\Lambda\cap\nabla_{\lambda,\epsilon}} \langle \mathcal{L}\psi_{\lambda'},\psi_\lambda\rangle u_{\lambda'}\Big|, \quad \lambda\in\nabla\setminus\Lambda.$$

Note that, in view of (11.13),

$$d_\lambda(\Lambda,\epsilon) = 0, \quad \lambda\in\nabla\setminus\Lambda, \quad \lambda\notin N_{\lambda,\epsilon}. \tag{11.14}$$

The main result can now be formulated as follows (Dahlke et al. 1997*b*).

Theorem 11.1 Under the above assumptions, one has

$$\|u-u_\Lambda\|_{H^t} \le c_2c_4\Big(\Big(\sum_{\lambda\in N_{\Lambda,\epsilon}} d_\lambda(\Lambda,\epsilon)^2\Big)^{\frac12} + c_5'\epsilon\,\|f\|_{H^{-t}} + c_6\inf_{v\in\tilde S_\Lambda}\|f-v\|_{H^{-t}}\Big)$$

as well as

$$\Big(\sum_{\lambda\in N_{\Lambda,\epsilon}} d_\lambda(\Lambda,\epsilon)^2\Big)^{\frac12} \le \frac{1}{c_1c_3}\|u-u_\Lambda\|_{H^t} + c_5'\epsilon\,\|f\|_{H^{-t}} + c_6\inf_{v\in\tilde S_\Lambda}\|f-v\|_{H^{-t}}.$$

Moreover, for any $\tilde\Lambda\subset\nabla$, $\Lambda\subset\tilde\Lambda$, one has

$$\Big(\sum_{\lambda\in\tilde\Lambda\cap N_{\Lambda,\epsilon}} d_\lambda(\Lambda,\epsilon)^2\Big)^{\frac12} \le \frac{1}{c_1c_3}\|u_{\tilde\Lambda}-u_\Lambda\|_{H^t} + c_5'\epsilon\,\|f\|_{H^{-t}} + c_6\inf_{v\in\tilde S_\Lambda}\|f-v\|_{H^{-t}}.$$

This result provides, up to the controllable tolerance

$$\tau(\Lambda,\epsilon) := c_5'\mathrm{eps}\,\|f\|_{H^{-t}} + c_6\inf_{v\in\tilde S_\Lambda}\|f-v\|_{H^{-t}},$$

computable lower and upper bounds for the error $\|u - u_\Lambda\|_{H^t}$. For second-order two-point boundary value problems, estimates of the above type were first obtained by Bertoluzza (1994). Under much more specialized assumptions, results of similar nature have also been established in the finite element context; see, for example, Dörfler (1996).

11.4. Convergence of an adaptive refinement scheme

In the present setting, it can be shown with the aid of Theorem 11.1 that, under mild assumptions on the right-hand side f, a suitable adaptive choice of $\tilde{\Lambda}$ *enforces* the validity of the saturation property (11.6). We continue with the notation of Section 11.3. However, for simplicity we confine the discussion to the selfadjoint case (11.3), (11.4), that is, the norm $\|\cdot\|_{H^t}$ is replaced by the *energy norm* $\|\cdot\|$. The constants c_i have to be properly adjusted. The following theorem was proved by Dahlke et al. (1997*b*).

Theorem 11.2 Let tol > 0 be a given tolerance and fix $\theta \in (0,1)$. Define

$$C^* := \left(\frac{1}{c_1 c_3} + \frac{1-\theta}{2c_2 c_4}\right), \tag{11.15}$$

choose $\mu > 0$ such that

$$\mu C^* \leq \frac{1-\theta}{2(2-\theta)c_2 c_4}, \tag{11.16}$$

and set

$$\epsilon := \frac{\mu \text{ tol}}{2c_5' \|f\|_{H^{-t}}}. \tag{11.17}$$

Suppose that for $\Lambda \subset \nabla$, one has

$$c_6 \inf_{v \in \tilde{S}_\Lambda} \|f - v\|_{H^{-t}} < \frac{1}{2}\mu \text{ tol}.$$

Then, whenever $\tilde{\Lambda} \subset \nabla$, $\Lambda \subset \tilde{\Lambda}$ is chosen so that

$$\Big(\sum_{\lambda \in \tilde{\Lambda} \cap N_{\Lambda,\epsilon}} d_\lambda(\Lambda,\epsilon)^2\Big)^{\frac{1}{2}} \geq (1-\theta)\Big(\sum_{\lambda \in N_{\Lambda,\epsilon}} d_\lambda(\Lambda,\epsilon)^2\Big)^{\frac{1}{2}},$$

there exists a constant $\beta \in (0,1)$ depending only on the constants μ, θ, c_i, $i = 1, \ldots, 6$, such that either

$$\|u - u_{\tilde{\Lambda}}\| \leq \beta \|u - u_\Lambda\|$$

or

$$\Big(\sum_{\lambda \in N_{\lambda,\epsilon}} d_\lambda(\Lambda,\epsilon)^2\Big)^{\frac{1}{2}} = \Big(\sum_{\lambda \in \nabla \setminus \Lambda} d_\lambda(\Lambda,\epsilon)^2\Big)^{\frac{1}{2}} < \text{tol}.$$

For the discussion of unsymmetric problems see Dahlke et al. (1997*b*) and Hochmuth (1996).

Of course, the idea is to choose $\tilde{\Lambda} \supset \Lambda$ as small as possible, that is, in any case $\tilde{\Lambda} \setminus \Lambda \subset N_{\Lambda,\epsilon}$. This leads to the following.

ALGORITHM 7
Choose $\Lambda_0 = \emptyset$, eps > 0, tol $>$ eps, $\theta \in (0,1)$.

(1) Compute C^*, μ according to (11.15), (11.16).
(2) Compute $\epsilon = \epsilon(\mu, \text{tol})$ by (11.17).
(3) Determine $\Lambda \subset \nabla$, $\Lambda_0 \subset \Lambda$ such that

$$c_6 \inf_{v \in \tilde{S}_\Lambda} \|f - v\|_{H^{-t}} < \frac{1}{2}\mu \text{ tol}.$$

(4) Solve

$$\langle \mathcal{L} u_\Lambda, v \rangle = \langle f, v \rangle, \quad \text{for all } v \in S_\Lambda.$$

(5) Compute

$$\eta_{\Lambda,\epsilon} := \Big(\sum_{\lambda \in N_{\Lambda,\epsilon}} d_\lambda(\Lambda,\epsilon)^2 \Big)^{\frac{1}{2}}.$$

If $\eta_{\Lambda,\epsilon} <$ tol:

- If tol $\leq$ eps, accept u_Λ as solution and stop.
- Otherwise set $\Lambda \to \Lambda_0$, $\frac{\text{tol}}{2} \to$ tol, and go to (2).

Otherwise, go to (6).

(6) Determine $\tilde{\Lambda}$ with $\Lambda \subset \tilde{\Lambda} \subset \Lambda \cup N_{\Lambda,\epsilon}$ such that

$$\Big(\sum_{\lambda \in \tilde{\Lambda}} d_\lambda(\Lambda,\epsilon)^2 \Big)^{\frac{1}{2}} \geq (1-\theta)\eta_{\Lambda,\epsilon}.$$

Set $\tilde{\Lambda} \to \Lambda$ and go to (4).

Although quite different with regard to its technical ingredients, the above algorithm is very similar in spirit to the adaptive scheme proposed by Dörfler (1996) for bivariate piecewise linear finite element discretizations of Poisson's equation. As above, Dörfler (1996) chooses the coarsest grid in such a way that all errors stemming from data are kept below any desired tolerance.

A brief comment on step (4) in Algorithm 7 is in order. By Theorem 6.1, the principal sections of the matrix $\mathbf{B}_\Lambda := \mathbf{D}^{-t} \langle \mathcal{L}\Psi_\Lambda, \Psi_\Lambda \rangle^{\mathrm{T}} \mathbf{D}^{-t}$ are well conditioned. This can be exploited to update a current Galerkin approximation u_Λ, as follows. Let $\mathbf{u}_\Lambda := \langle u_\Lambda, \tilde{\Psi}_\Lambda \rangle^T$ be the vector of wavelet coefficients of u_Λ. To compute the coefficient vector $\mathbf{u}_{\tilde{\Lambda}}$ of $u_{\tilde{\Lambda}}$ we choose an initial approximation $\mathbf{v}$ according to

$$v_\lambda = \begin{cases} u_\lambda, & \lambda \in \Lambda \\ w_\lambda, & \lambda \in \tilde{\Lambda} \setminus \Lambda, \end{cases} \tag{11.18}$$

where $\mathbf{w}_{\tilde{\Lambda}\setminus\Lambda}$ are the coefficients of the Galerkin solution $w_{\tilde{\Lambda}\setminus\Lambda}$ of the complement system

$$\langle \mathcal{L} w_{\tilde{\Lambda}\setminus\Lambda}, v\rangle = \langle f, v\rangle, \quad v \in S_{\tilde{\Lambda}\setminus\Lambda} := \operatorname{span}\{\psi_\lambda : \lambda \in \tilde{\Lambda}\setminus\Lambda\}.$$

The corresponding matrix entries have to be determined anyway for the adaptive refinement. Since, by (6.16), the corresponding section $\mathbf{B}_{\tilde{\Lambda}\setminus\Lambda}$ of $\mathbf{B}_{\tilde{\Lambda}}$ is well conditioned, only a few conjugate gradient iterations are expected to be necessary to approximate $\mathbf{w}_{\tilde{\Lambda}\setminus\Lambda}$ well enough to provide a good starting approximation of the form (11.18). This will then have to be improved by (a few) further iterations on the system matrix $\mathbf{B}_{\tilde{\Lambda}}$.

11.5. Besov regularity

The results of the previous section imply convergence of the adaptic scheme but do not provide any concrete information about the efficiency, for instance by relating the final accuracy to the number $\#\Lambda$ needed to realize it by the scheme. The ideal case would be that the scheme picks at each stage the *minimal* number of additional indices needed to reduce the current error by a fixed fraction. This cannot be concluded, since the scheme selects the indices with respect to bounds, not with respect to the true error. Nevertheless, since these bounds are lower and upper ones, one expects that the selected index sets are close to minimal ones. Given this assumption, the question of for which circumstances the above adaptive scheme is significantly more efficient than working simply with uniform refinements is closely related to characterizing the efficiency of so-called *best N-term approximation*, or *nonlinear approximation*. A beautiful theory for these issues has been developed in a number of papers; see, for instance, DeVore and Popov (1988*a*), DeVore et al. (1992) and DeVore and Lucier (1992). Here we indicate very briefly some typical facts suited to the present context. To this end, consider

$$\sigma_{N,t}(g) := \inf\left\{ \|g - \sum_{\lambda\in\Lambda} d_\lambda \psi_\lambda\|_{H^t} : d_\lambda \in \mathbb{R},\ \lambda \in \Lambda \subset \nabla,\ \#\Lambda = N \right\}.$$

Employing the norm equivalence (5.38) yields

$$\sigma_{N,t}(v) \sim \sigma_{N,0}(\Sigma_t v) := \sigma_N(\Sigma_t v), \tag{11.19}$$

which in turn leads to the following (Dahlke, Dahmen and DeVore 1997*a*).

Remark 11.3 Let $v \in H^t$. We take Λ_N to be a set of N indices λ for which $2^{t|\lambda|}|\langle v, \tilde{\psi}_\lambda\rangle|$ is largest. Then one has

$$\sigma_{N,t}(v) \sim \|v - Q_{\Lambda_N} v\|_{H^t}, \quad N \in \mathbb{N}. \tag{11.20}$$

Thus, picking the N first largest *weighted* coefficients realizes asymptotically the best N-term approximation relative to the norm $\|\cdot\|_{H^t}$ and hence, in case (11.4), also relative to the energy norm $\|\cdot\|$.

Combining (11.20) with analogous results about $\sigma_{N,t}$ for $t = 0$, the best N-term approximation of a function v relative to $\|\cdot\|_{H^t}$ can be characterized in terms of its *Besov regularity* (Dahlke et al. 1997*a*).

Proposition 11.4 Assume that $\alpha - t < \gamma$ and let for $t \leq \alpha$

$$\frac{1}{\tau^*} := \frac{\alpha - t}{n} + \frac{1}{2}. \tag{11.21}$$

Then one has

$$\sum_{N=1}^{\infty} \left(N^{(\alpha-t)/n}\sigma_{N,t}(v)\right)^{\tau^*} < \infty \tag{11.22}$$

(where n is again the spatial dimension of the underlying domain Ω), if and only if $v \in B^{\alpha}_{\tau^*}(L_{\tau^*}(\Omega))$. Recall the characterization of Besov norms (5.46).

Proposition 11.4 has an interesting application to the Poisson equation

$$-\Delta u = f \quad \text{in} \quad \Omega, \quad u = 0 \quad \text{on} \quad \partial\Omega, \tag{11.23}$$

when Ω is a bounded *Lipschitz domain* in $\mathbb{R}^n$. The efficiency of the best N-term approximation when applied to the solution of Laplace's equation has been studied by Dahlke and DeVore (1995). However, these results were formulated with respect to approximation in $L_2(\Omega)$. For elliptic equations, the energy norm is more natural. A combination of Proposition 11.4 and the results of Dahlke and DeVore (1995) provides the following result concerning approximation relative to $\|\cdot\|_{H^1}$ (Dahlke et al. 1997*a*).

Proposition 11.5 Let Ω be a bounded Lipschitz domain in $\mathbb{R}^n$, and let u denote the solution of (11.23) with $f \in B_2^{\alpha-1}(L_2(\Omega))$, $\alpha \geq 1$. Then

$$\sum_{N=1}^{\infty} \left(N^{s/n}\sigma_{N,1}(u)\right)^{\tau} < \infty \quad \text{for all} \quad 0 < s < s^*/3, \tag{11.24}$$

where $s^* = \min\{\frac{3n}{2(n-1)}, \alpha + 1\}$ and $\tau = (s-1)/n + 1/2$.

To illustrate this result, consider the example where $n = 2$. If $\alpha \geq 2$, then $s^* = 3$. Hence, in this case, the nonlinear method gives an H^1–approximation to u of order up to $N^{-1/2}$, whereas a linear method, that is, uniform refinements, using N terms could only give $N^{-1/4}$ in the worst case.

These facts indicate that adaptive refinements will generally perform significantly better. Establishing a closer connection between the adaptive scheme discussed in the preceding section and N-term approximation is an interesting question under current investigation.

12. What else?

Evidently a lot more than could be included in a survey. Therefore I would like to add only a few brief comments on further interesting directions.

Collocation
It has been pointed out that collocation plays an important role in connection with the fast evaluation of nonlinear terms. Bertoluzza (1997) has discussed some promising features of collocation in connection with highly accurate discretizations. There, interpolatory scaling functions are employed and corresponding analogues to hierarchical bases are established. Again, interpolatory representations are helpful with regard to evaluating nonlinear terms.

Transport problems
The above concepts are more or less tailored to *elliptic* problems. It is less clear how to treat transport terms. Typical model problems are *convection–diffusion* problems of the form

$$-\Delta u + \beta(x) \cdot \nabla u = f \quad \text{in} \quad \Omega, \quad u = 0 \quad \text{on} \quad \partial\Omega, \tag{12.1}$$

where the convection term is strongly dominant. Canuto and Cravero (1996) have proposed discretizing (12.1) with a conventional finite element method and use wavelet expansions of the current solution to determine successive mesh refinements at locations where wavelet coefficients are large. First results by Dahmen, Müller and Schlinkmann (199x) indicate that the concept of stable completions can be successfully employed to design *level-dependent* Petrov–Galerkin discretizations, which in a multigrid context recover the usual multigrid efficiency for elliptic problems even in the case of strong convection terms.

Discrete multiresolution concepts have been developed by Harten (1995), with special emphasis on the treatment of *hyperbolic conservation* laws. It is well known that such systems can be viewed as evolution equations for *cell averages*. This fact serves as the basis for *finite volume discretizations*. However, when advancing in time, these schemes require at some stage the computation of fluxes across cell boundaries, which in turn need pointwise values of the conservative variables. To realize high accuracy, one therefore has to design highly accurate reconstruction schemes to recover the pointwise values from the cell averages, which is actually the only place where a discretization error is introduced. Unfortunately, in realistic problems the evaluation of corresponding numerical fluxes is very costly. The main thrust of Harten's concept therefore aims at reducing the cost of numerical flux computations according to the following idea. Fluxes are initially computed only on a very coarse grid (using, however, data corresponding to the highest

level of resolution). The flux values on successively finer grids are then determined either by cheap interpolation schemes from those on coarser levels, whenever they are smooth, or otherwise by expensive accurate schemes. This decision is based on a suitable multiscale representation of the data. This is highly reminiscent of data compression techniques. The underlying multiscale decomposition concept proposed by Harten is very flexible and has to be made concrete in each application. The multiscale transformations have the format (3.26) and (3.28), although, in principle, no explicit knowledge of underlying bases Φ_j, Ψ_j is required. Nevertheless, many technical as well as conceptual problems arise when applying this methodology to concrete problems, in particular, when dealing with several space variables. Some recent contributions can be found in Gottschlich-Müller and Müller (1996), Sjögreen (1995) and Sonar (1995), for instance.

Software

To unfold the full efficiency of most of the concepts discussed so far, rather new data structures are needed. It does not seem to be possible to simply hook wavelet components to existing software for conventional discretization schemes. Existing codes still seem to be confined to model problems. The beginnings of a systematic software development for wavelet schemes in a PDE context are discussed by Barsch et al. (1997), for example.

Wavelets as analysis tools

The primary objective of the developments detailed in this paper is the understanding and design of highly efficient solvers for large-scale problems. I find the variety of contributions very promising and interesting. However, because of the state of the software development, and for conceptual reasons mainly in connection with geometry constraints, it is fair to say that wavelet schemes have not yet become quite competitive with well tuned multigrid codes for realistic problems. On the other hand, the discussion also indicates that the potential of wavelets has not yet been exhausted, and that the results that have been achieved so far provide a highly stimulating source of ideas and further progress. In fact, the above comments on the convection–diffusion problem suggest that true benefit for future generations of multiscale techniques may result from a marriage of different methodologies. I would be very pleased if the present paper could be of some help in this regard.

On the other hand, it has already been indicated that, aside from algorithmic developments, wavelets offer powerful analysis tools. An example, namely investigating the *boundedness of Galerkin projectors in L_p-Sobolev spaces* has already been mentioned by Angeletti et al. (1997). The determination of *Besov regularity* of solutions to elliptic boundary value problems (Dahlke and DeVore 1995, Dahlke 1996) is another intriguing instance which

is important for the understanding of adaptivity. The multiresolution approach to *homogenization* (Brewster and Beylkin 1995, Dorobantu 1995) opens further startling perspectives. Wavelets have recently been employed in the study of turbulence and multiscale interaction of flow phenomena (Elezgaray, Berkooz, Dankowicz, Holmes and Myers 1997, Wickerhauser, Farge and Goirand 1997, Farge et al. 1992).

In summary, it seems that wavelets have become indispensible as a conceptual source for understanding multiscale phenomena and corresponding solution schemes.

Acknowledgements

I would like to thank A. Kunoth and S. Müller for their assistance and valuable suggestions during the preparation of the manuscript. I am also indebted to T. Bronger for turning my terrible handwriting into the above result.

REFERENCES

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

A. Alpert, G. Beylkin, R. Coifman and V. Rokhlin (1993), 'Wavelet-like bases for the fast solution of second-kind integral equations', *SIAM J. Sci. Statist. Comput.* **14**, 159–184.

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

L. Andersson, N. Hall, B. Jawerth and G. Peters (1994), Wavelets on closed subsets of the real line, in *Topics in the Theory and Applications of Wavelets* (L. L. Schumaker and G. Webb, eds), Academic Press, Boston, pp. 1–61.

J. M. Angeletti, S. Mazet and P. Tchamitchian (1997), Analysis of second order elliptic operators without boundary conditions and with VMO or Hölderian coefficients, in *Multiscale Wavelet Methods for PDEs* (W. Dahmen, A. J. Kurdila and P. Oswald, eds), Academic Press. To appear.

A. Averbuch, G. Beylkin, R. Coifman and M. Israeli (1995), Multiscale inversion of elliptic operators, in *Signal and Image Representation in Combined Spaces* (J. Zeevi and R. Coifman, eds), Academic Press, pp. 1–16.

I. Babuška (1973), 'The finite element method with Lagrange multipliers', *Numer. Math.* **20**, 179–192.

I. Babuška and A. Miller (1987), 'A feedback finite element method with *a posteriori* error estimation: Part I. The finite element method and some basic properties of the *a posteriori* error estimator', *Comput. Meth. Appl. Mech. Eng.* **61**, 1–40.

I. Babuška and W. C. Rheinboldt (1978), 'Error estimates for adaptive finite element computations', *SIAM J. Numer. Anal.* **15**, 736–754.

R. E. Bank and A. Weiser (1985), 'Some *a posteriori* error estimates for elliptic partial differential equations', *Math. Comput.* **44**, 283–301.

R. E. Bank, A. H. Sherman and A. Weiser (1983), Refinement algorithms and data structures for regular local mesh refinement, in *Scientific Computing* (R. Stepleman and et al., eds), IMACS, North-Holland, Amsterdam, pp. 3–17.

T. Barsch, A. Kunoth and K. Urban (1997), Towards object oriented software tools for numerical multiscale methods for PDEs using wavelets, in *Multiscale Wavelet Methods for PDEs* (W. Dahmen, A. J. Kurdila and P. Oswald, eds), Academic Press. To appear.

J. Bergh and J. Löfström (1976), *Interpolation Spaces: An Introduction*, Springer.

G. Berkooz, J. Elezgaray and P. Holmes (1993), Wavelet analysis of the motion of coherent structures, in *Progress in Wavelet Analysis and Applications* (Y. Meyer and S. Roques, eds), Editions Frontières, pp. 471–476.

S. Bertoluzza (1994), '*A posteriori* error estimates for wavelet Galerkin methods', Istituto di Analisi Numerica, Pavia. Preprint Nr. 935.

S. Bertoluzza (1997), An adaptive collocation method based on interpolating wavelets, in *Multiscale Wavelet Methods for PDEs* (W. Dahmen, A. J. Kurdila and P. Oswald, eds), Academic Press. To appear.

G. Beylkin (1992), 'On the representation of operators in bases of compactly supported wavelets', *SIAM J. Numer. Anal.* **29**, 1716–1740.

G. Beylkin (1993), Wavelets and fast numerical algorithms, in *Different Perspectives on Wavelets* (I. Daubechies, ed.), Vol. 47 of *Proc. Symp. Appl. Math.*, pp. 89–117.

G. Beylkin and J. M. Keiser (1997), An adaptive pseudo-wavelet approach for solving nonlinear partial differential equations, in *Multiscale Wavelet Methods for PDEs* (W. Dahmen, A. J. Kurdila and P. Oswald, eds), Academic Press. To appear.

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

G. Beylkin, J. M. Keiser and L. Vozovoi (1996), 'A new class of stable time discretization schemes for the solution of nonlinear PDE's'. Preprint.

F. Bornemann and H. Yserentant (1993), 'A basic norm equivalence for the theory of multilevel methods', *Numer. Math.* **64**, 455–476.

F. Bornemann, B. Erdmann and R. Kornhuber (1996), '*A posteriori* error estimates for elliptic problems in two and three space dimensions', *SIAM J. Numer. Anal.* **33**, 1188–1204.

D. Braess (1997), *Finite Elements: Theory, Fast Solvers and Applications in Solid Mechanics*, Cambridge University Press.

J. H. Bramble (1981), 'The Lagrange multiplier method for Dirichlet's problem', *Math. Comput.* **37**, 1–11.

J. H. Bramble (1993), *Multigrid Methods*, Vol. 294 of *Pitman Research Notes in Mathematics*, Longman, London. Co-published in the USA with Wiley, New York.

J. H. Bramble and J. Pasciak (1988), 'A preconditioning technique for indefinite systems resulting from mixed approximations for elliptic problems', *Math. Comput.* **50**, 1–17.

J. H. Bramble and J. Pasciak (1994), Iterative techniques for time dependent Stokes problems. Preprint.

J. H. Bramble, Z. Leyk and J. E. Pasciak (1994), 'The analysis of multigrid algorithms for pseudo-differential operators of order minus one', *Math. Comput.* **63**, 461–478.

J. H. Bramble, J. E. Pasciak and J. Xu (1990), 'Parallel multilevel preconditioners', *Math. Comput.* **55**, 1–22.

A. Brandt and A. A. Lubrecht (1990), 'Multilevel matrix multiplication and the fast solution of integral equations', *J. Comput. Phys.* **90**, 348–370.

A. Brandt and K. Venner (preprint), Multilevel evaluation of integral transforms on adaptive grids.

M. E. Brewster and G. Beylkin (1995), 'A multiresolution strategy for numerical homogenization', *Appl. Comput. Harm. Anal.* **2**, 327–349.

F. Brezzi and M. Fortin (1991), *Mixed and Hybrid Finite Element Methods*, Springer, New York.

C. Canuto and I. Cravero (1996), Wavelet-based adaptive methods for advection–diffusion problems. Preprint, University of Torino.

J. M. Carnicer, W. Dahmen and J. M. Peña (1996), 'Local decomposition of refinable spaces', *Appl. Comput. Harm. Anal.* **3**, 127–153.

J. Carrier, L. Greengard and V. Rokhlin (1988), 'A fast adaptive multipole algorithm for particle simulations', *SIAM J. Sci. Statist. Comput.* **9**, 669–686.

C. Carstensen (1996), 'Efficiency of *a posteriori* BEM-error estimates for first-kind integral equations on quasi-uniform meshes', *Math. Comput.* **65**, 69–84.

A. S. Cavaretta, W. Dahmen and C. A. Micchelli (1991), *Stationary Subdivision.* Mem. Amer. Math. Soc., **93**, #453.

T. Chan and T. Mathew (1994), Domain decomposition algorithms, in *Acta Numerica*, Vol. 3, Cambridge University Press, pp. 61–143.

P. Charton and V. Perrier (1995), Towards a wavelet-based numerical scheme for the two-dimensional Navier–Stokes equations, in *Proceedings of the ICIAM*, Hamburg.

P. Charton and V. Perrier (1996), A pseudo-wavelet scheme for the two-dimensional Navier–Stokes equations. Preprint.

C. K. Chui and E. Quak (1992), Wavelets on a bounded interval, in *Numerical Methods of Approximation Theory* (D. Braess and L. Schumaker, eds), Birkhäuser, Basel, pp. 1–24.

Z. Ciesielski and T. Figiel (1983), 'Spline bases in classical function spaces on compact C^∞ manifolds, parts I & II', *Studia Math.* **76**, 1–58, 95–136.

A. Cohen (1994). Private communication.

A. Cohen and I. Daubechies (1993), 'Non-separable bidimensional wavelet bases', *Revista Mat. Iberoamericana* **9**, 51–137.

A. Cohen and J.-M. Schlenker (1993), 'Compactly supported bi-dimensional wavelet bases with hexagonal symmetry', *Constr. Appr.* **9**, 209–236.

A. Cohen, W. Dahmen and R. DeVore (1995), 'Multiscale decompositions on bounded domains'. IGPM-Report 113, RWTH Aachen. To appear in *Trans. Amer. Math. Soc.*

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

A. Cohen, I. Daubechies and P. Vial (1993), 'Wavelets on the interval and fast wavelet transforms', *Appl. Comput. Harm. Anal.* **1**, 54–81.

R. R. Coifman, Y. Meyer and M. V. Wickerhauser (1992), Size properties of the wavelet packets, in *Wavelets and their Applications* (G. Beylkin, R. R. Coifman, I. Daubechies, S. Mallat, Y. Meyer, L. A. Raphael and M. B. Ruskai, eds), Jones and Bartlett, Cambridge, MA, pp. 453–470.

R. R. Coifman, Y. Meyer, S. R. Quake and M. V. Wickerhauser (1993), Signal processing and compression with wavelet packets, in *Progress in Wavelet Analysis and Applications* (Y. Meyer and S. Roques, eds), Editions Frontières, Paris, pp. 77–93.

S. Dahlke (1996), 'Wavelets: Construction principles and applications to the numerical treatment of operator equations'. Habilitationsschrift, Aachen.

S. Dahlke and R. DeVore (1995), 'Besov regularity for elliptic boundary value problems'. IGPM-Report 116, RWTH Aachen.

S. Dahlke and I. Weinreich (1993), 'Wavelet-Galerkin methods: An adapted biorthogonal wavelet basis', *Constr. Approx.* **9**, 237–262.

S. Dahlke and I. Weinreich (1994), 'Wavelet bases adapted to pseudo-differential operators', *Appl. Comput. Harm. Anal.* **1**, 267–283.

S. Dahlke, W. Dahmen and R. DeVore (1997*a*), Nonlinear approximation and adaptive techniques for solving elliptic operator equations, in *Multiscale Wavelet Methods for PDEs* (W. Dahmen, A. Kurdila and P. Oswald, eds), Academic Press. To appear.

S. Dahlke, W. Dahmen and V. Latour (1995), 'Smooth refinable functions and wavelets obtained by convolution products', *Appl. Comput. Harm. Anal.* **2**, 68–84.

S. Dahlke, W. Dahmen, R. Hochmuth and R. Schneider (1997*b*), 'Stable multiscale bases and local error estimation for elliptic problems', *Appl. Numer. Math.* **23**, 21–48.

W. Dahmen (1994), Some remarks on multiscale transformations, stability and biorthogonality, in *Wavelets, Images and Surface Fitting* (P. J. Laurent, A. le Méhauté and L. L. Schumaker, eds), A. K. Peters, Wellesley, MA, pp. 157–188.

W. Dahmen (1995), Multiscale analysis, approximation, and interpolation spaces, in *Approximation Theory VIII, Wavelets and Multilevel Approximation* (C. K. Chui and L. L. Schumaker, eds), World Scientific, pp. 47–88.

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

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

W. Dahmen and C. A. Micchelli (1993), 'Using the refinement equation for evaluating integrals of wavelets', *SIAM J. Numer. Anal.* **30**, 507–537.

W. Dahmen and R. Schneider (1996), 'Composite wavelet bases'. IGPM-Report 133, RWTH Aachen.

W. Dahmen and R. Schneider (1997*a*), Wavelets on manifolds I. Construction and domain decomposition. In preparation.

W. Dahmen and R. Schneider (1997*b*), Wavelets on manifolds II. Applications to boundary integral equations. In preparation.

W. Dahmen, B. Kleemann, S. Prößdorf and R. Schneider (1994*a*), A multiscale method for the double layer potential equation on a polyhedron, in *Advances in Computational Mathematics* (H. P. Dikshit and C. A. Micchelli, eds), World Scientific, pp. 15–57.

W. Dahmen, B. Kleemann, S. Prößdorf and R. Schneider (1996*a*), Multiscale methods for the solution of the Helmholtz and Laplace equation, in *Boundary*

Element Methods: Report from the Final Conference of the Priority Research Programme 1989–1995 of the German Research Foundation, Oct. 2–4, 1995 in Stuttgart (W. Wendland, ed.), Springer, pp. 180–211.

W. Dahmen, A. Kunoth and R. Schneider (1997), Operator equations, multiscale concepts and complexity, in *Mathematics of Numerical Analysis: Real Number Algorithms* (J. Renegar, M. Shub and S. Smale, eds), Vol. 32 of *Lectures in Applied Mathematics*, AMS, Providence, RI, pp. 225–261.

W. Dahmen, A. Kunoth and K. Urban (1996*b*), 'Biorthogonal spline-wavelets on the interval: Stability and moment conditions'. IGPM-Report 129, RWTH Aachen.

W. Dahmen, A. Kunoth and K. Urban (1996*c*), 'A wavelet Galerkin method for the Stokes problem', *Computing* **56**, 259–302.

W. Dahmen, S. Müller and T. Schlinkmann (199x), Multiscale techniques for convection-dominated problems. In preparation.

W. Dahmen, P. Oswald and X.-Q. Shi (1993*a*), 'C^1-hierarchical bases', *J. Comput. Appl. Math.* **9**, 263–281.

W. Dahmen, S. Prößdorf and R. Schneider (1993*b*), 'Wavelet approximation methods for pseudodifferential equations II: Matrix compression and fast solution', *Advances in Computational Mathematics* **1**, 259–335.

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

W. Dahmen, S. Prößdorf and R. Schneider (1994*c*), 'Wavelet approximation methods for pseudodifferential equations I: Stability and convergence', *Math. Z.* **215**, 583–620.

R. Danchin (1997), PhD thesis, Université Paris VI. In preparation.

I. Daubechies (1988), 'Orthonormal bases of compactly supported wavelets', *Comm. Pure Appl. Math.* **41**, 909–996.

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

C. de Boor, R. DeVore and A. Ron (1993), 'On the construction of multivariate (pre-) wavelets', *Constr. Approx.* **9**(2), 123–166.

R. DeVore and B. Lucier (1992), Wavelets, in *Acta Numerica*, Vol. 1, Cambridge University Press, pp. 1–56.

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

R. DeVore and V. Popov (1988*b*), Interpolation spaces and nonlinear approximation, in *Function Spaces and Approximation* (M. Cwikel, J. Peetre, Y. Sagher and H. Wallin, eds), Lecture Notes in Math., Springer, pp. 191–205.

R. DeVore and B. Sharpley (1993), 'Besov spaces on domains in $\mathbb{R}^d$', *Trans. Amer. Math. Soc.* **335**, 843–864.

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

W. Dörfler (1996), 'A convergent adaptive algorithm for Poisson's equation', *SIAM J. Numer. Anal.* **33**, 1106–1124.

M. Dorobantu (1995), Wavelet-Based Algorithms for Fast PDE Solvers, PhD thesis, Royal Institute of Technology, Stockholm University.

J. Elezgaray, G. Berkooz, H. Dankowicz, P. Holmes and M. Myers (1997), Local models and large scale statistics of the Kuramoto–Sivansky equation, in *Multiscale Wavelet Methods for PDEs* (W. Dahmen, A. J. Kurdila and P. Oswald, eds), Academic Press. To appear.

B. Enquist, S. Osher and S. Zhong (1994), 'Fast wavelet-based algorithms for linear evolution operators', *SIAM J. Sci. Comput.* **15**, 755–775.

K. Eriksson, D. Estep, P. Hansbo and C. Johnson (1995), Introduction to adaptive methods for differential equations, in *Acta Numerica*, Vol. 4, Cambridge University Press, pp. 105–158.

M. Farge, E. Goirand, Y. Meyer, F. Pascal and M. V. Wickerhauser (1992), 'Improved predictability of two-dimensional turbulent flows using wavelet packet compression', *Fluid Dynam. Res.* **10**, 229–250.

B. Fornberg and G. B. Whitham (1978), 'A numerical and theoretical study of certain nonlinear wave phenomena', *Philos. Trans. R. Soc. London, Ser. A* **289**, 373–404.

M. Fortin (1977), 'An analysis of convergence of mixed finite element methods', *R.A.I.R.O. Anal. Numer.* **11 R3**, 341–354.

J. Fröhlich and K. Schneider (1995), 'An adaptive wavelet-vaguelette algorithm for the solution of nonlinear PDEs'. Preprint SC 95-28, ZIB.

J. Fröhlich and K. Schneider (1996), 'Numerical simulation of decaying turbulence in an adaptive wavelet basis'. Preprint, Universität Kaiserslautern, Fachbereich Chemie.

V. Girault and P.-A. Raviart (1986), *Finite Element Methods for Navier–Stokes Equations: Theory and Algorithms*, Series in Computational Mathematics, Springer.

R. Glowinski, T. W. Pan, R. O. Wells and X. Zhou (1996), 'Wavelet and finite element solutions for the Neumann problem using fictitious domains', *J. Comput. Phys.* **126**, 40–51.

R. Glowinski, A. Rieder, R. O. Wells and X. Zhou (1993), A wavelet multigrid preconditioner for Dirichlet boundary value problems in general domains, Technical Report 93-06, Rice University, Houston.

B. Gottschlich-Müller and S. Müller (1996), 'Multiscale concept for conservation laws'. IGPM-Report 128, RWTH Aachen.

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

M. Griebel (1994), *Multilevelmethoden als Iterationsverfahren über Erzeugendensystemen*, Teubner Skripten zur Numerik, Teubner, Stuttgart.

M. Griebel and P. Oswald (1995*a*), 'Remarks on the abstract theory of additive and multiplicative Schwarz algorithms', *Numer. Math.* **70**, 163–180.

M. Griebel and P. Oswald (1995*b*), 'Tensor product type subspace splittings and multilevel iterative methods for anisotropic problems', *Advances in Computational Mathematics* **4**, 171–206.

K. H. Gröchenich and W. R. Madych (1992), 'Haar bases and self-affine tilings of $\mathbb{R}^n$', *IEEE Trans. Inform. Theory* **38**, 556–568.

W. Hackbusch (1985), *Multigrid Methods and Applications*, Springer, New York.

W. Hackbusch (1989), 'The frequency-decomposition multigrid method I, Applications to anisotropic equations', *Numer. Math.* **56**, 229–245.

W. Hackbusch (1992), 'The frequency-decomposition multigrid method II, Convergence analysis based on the additive Schwarz method', *Numer. Math.* **63**, 433–453.

W. Hackbusch and Z. P. Nowak (1984), 'On the fast matrix multiplication in the boundary element method by panel clustering', *Numer. Math.* **54**, 463–491.

W. Hackbusch and S. Sauter (1993), 'On the efficient use of the Galerkin method to solve Fredholm integral equations', *Appl. Math.* **38**, 301–322.

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

S. Hildebrandt and E. Wienholtz (1964), 'Constructive proofs of representation theorems in separable Hilbert spaces', *Comm. Pure Appl. Math.* **17**, 369–373.

R. Hochmuth (1996), '*A posteriori* estimates and adaptive schemes for transmission problems'. IGPM Report 131, RWTH Aachen.

S. Jaffard (1992), 'Wavelet methods for fast resolution of elliptic equations', *SIAM J. Numer. Anal.* **29**, 965–986.

R. Q. Jia and C. A. Micchelli (1991), Using the refinement equation for the construction of pre-wavelets II: Powers of two, in *Curves and Surfaces* (P. J. Laurent, A. le Méhauté and L. L. Schumaker, eds), Academic Press, pp. 209–246.

H. Johnen and K. Scherer (1977), On the equivalence of the K-functional and moduli of continuity and some applications, in *Constructive Theory of Functions of Several Variables*, Vol. 571 of *Lecture Notes in Math.*, Springer, pp. 119–140.

P. Joly, Y. Maday and V. Perrier (1994), 'Towards a method for solving partial differential equations by using wavelet packet bases', *Comput. Meth. Appl. Mech. Engrg.* **116**, 301–307.

P. Joly, Y. Maday and V. Perrier (1997), A dynamical adaptive concept based on wavelet packet best bases: Application to convection diffusion partial differential equations, in *Multiscale Wavelet Methods for PDEs* (W. Dahmen, A. Kurdila and P. Oswald, eds), Academic Press. To appear.

A. Jouini (1992), 'Constructions de bases d'ondelettes sur les variétés'. Dissertation, Université Paris Sud–Centre d'Orsay.

A. Jouini and P. G. Lemarié-Rieusset (1992), Ondelettes sur un ouvert borné du plan. Preprint.

A. Jouini and P. G. Lemarié-Rieusset (1993), 'Analyses multirésolutions biorthogonales et applications', *Ann. Inst. Henri Poincaré, Anal. Non Lineaire* **10**, 453–476.

J. Ko, A. J. Kurdila and P. Oswald (1997), Scaling function and wavelet preconditioners for second order elliptic problems, in *Multiscale Wavelet Methods for PDEs* (W. Dahmen, A. Kurdila and P. Oswald, eds), Academic Press. To appear.

U. Kotyczka and P. Oswald (1996), Piecewise linear pre-wavelets of small support, in *Approximation Theory VIII, vol. 2* (C. K. Chui and L. L. Schumaker, eds), World Scientific, Singapore, pp. 235–242.

H. Kumano-go (1981), *Pseudo-Differential Operators*, MIT Press, Boston.

A. Kunoth (1994), Multilevel Preconditioning, PhD thesis, FU Berlin. Shaker, Aachen.

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

O. A. Ladyshenskaya (1969), *The Mathematical Theory of Viscous Incompressible Flow*, 2nd edn, Gordon and Breach, New York.

A. Latto, H. L. Resnikoff and E. Tenenbaum (1992), The evaluation of connection coefficients of compactly supported wavelets, in *Proceedings of French–USA Workshop on Wavelets and Turbulence* (Y. Maday, ed.), Springer.

S. Lazaar (1995), Algorithmes à base d'ondelettes et résolution numérique de problèmes elliptiques à coefficients variables, PhD thesis, Université d'Aix-Marseille I.

S. Lazaar, J. Liandrat and P. Tchamitchian (1994), 'Algorithme à base d'ondelettes pour la résolution numérique d'équations aux dérivées partielles à coefficients variables', *C.R. Acad. Sci., Série I* **319**, 1101–1107.

P. G. Lemarié (1984), *Algèbre d'opérateurs et semi-groupes de Poisson sur un espace de nature homogène*, Publ. Math. d'Orsay.

P. G. Lemarié-Rieusset (1992), 'Analyses, multi-résolutions nonorthogonales, commutation entre projecteurs et derivation et ondelettes vecteurs à divergence nulle', *Revista Mat. Iberoamericana* **8**, 221–236.

P. G. Lemarié-Rieusset (1994), 'Un théorème d'inexistence pour des ondelettes vecteurs à divergence nulle', *C. R. Acad. Sci. Paris I* **319**, 811–813.

J. Liandrat and P. Tchamitchian (1997), 'Elliptic operators, adaptivity and wavelets', *SIAM J. Numer. Anal.* To appear.

R. Lorentz and P. Oswald (1996), 'Constructing 'economic' Riesz bases for Sobolev spaces', GMD-Birlinghoven. Preprint.

R. Lorentz and P. Oswald (1997), Multilevel finite element Riesz bases in Sobolev spaces, in *DD9 Proceedings* (P. Bjorstad, M. Espedal and D. Keyes, eds), Wiley. To appear.

Y. Maday, V. Perrier and J. C. Ravel (1991), 'Adaptivité dynamique sur base d'ondelettes pour l'approximation d'équations aux dérivées partielles', *C. R. Acad. Sci. Paris Sér. I Math.* **312**, 405–410.

S. Mallat (1989), 'Multiresolution approximations and wavelet orthonormal bases of $L_2(\mathbb{R})$', *Trans. Amer. Math. Soc.* **315**, 69–87.

Y. Meyer (1990), *Ondelettes et opérateurs 1–3: Ondelettes*, Hermann, Paris.

Y. Meyer (1994). Private communication.

C. A. Micchelli and Y. Xu (1994), 'Using the matrix refinement equation for the construction of wavelets on invariant set', *Appl. Comput. Harm. Anal.* **1**, 391–401.

S. G. Michlin (1965), *Multidimensional Singular Integral Equations*, Pergamon Press, Oxford.

J. C. Nedelec (1982), 'Integral equations with non-integrable kernels', *Integral Equations Operator Theory* **5**, 562–572.

S. V. Nepomnyaschikh (1990), 'Fictitious components and subdomain alternating methods', *Sov. J. Numer. Anal. Math. Modelling* **5**, 53–68.

G. Nießen (1995), 'An explicit norm representation for the analysis of multilevel methods'. IGPM-Preprint 115, RWTH Aachen.

P. Oswald (1990), 'On function spaces related to finite element approximation theory', *Z. Anal. Anwendungen* **9**, 43–64.

P. Oswald (1992), On discrete norm estimates related to multilevel preconditioners in the finite element method, in *Constructive Theory of Functions (Proc. Int. Conf. Varna, 1991)* (K. G. Ivanov, P. Petrushev and B. Sendov, eds), Bulg. Acad. Sci., Sofia, pp. 203–214.

P. Oswald (1994), *Multilevel Finite Element Approximations*, Teubner Skripten zur Numerik, Teubner, Stuttgart.

P. Oswald (1997), Multilevel solvers for elliptic problems on domains, in *Multiscale Wavelet Methods for PDEs* (W. Dahmen, A. Kurdila and P. Oswald, eds), Academic Press. To appear.

J. Peetre (1978), *New Thoughts on Besov Spaces*, Duke University Press, Durham, NC.

V. Perrier (1996), Numerical schemes for 2D-Navier–Stokes equations using wavelet bases. Preprint.

P. J. Ponenti (1994), Algorithmes en ondelettes pour la résolution d'équations aux dérivées partielles, PhD thesis, Université Aix-Marseille I.

L. Quartapelle (1993), *Numerical Solution of the Incompressible Navier–Stokes Equations*, Vol. 113 of *International Series of Numerical Mathematics*, Birkhäuser.

V. Rokhlin (1985), 'Rapid solution of integral equations of classical potential theory', *J. Comput. Phys.* **60**, 187–207.

S. Sauter (1992), Über die effiziente Verwendung des Galerkinverfahrens zur Lösung Fredholmscher Integralgleichungen, PhD thesis, Universität Kiel.

R. Schneider (1995), 'Multiskalen- und Wavelet-Matrixkompression: Analysisbasierte Methoden zur effizienten Lösung großer vollbesetzter Gleichungssysteme'. Habilitationsschrift, Technische Hochschule Darmstadt.

P. Schröder and W. Sweldens (1995), Spherical wavelets: Efficiently representing functions on the sphere, in *Computer Graphics Proceedings (SIGGRAPH 95)*, ACM SIGGRAPH, pp. 161–172.

C. Schwab (1994), 'Variable order composite quadrature of singular and nearly singular integrals', *Computing* **53**, 173–194.

B. Sjögreen (1995), 'Numerical experiments with the multiresolution scheme for the compressible Euler equations', *J. Comput. Phys.* **117**, 251–261.

T. Sonar (1995), 'Multivariate Rekonstruktionsverfahren zur numerischen Berechnung hyperbolischer Erhaltungsgleichungen'. Habilitationsschrift, Technische Hochschule Darmstadt.

R. P. Stevenson (1995*a*), A robust hierarchical basis preconditioner on general meshes. Preprint, University of Nijmegen.

R. P. Stevenson (1995*b*), 'Robustness of the additive and multiplicative frequency decomposition multilevel method', *Computing* **54**, 331–346.

R. P. Stevenson (1996), 'The frequency decomposition multilevel method: A robust additive hierarchical basis preconditioner', *Math. Comput.* **65**, 983–997.

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

W. Sweldens (1997), 'The lifting scheme: A construction of second generation wavelets', *SIAM J. Math. Anal.* To appear.

W. Sweldens and R. Piessens (1994), 'Quadrature formulae and asymptotic error expansions for wavelet approximations of smooth functions', *SIAM J. Num. Anal.* **31**, 2140–2164.

P. Tchamitchian (1987), 'Biorthogonalité et théorie des opérateurs', *Revista Mat. Iberoamericana* **3**, 163–189.

P. Tchamitchian (1996), Wavelets, functions, and operators, in *Wavelets: Theory and Applications* (G. Erlebacher, M. Y. Hussaini and L. Jameson, eds), ICASE/LaRC Series in Computational Science and Engineering, Oxford University Press, pp. 83–181.

P. Tchamitchian (1997), 'Inversion de certains opérateurs elliptiques à coefficients variables', *SIAM J. Math. Anal.* To appear.

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

K. Urban (1995*a*), Multiskalenverfahren für das Stokes-Problem und angepaßte Wavelet-Basen, PhD thesis, RWTH Aachen. Aachener Beiträge zur Mathematik.

K. Urban (1995*b*), 'On divergence free wavelets', *Advances in Computational Mathematics* **4**, 51–82.

K. Urban (1995*c*), A wavelet-Galerkin algorithm for the driven cavity Stokes problem in two space dimension, in *Numerical Modelling in Continuum Mechanics* (M. Feistauer, R. Rannacher and K. Korzel, eds), Charles University, Prague, pp. 278–289.

K. Urban (1996), Using divergence free wavelets for the numerical solution of the Stokes problem, in *AMLI '96: Proceedings of the Conference on Algebraic Multilevel Iteration Methods with Applications* (O. Axelsson and B. Polman, eds), University of Nijmegen, pp. 261–278.

P. S. Vassilevski and J. Wang (1997*a*), 'Stabilizing the hierarchical basis by approximate wavelets, I: Theory', *Numer. Lin. Alg. Appl.* To appear.

P. S. Vassilevski and J. Wang (1997*b*), 'Stabilizing the hierarchical basis by approximate wavelets, II: Implementation', *SIAM J. Sci. Comput.* To appear.

R. Verfürth (1994), '*A posteriori* error estimation and adaptive mesh refinement techniques', *J. Comput. Appl. Math.* **50**, 67–83.

L. F. Villemoes (1993), Sobolev regularity of wavelets and stability of iterated filter banks, in *Progress in Wavelet Analysis and Applications* (Y. Meyer and S. Roques, eds), Editions Frontières, Paris, pp. 243–251.

T. von Petersdorff and C. Schwab (1997*a*), Fully discrete multiscale Galerkin BEM, in *Multiscale Wavelet Methods for PDEs* (W. Dahmen, A. Kurdila and P. Oswald, eds), Academic Press. To appear.

T. von Petersdorff and C. Schwab (1997*b*), 'Wavelet approximation for first kind integral equations on polygons', *Numer. Math.* To appear.

T. von Petersdorff, R. Schneider and C. Schwab (1997), 'Multiwavelets for second kind integral equations', *SIAM J. Numer. Anal.* To appear.

W. L. Wendland (1987), Strongly elliptic boundary integral equations, in *The State of the Art in Numerical Analysis* (A. Iserles and M. J. D. Powell, eds), Clarendon Press, Oxford, pp. 511–561.

M. V. Wickerhauser, M. Farge and E. Goirand (1997), Theoretical dimension and the complexity of simulated turbulence, in *Multiscale Wavelet Methods for*

PDEs (W. Dahmen, A. J. Kurdila and P. Oswald, eds), Academic Press. To appear.

J. Xu (1992), 'Iterative methods by space decomposition and subspace correction', *SIAM Review* **34**, 581–613.

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

H. Yserentant (1990), 'Two preconditioners based on the multi-level splitting of finite element spaces', *Numer. Math.* **58**, 163–184.

H. Yserentant (1993), Old and new proofs for multigrid algorithms, in *Acta Numerica*, Vol. 2, Cambridge University Press, pp. 285–326.

X. Zhang (1992), 'Multilevel Schwarz methods', *Numer. Math.* **63**, 521–539.