Adaptive Application of Operators in Standard Representation *

Arne Barinka, Stephan Dahlke, Wolfgang Dahmen

October 30, 2003

Dedicated to Charles A. Micchelli on the occasion of his 60th birthday

Abstract

Recently adaptive wavelet methods have been developed which can be shown to exhibit an asymptotically optimal accuracy/work balance for a wide class of variational problems including classical elliptic boundary value problems, boundary integral equations as well as certain classes of non coercive problems such as saddle point problems [8, 9, 12]. A core ingredient of these schemes is the approximate application of the involved operators in standard wavelet representation. Optimal computational complexity could be shown under the assumption that the entries in properly compressed standard representations are known or computable in average at unit cost. In this paper we propose concrete computational strategies and show under which circumstances this assumption is justified in the context of elliptic boundary value problems.

AMS subject classification: 41A25, 41A46, 65F99, 65N12, 65N55.
Key Words: Operator equations, quasi sparse matrices and vectors, best N-term approximation, fast matrix/vector multiplication.

1 Introduction

Recently a class of adaptive wavelet methods has been developed that exhibits asymptotically optimal *accuracy/work balance* for a wide class of variational problems including classical linear elliptic boundary value problems, boundary integral equations, certain classes of non coercive problems such as saddle point problems as well as nonlinear problems [8, 9, 10, 12]. By this we mean that the adaptive scheme produces for every target accuracy $\varepsilon > 0$ an approximate solution as a linear combination of $N(\varepsilon)$ adaptively chosen

*This work has been supported in part by the Deutsche Forschungsgemeinschaft SFB 401, the first and third author are supported in part by the European Community's Human Potential Programme under contract HPRN-CT-202-00286, (BRAEKING COMPLEXITY). The second author acknowledges the financial support provided through the European Union's Human Potential Programme, under contract HPRN-CT-2002-00285 (HASSIP), and through DFG, Grant Da 360/4–1.

wavelets realizing this accuracy in the energy norm where the number of degrees of freedom $N(\varepsilon)$ grows asymptotically at a lowest possible rate, namely that of the *best wavelet N-term approximation*, see Section 2.3 and [9] for details. Moreover, under certain assumptions concerning the computational accessibility of the wavelet representations of the involved operators, the corresponding computational work could be shown to stay essentially proportional to the number $N(\varepsilon)$ of *significant coefficients* in the underlying wavelet expansions.

A key ingredient of such schemes is the *adaptive application* of full *infinite dimensional* operators in standard wavelet representation. To explain this a brief recollection of the relevant setting is in order. The scope of feasible problems can be described in a (variationally defined) system of operator equations

$$\mathcal{L}u = f \tag{1.1}$$

which is *well-posed* in the sense that $\mathcal{L}$ is an *isomorphism* from some Hilbert space $\mathcal{H}$ into its dual $\mathcal{H}'$, see [9] for examples. $\mathcal{H}$ is usually a product space whose components are closed subspaces of Sobolev spaces. It has been shown in [9] that whenever these component spaces admit a wavelet characterization (i.e., weighted sequence norms of wavelet coefficient arrays are equivalent to the respective function space norms) this combined with the mapping property of $\mathcal{L}$ allows one to reformulate (1.1) equivalently as an infinite dimensional system

$$\mathbf{L}\mathbf{u} = \mathbf{f}, \tag{1.2}$$

where $\mathbf{u}$ is the unknown array of wavelet coefficients of the solution u and $\mathbf{L}$ is the *standard wavelet representation* of $\mathcal{L}$. Moreover, as a consequence of the well-posedness of (1.1) and the Riesz basis property of the wavelet basis, (1.2) can be shown to be now well-posed in the Euclidean metric, i.e., $\mathbf{L}$ is an *automorphism* on ℓ_2

$$\|\mathbf{L}\mathbf{v}\|_{\ell_2} \sim \|\mathbf{v}\|_{\ell_2}, \tag{1.3}$$

where here and below $A \sim B$ means that A can be bounded by a constant times B and vice versa independently of the parameters on which A, B may depend.

In the special case when $\mathcal{L}$ is $\mathcal{H}$-elliptic and, in particular, symmetric positive definite, i.e.

$$\langle v, \mathcal{L}w\rangle = a(v,w), \quad v, w \in \mathcal{H}, \quad \text{with} \;\; a(v,v) \sim \|v\|_{\mathcal{H}}^2, \tag{1.4}$$

$\mathbf{L}$ is also symmetric positive definite. Now (1.3) ensures that a simple iterative scheme like Richardson or gradient iterations

$$\mathbf{u}^{n+1} = \mathbf{u}^n - \alpha(\mathbf{L}\mathbf{u}^n - \mathbf{f}), \quad n = 0, 1, 2, \ldots, \tag{1.5}$$

for the full infinite dimensional problem (1.2) converges with a uniform error reduction per step.

One then mimics this ideal iteration by *approximately* applying $\mathbf{L}$ to the current *finitely supported* approximations to $\mathbf{u}$ with a certain stage dependent accuracy. The overall complexity of the whole scheme depends then on the complexity or better work/accuracy balance of this approximate (infinite) matrix/vector multiplication, i.e. the computational cost needed for attaining a given target accuracy. Note that the main distinction from more conventional approaches is the fact that one has, in principle, a representation of the full infinite dimensional operator independent of any fixed a priory mesh.

But even when $\mathbf{L}$ is not symmetric positive definite but only (1.3) holds, one can transform (1.2) into yet another equivalent form $\mathbf{Mp} = \mathbf{g}$ where $\mathbf{M}$ still satisfies (1.3) and is symmetric positive definite so that an iterative scheme applies successfully to this new form. This new form could simply be a least squares formulation, i.e. $\mathbf{M} = \mathbf{L}^T\mathbf{L}$, $\mathbf{g} = \mathbf{L}^T\mathbf{f}$, $\mathbf{p} = \mathbf{u}$, or, when dealing with saddle point problems, the Schur complement equation, see [9, 12] for details. The point is that each approximate application of $\mathbf{M}$ can then still be reduced to applications of the wavelet representations $\mathbf{L}$. Therefore the resulting scheme in turn can be realized with an asymptotically optimal work/accuracy balance whenever the wavelet representation $\mathbf{L}$ is *compressible* in a certain sense, [8, 9, 12]. Nevertheless, the optimal work count holds only under the assumption that the relevant entries in compressed versions of the wavelet representation are available or can be computed essentially at unit cost. While this assumption is justified for constant coefficient differential operators it is much less clear for other more realistic cases [3].

The main computational obstruction is the fact that these entries involve inner products of wavelets on possibly very different scales. When dealing with non-constant coefficients the naive approach of approximating the entries individually by quadrature will then increase the computational cost prohibitively. This usually motivates the use of the so called *nonstandard representation*, see e.g. [5]. However, variable coefficients would still require higher computational efforts for computing coarse scale quantities with sufficient accuracy. Moreover, the standard representation is smaller in size and has better compressibility, in the sense that more entries can be neglected without sacrificing accuracy, which is used in the above mentioned schemes in an essential way.

The objective of this paper is therefore to develop strategies for the efficient calculation and application of *standard representations* in the above adaptive context that still apply with optimal or good complexity under realistic and less stringent assumptions on the underlying operators. The core ingredient is a combination of the matrix/vector multiplication scheme from [8] with an efficient recovery scheme developed in [19] that was originally designed for treating nonlinear compositions of wavelet expansions. Since, by the above remarks, the essential difficulty arises already when dealing with scalar elliptic boundary value problems, we will focus in this paper only on elliptic PDEs with periodic boundary conditions to keep the exposition technically as simple as possible.

The outline of the paper is as follows. In Section 2 we formulate the periodic elliptic model problem problem, collect the necessary prerequisits for its formulation in wavelet coordinates and outline the main features of adaptive solution concepts. This provides the background for the objectives of the present studies and identifies the particular demands on the central ingredient, namely the adaptive application of standard wavelet representations of operators. In Section 3 we recall some basic facts about the matrix/vector multiplication scheme from [8] which the computational realization to be developed here will be based upon. Section 4 is concerned with the construction of such a realization. After outlining briefly the main strategy we specify the wavelet bases and take a closer look at the structure of the corresponding representation of the operator. Special emphasis is placed on a particular mechanism of generating new pairs of Riesz bases by integrating and differentiating a given pair of biorthogonal wavelet bases since this will play a central role for the error analysis of the adaptive evaluation scheme presented in Section 4.5. We conclude in Section 5 with some numerical experiments that support and illustrate the preceding findings.

2 Problem Formulation and Background

2.1 A Periodic Model Problem

We will be concerned with periodic elliptic boundary value problems on $\Omega = \Box := (0,1)^d$. Let C^∞_Π denote the space of infinitely often differentiable 1-periodic functions, i.e. $v(x) = v(x+k)$ for $k \in \mathbb{Z}^d$. For $s \geq 0$ we define H^s as the closure of C^∞_Π with respect to the usual s-order Sobolev norm, so that, in particular, $H^0 = L_2$. Moreover, the normed dual $(H^s)'$ of H^s will be denoted by H^{-s}.

We consider for a fixed $t \in \mathbb{N}$ the following variational problem: Find for a given $f \in H^{-t}$ a $u \in H^t$ such that

$$a(v,u) = \langle v, f\rangle_\Box, \quad v \in H^t, \tag{2.1.1}$$

with

$$a(v,w) := \sum_{|\alpha|,|\beta|=t} \int_\Box a_{\alpha,\beta}(x)\partial^\alpha v(x)\partial^\beta w(x) + v(x)w(x)dx. \tag{2.1.2}$$

Here $\mathbf{a}(x) := (a_{\alpha,\beta}(x))_{|\alpha|,|\beta|=t}$ is uniformly positive definite on Ω and $t \in \mathbb{N}$. We shall always assume that $\mathbf{a} \in L_\infty$ and is piecewise at least as smooth as the wavelets to be employed. Thus the operator $\mathcal{L}$ defined by $\langle v, \mathcal{L}w\rangle = a(v,w)$ for all $v, w \in H^t$ (where $\langle\cdot,\cdot\rangle$ denotes the standard duality pairing for H^t and H^{-t} induced by the L_2-inner product) is here a differential operator of order $2t$. It is well known that (2.1.1) possesses a unique solution in H^t for any $f \in H^{-t}$ due to the ellipticity of $a(\cdot,\cdot)$, i.e. there exist positive constants c_a, C_a such that

$$c_a\|v\|_{H^t} \leq a(v,v)^{1/2} \leq C_a\|v\|_{H^t}, \quad v \in H^t. \tag{2.1.3}$$

The periodic setting is chosen here primarily in order to minimize technicalities. More general domains and other boundary conditions could be treated as well. We shall add some remarks in this regard later.

2.2 Wavelet Bases and an Equivalent ℓ_2-Problem

Instead of taking the usual route of restricting (2.1.1) to some finite dimensional subspace of H^t to arrive at a finite dimensional linear problem we follow [9] and transform first (2.1.1) into an equivalent problem over an ℓ_2 sequence space. This can be accomplished with the aid of a suitable *wavelet basis* for H^t

$$\Psi = \{\psi_I : I \in \mathcal{I}\} \subset H^t,$$

that will be specified later in more detail. Let us remark at this point only that the indices I are comprised of several components. In particular, the component $k = k(I)$ specifies the spatial location of ψ_I and $j = j(I) =: |I|$ gives its scale in terms of powers of two. The meaning of "suitable" can be summarized as follows. We shall only employ compactly supported wavelets obtained here by periodizing tensor products of compactly supported wavelets on $I\!R$. Thus we have

$$\operatorname{diam}\operatorname{supp}\psi_I \sim 2^{-|I|}.$$

The collection Ψ will be assumed to *characterize* the space H^t in the following sense: There exists a diagonal matrix $\mathbf{D} = \mathbf{D}_\Psi$ such that $\mathbf{D}^{-1}\Psi := (D_I^{-1}\psi_I : I \in \mathcal{I})^T$ (which will be often viewed as an infinite vector of functions) is a *Riesz basis* for H^t, i.e., every $v \in H^t$ possesses a unique expansion $v = \sum_{I\in\mathcal{I}} v_I D_I^{-1}\psi_I =: \mathbf{v}^T\mathbf{D}^{-1}\Psi$ such that for some posistive finite constants c, C one has

$$c\|\mathbf{v}\|_{\ell_2(\mathcal{I})} \leq \|v\|_{H^t} \leq C\|\mathbf{v}\|_{\ell_2(\mathcal{I})}, \qquad v \in H^t. \tag{2.2.1}$$

Note that a duality argument yields then for c, C from (2.2.1)

$$C^{-1}\|\mathbf{D}^{-1}\langle\Psi, w\rangle\|_{\ell_2} \leq \|w\|_{H^{-t}} \leq c^{-1}\|\mathbf{D}^{-1}\langle\Psi, w\rangle\|_{\ell_2}, \quad w \in H^{-t}. \tag{2.2.2}$$

Usually a quantitatively good choice for the scaling matrix $\mathbf{D}$, that takes the coefficients in (2.1.2) into proper account, is $D_{I,I} := a(\psi_I, \psi_I)^{1/2}$. For moderately varying coefficients in $\mathbf{a}(x)$, $D_{I,I} := 2^{-t|I|}$ will do as well, which has the same asymptotic growth in $|I|$. We shall for simplicity mainly work with this simple choice and write $\mathbf{D} = \mathbf{D}^t$.

Finally, Ψ is assumed to have *cancellation properties* of order $\tilde{m}$ which here means that the unperiodized wavelets are orthogonal to all polynomials of order $\tilde{m}$.

The (infinite) matrix

$$\mathbf{L} := a(\mathbf{D}^{-1}\Psi, \mathbf{D}^{-1}\Psi) := \left(D_I^{-1}a(\psi_I, \psi_J)D_J^{-1}\right)_{I\in\mathcal{I}, J\in\mathcal{I}} = \mathbf{D}_\Psi^{-1}\langle\Psi, \mathcal{L}\Psi\rangle\mathbf{D}^{-1} \tag{2.2.3}$$

will be referred to as *standard (preconditioned) representation* of $\mathcal{L}$ with respect to $\mathbf{D}^{-1}\Psi$. The following fact is well known, see e.g. [13].

Theorem 2.1 $u = \mathbf{u}^T\mathbf{D}^{-1}\Psi$ *is the solution of (2.1.1) if and only if the coefficient array* $\mathbf{u}$ *solves*

$$\mathbf{L}\mathbf{u} = \mathbf{f} := \mathbf{D}^{-1}\langle\Psi, f\rangle. \tag{2.2.4}$$

Moreover, one has

$$c^2c_a\|\mathbf{v}\|_{\ell_2(\mathcal{I})} \leq \|\mathbf{L}\mathbf{v}\|_{\ell_2(\mathcal{I})} \leq C^2C_a\|\mathbf{v}\|_{\ell_2(\mathcal{I})}, \quad \mathbf{v} \in \ell_2(\mathcal{I}). \tag{2.2.5}$$

2.3 An Iteration Scheme and Best N-Term Approximation

Thus we have exactly the situation described in the introduction, i.e, the iteration

$$\mathbf{u}^{n+1} = \mathbf{u}^n - \alpha(\mathbf{L}\mathbf{u}^n - \mathbf{f}), \quad n = 0, 1, 2, \ldots, \tag{2.3.1}$$

converges with a fixed error reduction $\rho \leq \max\{1 - \alpha c^2c_a, \alpha C^2C_a - 1\}$ whenever $\alpha < 2/(C^2C_a)$.

The schemes developed and analyzed in [9, 12] aim at realizing (2.3.1) approximately by executing each step within some stage dependent accuracy tolerance. Aside from approximating the right hand side data, the central task lies in the approximate application of $\mathbf{L}$ to a given finitely supported input. Thus, one needs a scheme of the following type.

APPLY $[\eta, \mathbf{L}, \mathbf{v}] \rightarrow \mathbf{w}_\eta$ DETERMINES FOR ANY TARGET ACCURACY $\eta > 0$ AND ANY FINITELY SUPPORTED VECTOR $\mathbf{v}$ A FINITELY SUPPORTED VECTOR $\mathbf{w}_\eta$ SUCH THAT

$$\|\mathbf{L}\mathbf{v} - \mathbf{w}_\eta\|_{\ell_2(\mathcal{I})} \leq \eta. \tag{2.3.2}$$

The basic idea put forward in [9] is to realize (2.3.1) approximately by replacing at each step the exact evaluation of the residual $\mathbf{L}\mathbf{u}^n - \mathbf{f}$ by an approximation based essentially on APPLY. Given such a scheme APPLY, it has been shown in [9] how to choose the corresponding dynamic tolerances and how to interlace the perturbed iterations by *coarsening* steps so that for every target accuracy $\varepsilon > 0$ a finitely supported array $\mathbf{u}_\varepsilon$ is produced after finitely many steps that satisfies $\|\mathbf{u} - \mathbf{u}_\varepsilon\|_{\ell_2(\mathcal{I})} \leq \varepsilon$. Of course, by (2.2.1) this means that the corresponding finite wavelet expansion approximates u in the energy norm within accuracy $C\varepsilon$. Obviously, a *lower bound* for the computational cost to determine $\mathbf{u}_\varepsilon$ will depend, among other things, on its support $\Lambda_\varepsilon = \mathrm{supp}\,\mathbf{u}_\varepsilon$. The best one can hope for is therefore that not only the computational cost stays proportional to the number $\#\Lambda_\varepsilon$ of adaptively generated degrees of freedom but also that this number of degrees of freedom stays in some sense asymptotically as small as possible when $\varepsilon \rightarrow 0$.

To make this more precise, note that, for any $\mathbf{v} \in \ell_2(\mathcal{I})$, the best possible rate between degrees of freedom and accuracy in $\ell_2(\mathcal{I})$ is given by the *best N-term approximation* $\mathbf{v}_N$ which by definition satisfies

$$\sigma_N(\mathbf{v}) = \|\mathbf{v} - \mathbf{v}_N\|_{\ell_2(\mathcal{I})} = \min\{\|\mathbf{v} - \mathbf{w}\|_{\ell_2(\mathcal{I})} : \#\mathrm{supp}\,\mathbf{w} \leq N\},$$

and is simply obtained by retaining the N in modulus largest coefficients of $\mathbf{v}$. In the following we will not distinguish formally between a sequence $\mathbf{v} \in \ell_2$ and a finitely supported vector. Such a vector will always be tacitly assumed to be extended to an infinite sequence by placing zero coefficients with respect to a fixed ordering of the respective index set for ℓ_2. In particular, the application of an infinite matrix $\mathbf{C}$ to a finitely supported vector $\mathbf{v}$ is to be understood in this way.

Now for any $s > 0$ the *approximation class* $\mathcal{A}^s$ consists of those elements in $\ell_2(\mathcal{I})$ for which the (quasi-)norm

$$\|\mathbf{v}\|_{\mathcal{A}^s} := \sup_{N \in \mathbb{N}} N^s \sigma_N(\mathbf{v}) \tag{2.3.3}$$

stays finite. Hence, in particular, every finitely supported $\mathbf{v}$ belongs to $\mathcal{A}^s$ for all $s > 0$. Perturbed iteration schemes of the above type are said to be asymptotically optimal for $s < s^*$ if, whenever the exact solution $\mathbf{u}$ belongs to $\mathcal{A}^s$ for some $s < s^*$, then one has

$$\#\mathrm{supp}\,\Lambda_\varepsilon,\ \#\mathrm{flops}\ \lesssim\ \varepsilon^{-1/s}\|\mathbf{u}\|_{\mathcal{A}^s}^{1/s}, \quad \varepsilon \rightarrow 0. \tag{2.3.4}$$

It is shown in [9] that the whole perturbed iteration scheme is asymptotically optimal for some range s^* provided that the APPLY-scheme is asymptotically optimal in a similar sense for that range, namely

$$\#\mathrm{supp}\,\mathbf{w}_\eta,\ \#\mathrm{flops}\ \lesssim\ \mathrm{supp}\,\mathbf{v} + \varepsilon^{-1/\eta}\|\mathbf{v}\|_{\mathcal{A}^s}^{1/s}, \quad \|\mathbf{w}_\eta\|_{\mathcal{A}^s}\ \lesssim\ \|\mathbf{v}\|_{\mathcal{A}^s}. \tag{2.3.5}$$

This result puts into a clear focus the demands on the scheme APPLY that yield in the above sense optimal complexity. The main issue therefore is to develop concrete schemes APPLY with the above properties.

3 Fast Approximate Matrix/Vector Multiplication

We shall review next briefly a version of APPLY that has been developed in [8], see also [2] for first numerical realizations. A key role in this context is played by the following class of (infinite) *compressible matrices.*

Definition 3.1 *Let $s^* > 0$. A matrix $\mathbf{C}$ is called* s^*-compressible *if for each $0 < s < s^*$ and for some summable positive sequence $\{\alpha_k\}_{k\in\mathbb{N}_0}$ there exists for each $k \in \mathbb{N}_0$ a matrix $\mathbf{C}_k$ having at most $\alpha_k 2^k$ nonzero entries per row and column such that*

$$\|\mathbf{C} - \mathbf{C}_k\|_{\ell_2\to\ell_2} \le \alpha_k 2^{-sk}, \quad k \in \mathbb{N}_0. \tag{3.6}$$

The class of s^*-compressible matrices is denoted by $\mathcal{C}_{s^*}$.

We recall next from [8] how to approximate a matrix/vector product in an efficient way when the matrix is s^*-compressible. To this end, the following notation

$$\mathbf{v}_{[k]} := \mathbf{v}_{2^k} - \mathbf{v}_{2^{k-1}}, \quad \mathbf{v}_{[0]} := \mathbf{v}_{2^0} = \mathbf{v}_1, \quad \hat{\mathbf{v}}_{[k]} := \mathbf{v}_{[k]}/\|\mathbf{v}_{[k]}\|_{\ell_2}, \tag{3.7}$$

will be convenient. Thus $\mathbf{v}_{[k]}$ consists of the second half of the largest 2^k coefficients of $\mathbf{v}$. Note that the construction of $\mathbf{v}_{[k]}$ does not necessarily require to carry out the suboptimal operation of sorting the entries of $\mathbf{v}$ in the strict sense. Quasi-sorting based on so called *binary binning*, which re-groups $\mathbf{v}$ according to binary intervals, allows one to construct $\mathbf{v}_{[k]}$ in linear time, i.e. with $\mathcal{O}(\#\,\mathrm{supp}\,\mathbf{v})$ operations, while still providing asymptotical optimality, cf. [1], see also [23].

Defining now

$$\mathbf{w}_k := \mathbf{C}_k\mathbf{v}_{[0]} + \mathbf{C}_{k-1}\mathbf{v}_{[1]} + \cdots + \mathbf{C}_0\mathbf{v}_{[k]}, \tag{3.8}$$

one obviously has for any s^*-compressible matrix $\mathbf{C}$ and $s < s^*$ the error estimate

$$\begin{aligned}\|\mathbf{C}\mathbf{v} - \mathbf{w}_k\|_{\ell_2} \ &\le\ \|\mathbf{C}\|_{\ell_2\to\ell_2}\|\mathbf{v} - \mathbf{v}_{2^k}\|_{\ell_2} + \sum_{j=0}^{k}\|(\mathbf{C} - \mathbf{C}_{k-j})\hat{\mathbf{v}}_{[j]}\|_{\ell_2}\|\mathbf{v}_{[j]}\|_{\ell_2} \\ &\le\ \|\mathbf{C}\|_{\ell_2\to\ell_2}\|\mathbf{v} - \mathbf{v}_{2^k}\|_{\ell_2} + \sum_{j=0}^{k}\alpha_{k-j}2^{-s(k-j)}\|\mathbf{v}_{[j]}\|_{\ell_2}.\end{aligned} \tag{3.9}$$

Thus, given the a-priori assumption on $\mathbf{C}$ in terms of (3.6) and the a-posteriori information about the binary chunks $\|\mathbf{v}_{[j]}\|_{\ell_2}$, one can determine for any target accuracy $\eta > 0$ the smallest $K = K(\eta)$ such that the right hand side of (3.9) is smaller than η which always exists. For a finitely supported $\mathbf{v}$ this leads to the scheme APPLY defined and analyzed in [8]. Moreover, whenever $\mathbf{v} \in \mathcal{A}^s$ for some $s < s^*$, we immediately conclude from Definition 3.1 and (3.9) that

$$\|\mathbf{C}\mathbf{v} - \mathbf{w}_k\|_{\ell_2} \lesssim 2^{-ks}\|\mathbf{v}\|_{\mathcal{A}^s}, \tag{3.10}$$

which gives the following fact [8].

Remark 3.2 *Assume that* $\mathbf{C}$ *is* s^**-compressible. Then for any* $s < s^*$ *one has*

$$\|\mathbf{C}\|_{\mathcal{A}^s \to \mathcal{A}^s} < \infty, \tag{3.11}$$

i.e., $\mathbf{C}$ *is bounded on* $\mathcal{A}^s$.

Note that APPLY $[\eta, \mathbf{C}, \mathbf{v}]$ can be viewed as an adaptive evaluation of $\mathbf{C}\mathbf{v}$ using a-posteriori information provided by the $\mathbf{v}_{[j]}$.

We briefly summarize now the findings from [8] concerning the scheme APPLY.

Remark 3.3 *Assume that* $\mathbf{C}$ *is* s^**-compressible. Given any tolerance* $\eta > 0$ *and any vector* $\mathbf{v}$ *with finite support, the output* $\mathbf{w}_\eta := \mathbf{w}_{K(\eta)}$ *with support* $\mathcal{I}_\eta$ *of* **APPLY** $[\eta, \mathbf{C}, \mathbf{v}]$ *satisfies*

$$\|\mathbf{w}_\eta - \mathbf{C}\mathbf{v}\|_{\ell_2} \leq \eta. \tag{3.12}$$

Moreover, the following properties hold. There exists a positive constant C *depending only on* s *when* s *tends to infinity such that:*

(*i*) *The size of the support* $\mathcal{I}_\eta$ *is bounded by*

$$\#(\mathcal{I}_\eta) \leq C\|\mathbf{v}\|_{\mathcal{A}^s}^{1/s}\eta^{-1/s}. \tag{3.13}$$

(*ii*) *When the entries of* $\mathbf{C}$ *are known the number of* **arithmetic operations** *needed to compute* $\mathbf{w}_\eta$ *does not exceed* $C\left\{\eta^{-1/s}\|\mathbf{v}\|_{\mathcal{A}^s}^{1/s} + N\right\}$ *with* $N := \#\mathrm{supp}\,\mathbf{v}$.

(*iii*) *The number of* **quasi-sorting** *operations needed to compute* $\mathbf{w}_\eta$ *does not exceed* $\mathcal{O}(N)$.

(*iv*) *The output vector* $\mathbf{w}_\eta$ *satisfies*

$$\|\mathbf{w}_\eta\|_{\mathcal{A}^s} \leq C\|\mathbf{v}\|_{\mathcal{A}^s}. \tag{3.14}$$

Typical examples of compressible matrices are standard wavelet representations of a wide class of operators. Specifically, returning to (2.2.3), and taking $\mathbf{D} = (2^{t|I|}\delta_{I,J})_{I,J\in\mathcal{I}}$, the validity of the following estimates on the entries of $\mathbf{L}$ has been established in numerous settings (see e.g. [13, 4, 22, 24])

$$\mathbf{L}_{I,J} = 2^{-(|J|+|I|)t}|a(\psi_I, \psi_J)| \lesssim 2^{-||I|-|J||\sigma}(1 + d(I,J))^{-\beta}, \tag{3.15}$$

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

$$d(I,J) := 2^{\min(|I|,|J|)}\,\mathrm{dist}(\mathrm{supp}(\psi_I), \mathrm{supp}(\psi_J)). \tag{3.16}$$

Here σ depends on the *regularity* of the wavelets in Ψ while $\beta = d - 2t + \tilde{m}$ depends on the order $2t$ of $\mathcal{L}$ and on the order $\tilde{m}$ of the *cancellation properties* of Ψ, see e.g. [13]. Such decay estimates in turn have been shown in [8] to imply compressibility of $\mathbf{L}$.

Proposition 3.4 *For σ and β in* (3.15) *let*

$$s^* := \min\left\{\frac{\sigma}{d} - \frac{1}{2}, \frac{\beta}{d} - 1\right\}. \tag{3.17}$$

Then any matrix $\mathbf{L}$ *satisfying (3.15) belongs to* $\mathcal{C}_{s^*}$*.*

Moreover, one can identify explicitly the matrices $\mathbf{L}_k$ satisfying (3.6) for $s < s^*$ given by (3.17). In fact, given k, set

$$\tilde{a}_{I,J} := \begin{cases} a(\psi_I, \psi_J), & ||I| - |J|| \le k/d, \text{ and } d(I,J) \le 2^{k/d - ||I|-|J||}\, \gamma(||I| - |J||), \\ 0, & \text{else.} \end{cases} \tag{3.18}$$

Here γ denotes any summable sequence, e.g. $\gamma(n) := (1+n)^{-2/d}$.

Remark 3.5 *Note that once for a given target accuracy η the $K = K(\eta)$ according to (3.8) has been determined, one can reliably* **predict** *from the supports of the* $\mathbf{v}_{[i]}$ *and from the truncation rule (3.18) the support of the result of* APPLY $[\eta, \mathbf{L}, \mathbf{v}]$.

4 An Approximate Evaluation Scheme

4.1 The Basic Strategy

From the above findings we see that the scheme APPLY from [8], described above, does satisfy the optimality requirements (2.3.5) provided that the entries $a_{I,J} := a(\psi_I, \psi_J)$ are known or computable on the fly at unit cost. This is the case when the coefficients in (2.1.2) are constant and the wavelets are piecewise polynomial. But even then the computational cost for each entry is rather high, although uniformly bounded. Recalling (3.8), the fast approximate matrix/vector multiplication APPLY $[\eta, \mathbf{L}, \mathbf{v}]$ mentioned in Section 3, requires accessing those columns of the compressed versions $\mathbf{L}_{K-j}$ which are identified by the active indices in $\mathbf{v}_{[j]}$. This poses the following two difficulties.

(i) When (2.1.2) involves variable coefficients or when working with isoparametric bases, one would have to employ quadrature for the approximate computation of those $a_{I,J}$ needed in a current application of $\mathbf{L}$. So the relevant entries cannot be obtained exactly.

(ii) Since then one has to ensure sufficient accuracy relative to the target accuracy ε in the adaptive scheme, the quadrature cost for each $a_{I,J}$ when $|I|, |J|$ are small grows with decreasing ε, [3]. So keeping the cost proportional to the number of degrees of freedom is impossible along such lines.

As mentioned before the so called *nonstandard form* of operators in wavelet coordinates is often used to aleviate computational obstructions. In fact, the resulting matrix has the advantage that it involves only scalar products of basis functions on the *same level*. In the constant coefficient case and in the shift-invariant setting such entries can be evaluated at unit cost even for non-piecewise polynomial wavelets and scaling functions, see e.g.

[16]. However, it is no longer a representation of the operator in a strict sense, and one encounters similar problems as with the standard form when the differential equation involves variable coefficients, as then the quadrature effort for lower scale blocks increases.

But even in the case of constant coefficients, there are the following two difficulties. First, the size of these matrices relative to a fixed highest discretization level is larger than that for the standard representation and second, they have a weaker compressibility. Yet compressibility of the system matrix is crucial for the performance of the adaptive scheme, see [8, 9] Hence, lacking compressibility is a critical obstruction and therefore we will stay with the standard representation.

The obstruction (i) is easily dealt with in the context of (3.8) which will remain our principal guide line. Of course, the application of each summand $\mathbf{C}_{k-j}\mathbf{v}_{[j]}$ need not be exact either, as long as this further perturbation stays within appropriate tolerances. Since this will be used repeatedly it is worth formalizing it as follows. To this end, recall from (3.7) the notation $\hat{\mathbf{v}}_{[l]} = \mathbf{v}_{[l]}/\|\mathbf{v}_{[l]}\|_{\ell_2(\mathcal{J})}$.

Proposition 4.1 *Let* $\mathbf{C}$ *be an* s^**-compressible matrix and* $(\alpha_i)_{i\in\mathbb{N}_0}$ *the summable sequence from Definition 3.1. Suppose that for some fixed* $s' < s^*$ *one has a scheme*

$$\text{APPR}\,(\mathbf{C}, \mathbf{v}, j, k) \to \mathbf{d}_{k-j}$$

that produces for any $j \leq k, j, k \in \mathbb{N}_0$ *and any finitely supported vector* $\mathbf{v}$ *a finitely supported array* $\mathbf{d}_{k-j}$ *such that*

$$\|\mathbf{d}_{k-j} - \mathbf{C}_{k-j}\hat{\mathbf{v}}_{[j]}\|_{\ell_2(\mathcal{I})} \leq C\alpha_{k-j}2^{-s'(k-j)}, \tag{4.1.1}$$

where C *is independent of* k, j *and* $\mathbf{v}$*. Define*

$$(\mathbf{C} \circledast \mathbf{v})_k := \sum_{j=0}^{k} \|\mathbf{v}_{[j]}\|_{\ell_2(\mathcal{I})}\text{APPR}\,(\mathbf{C}, \mathbf{v}, j, k). \tag{4.1.2}$$

Then, whenever for any given $\eta > 0$ *the integer* $K = K(\eta)$ *renders the right hand side of (3.9) less than or equal to* η*, one also has*

$$\|\mathbf{C}\mathbf{v} - (\mathbf{C} \circledast \mathbf{v})_{K(\eta)}\|_{\ell_2(\mathcal{I})} \leq C'\eta, \tag{4.1.3}$$

where C' *depends only on* C *from (4.1.1). Moreover, for any* $s \leq s'$ *one has*

$$\|\mathbf{C}\mathbf{v} - (\mathbf{C} \circledast \mathbf{v})_k\|_{\ell_2(\mathcal{I})} \lesssim 2^{-sk}\|\mathbf{v}\|_{\mathcal{A}^s}. \tag{4.1.4}$$

The computational work $\mathrm{w}\,(\circledast)$ *needed to compute* $(\mathbf{C} \circledast \mathbf{v})_k$ *is bounded by*

$$\mathrm{w}\,(\circledast) \leq C\sum_{j\leq k} \mathrm{w}\,(\mathbf{C}, \mathbf{v}, j, k), \tag{4.1.5}$$

where $\mathrm{w}\,(\mathbf{C}, \mathbf{v}, j, k)$ *is the computational work required by* $\text{APPR}\,(\mathbf{C}, \mathbf{v}, j, k)$*.*

Proof: Given $\mathbf{v}$ and $k \in I\!\!N$, let $\mathbf{w}_k := \sum_{l=0}^{k} \mathbf{C}_{k-l}\mathbf{v}_{[l]}$ be according to (3.8). By definitions (3.7) and (4.1.2), we have

$$\begin{aligned}
\mathbf{C}\mathbf{v} - (\mathbf{C} \circledast \mathbf{v})_k &= \mathbf{C}\mathbf{v} - \mathbf{w}_k + \mathbf{w}_k - (\mathbf{C} \circledast \mathbf{v})_k \\
&= \mathbf{C}\mathbf{v} - \mathbf{w}_k + \sum_{j=0}^{k} \|\mathbf{v}_{[j]}\|_{\ell_2(\mathcal{I})} (\mathbf{C}_{k-j}\hat{\mathbf{v}}_{[j]} - \textsc{Appr}\,(\mathbf{C}, \mathbf{v}, j, k)).
\end{aligned}$$

Invoking (4.1.1) and recalling the definition of $K(\eta)$ based on the right hand side of (3.9), the estimate (4.1.3) is an immediate consequence (with $C' = 1 + C$). From this, we also conclude with (4.1.1) that

$$\begin{aligned}
\|\mathbf{C}\mathbf{v} - (\mathbf{C} \circledast \mathbf{v})_k\|_{\ell_2(\mathcal{I})} &\leq \|\mathbf{C}\mathbf{v} - \mathbf{w}_k\|_{\ell_2(\mathcal{I})} + C \sum_{j=0}^{k} \alpha_{k-j} 2^{-s(k-j)} \|\mathbf{v}_{[j]}\|_{\ell_2(\mathcal{I})} \\
&\leq \|\mathbf{C}\mathbf{v} - \mathbf{w}_k\|_{\ell_2(\mathcal{I})} + C \sum_{l=0}^{k} \alpha_l 2^{-s(l)} \|\mathbf{v} - \mathbf{v}_{2^{k-l}}\|_{\ell_2(\mathcal{I})} \\
&\leq C 2^{-sk} \|\mathbf{v}\|_{\mathcal{A}^s},
\end{aligned}$$

where we have used (3.10) in the last step. With (3.10), we confirm (4.1.4). The estimate (4.1.5) follows from (4.1.2). ∎

Obstruction (ii) is much more serious and calls for a completely different strategy. Our approach is motivated by earlier work concerning the evaluation of nonlinear compositions of wavelet expansions. It still exploits the better compressibility of standard representations but does not aim at computing the relevant entries $a_{I,J}$ individually. Instead one approximates with the desired precision the full result of applying $\mathbf{L}$ to a finitely supported input $\mathbf{v}$. This in turn is done by noting that for $v = \mathbf{v}^T\mathbf{D}^{-t}\Psi$, the array $\mathbf{L}\mathbf{v}$ consists just of the wavelet coefficients of the element $\mathcal{L}v \in H^{-t}$ with respect to the dual wavelet basis $\mathbf{D}\tilde{\Psi}$. So the idea would then be to approximate $\mathcal{L}v$ efficiently by a finite expansion of $\tilde{\Psi}$. More precisely, suppose that for $\mathcal{L}v = (\mathbf{D}\tilde{\Psi})^T(\mathbf{L}\mathbf{v})$ we have to construct an approximation $R(\mathcal{L}v) := (\mathbf{D}\tilde{\Psi})^T\mathbf{d}$. Since

$$\mathcal{L}v - R(\mathcal{L}v) = (\mathbf{D}\tilde{\Psi})^T \left(\langle \mathbf{D}^{-1}\Psi, \mathbf{L}\mathbf{v} \rangle - \mathbf{d} \right). \tag{4.1.6}$$

the norm equivalence (2.2.2) yields

$$\|\langle \mathbf{D}^{-1}\Psi, \mathbf{L}v \rangle - \mathbf{d}\|_{\ell_2(\mathcal{I})} \leq C\|\mathcal{L}v - R(\mathcal{L}v)\|_{H^{-t}}. \tag{4.1.7}$$

Hence, if $R(\mathcal{L}v)$ is a ε-accurate approximation to $\mathcal{L}v$ in the dual space H^{-t}, then its coefficients approximate the array $\mathbf{L}\mathbf{v}$ within accuracy $C\varepsilon$. The point is that the norm equivalences allow us to control the accuracy of the wavelet coefficients by the accuracy of an approximation on the "function side" which hopefully can be obtained at much lower cost than the individual computation of the desired coefficients. This in turn is helped by the fact that, according to Remark 3.5, we already know which coefficients will be needed.

Therefore, our strategy will be to develop inexpensive reconstructions $R(\mathcal{L}v)$ of $\mathcal{L}v$ by means of the recovery scheme from [19]. The output will be an array $\mathbf{d}$, namely the coefficients of $R(\mathcal{L}v)$ with respect to $\mathbf{D}\tilde{\Psi}$. However, one then faces the following further difficulty. The accuracy estimates for the approximation provided by the recovery scheme in [19] rely on *local estimates* which are natural in the L_2 metric. In the present contex the governing norm is that for H^{-t} which does not localize in a straightforward way. One possibility to overcome this is to exploit the fact that derivatives of wavelets are essentially still wavelets. This will allow us to reduce the whole reasoning to L_2-estimates. This is most conveniently exemplified in the periodic setting chosen here. In principle, the approach carries over, however, e.g. to composite wavelet bases on more general domains, see [17, 18]. We shall therefore collect next the relevant facts about periodic tensor product wavelet bases that are suitable for our purposes.

4.2 Biorthogonal Spline Wavelets

It is well-known how to derive *periodic wavelet bases* from wavelets on the real line. Here we focus on spline wavelets based on multiresolution spaces spanned by cardinal B-splines. To set notation we recall a few facts to be used below and refer the reader for more details e.g. to [11, 15].

For any fixed $m \in I\!N$ let $\phi = {}_m\phi$ denote the mth order B-spline with integer knots. It has been shown in [11] that for any $\tilde{m} \in I\!N$ such that $\tilde{m} \geq m$, $m + \tilde{m} \in 2I\!N$, there exists a compcatly supported function $\tilde{\phi} = {}_{m,\tilde{m}}\tilde{\phi} \in L_2(I\!R)$ which is *biorthogonal* to ϕ, i.e.,

$$\delta_{0,k} = \langle \phi, \tilde{\phi}(\cdot - k)\rangle_{I\!R} := \int_{I\!R} \phi(x)\overline{\tilde{\phi}(x-k)}dx = \delta_{0,k}, \quad k \in \mathbb{Z}.$$

Both functions satisfy the refinement relations

$$\phi(\cdot) = \sum_{k\in\mathbb{Z}} a_{0,k}\phi(2\cdot - k), \quad \tilde{\phi}(\cdot) = \sum_{k\in\mathbb{Z}} \tilde{a}_{0,k}\tilde{\phi}(2\cdot - k), \tag{4.2.1}$$

with *masks* $\mathbf{a}_0 = \{a_{0,k}\}_{k\in\mathbb{Z}}$ and $\tilde{\mathbf{a}}_0 = \{\tilde{a}_{0,k}\}_{k\in\mathbb{Z}}$ that are (necessarily) finitely supported. The functions

$$\psi(x) := \sum_{k\in\mathbb{Z}} a_{1,k}\phi(2x - k), \quad \tilde{\psi}(x) := \sum_{k\in\mathbb{Z}} \tilde{a}_{1,k}\tilde{\phi}(2x - k), \tag{4.2.2}$$

with $a_{1,k} := (-1)^k \tilde{a}_{0,m-k}$, $\tilde{a}_{1,k} := (-1)^k a_{0,m-k}$, are known to form corresponding *biorthogonal wavelets*, i.e.,

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

Wavelets on $I\!R^d$ are conveniently obtained by taking tensor products. It will be convenient to set $\psi_0 := \phi$, $\tilde{\psi}_0 := \tilde{\phi}$ and $\psi_1 := \psi$, $\tilde{\psi}_1 := \tilde{\psi}$. Setting for $e = (e_1, \ldots, e_d) \in E \setminus \{\mathbf{0}\} =: E^*$, $E := \{0,1\}^d$, and $x = (x_1, \ldots, x_d) \in I\!R^d$, we define

$$\psi_e(x) := \psi_{e_1}(x_1) \cdots \psi_{e_d}(x_d)$$

and analogously $\tilde{\psi}_e$ we obtain again two-scale relations

$$\psi_e(\cdot) = \sum_{k\in \mathbb{Z}^d} a_{e,k}\phi(2\cdot -k), \quad \tilde{\psi}_e(\cdot) = \sum_{k\in \mathbb{Z}^d} \tilde{a}_{e,k}\tilde{\phi}(2\cdot -k), \tag{4.2.3}$$

where the masks $\mathbf{a}_e, \tilde{\mathbf{a}}_e$ are tensor products of the univariate masks.

The next step is periodization. Given $\xi \in L_2(\mathbb{R}^d)$, define

$$\xi_{j,k}(\cdot) := 2^{dj/2}\sum_{l\in \mathbb{Z}^d} \xi(2^j(\cdot + l) - k), \quad \Xi_j := \{\xi_{j,k} : k \in \mathbb{Z}^d/2^j\mathbb{Z}^d\}$$

and denote the analogous collections generated by $\xi = \psi_e, \tilde{\psi}_e$ by $\Psi_{e,j}$, respectively $\tilde{\Psi}_{e,j}$. Viewing these collections as before also as vectors, the biorthogonality relations (that carry over to the periodic case) can be formulated as

$$\langle \Phi_j, \tilde{\Phi}_j\rangle_\Box = \mathbf{I}, \quad \langle \Psi_{e,j}, \tilde{\Psi}_{e',j'}\rangle_\Box = \delta_{e,e'}\delta_{j,j'}\mathbf{I}, \quad j \in \mathbb{Z}_+,\ e, e' \in E\setminus\{0\}, \tag{4.2.4}$$

where we have set $\Phi_j := \Psi_{\mathbf{0},j}$, $\tilde{\Phi}_j := \tilde{\Psi}_{\mathbf{0},j}$.

The collections

$$\Psi := \Phi_0 \bigcup_{j=0}^{\infty}\bigcup_{e\in E^*} \Psi_{e,j}, \quad \tilde{\Psi} := \tilde{\Phi}_0 \bigcup_{j=0}^{\infty}\bigcup_{e\in E^*} \tilde{\Psi}_{e,j}, \tag{4.2.5}$$

are known to form a pair of biorthogonal Riesz bases for the (one-periodic) space L_2, i.e.

$$\langle \Psi, \tilde{\Psi}\rangle_\Box = \mathbf{I}. \tag{4.2.6}$$

The corresponding index is here given by

$$\mathcal{I} = \mathcal{I}_\Psi = \{0\} \times \{(0,\dots,0)\} \bigcup_{j=0}^{\infty} \{j\} \times E^* \times (\mathbb{Z}^d/2^j\mathbb{Z}^d),$$

i.e., a typical index has here the form $I = (j, e, k)$. Accordingly, we will sometimes denote for a given index I by $k(I)$ or $e(I)$ the corresponding components.

Furthermore, the spaces

$$S_j := \operatorname{span}\Phi_j, \quad \tilde{S}_j := \operatorname{span}\tilde{\Phi}_j$$

(for the univariate generators ${}_m\phi, {}_{m,\tilde{m}}\tilde{\phi}$ as above) are mth, respectively $\tilde{m}$th order accurate, i.e.,

$$\inf_{v_j\in V_j} \|v - v_j\|_{L_2} \ \lesssim\ 2^{-js}\|v\|_{H^s}, \quad v \in H^s, \tag{4.2.7}$$

holds for $V_j = S_j, \tilde{S}_j$ with $s \le m, \tilde{m}$, respectively. Moreover, suppose that

$$\tilde{\gamma} := \sup\{s \in \mathbb{R} : {}_{m,\tilde{m}}\tilde{\phi} \in H^s(\mathbb{R})\},$$

it is well known that the following *norm equivalences* hold, see e.g. [13]

$$\begin{aligned} \|v\|_{H^s} &\sim \|\langle v, \tilde{\Psi}\rangle_\Box \mathbf{D}^s\|_{\ell_2}, \quad s \in (-\tilde{\gamma}, m - 1/2), \\ \|v\|_{H^s} &\sim \|\langle v, \Psi\rangle_\Box \mathbf{D}^s\|_{\ell_2}, \quad s \in (-m + 1/2, \tilde{\gamma}), \end{aligned} \tag{4.2.8}$$

where $\mathbf{D}^s$ is the infinite diagonal matrix $(\mathbf{D}^s)_{I,J} = 2^{s|I|}\delta_{I,J}$ and we have used $\langle v, \tilde{\Psi}\rangle_\Box \mathbf{D}^s =: \{D^s_I \langle v, \psi_I\rangle_\Box : I \in \mathcal{I}\}$ to denote the product of the row vector $\langle v, \Psi\rangle_\Box$ with the diagonal matrix $\mathbf{D}^s$. Thus (2.2.1) is valid for these bases.

4.3 Differentiating and Integrating Wavelets

We shall describe next another important feature of wavelets, namely that differentiating and integrating a given dual pair of the above type gives rise to another dual pair. This will serve as the main vehicle to reduce approximation estimates in H^{-t} to estimating errors in L_2.

To describe this, consider the *symbols*

$$a^e(z) := \sum_{k\in\mathbb{Z}^d} a_{e,k} z^k, \quad \text{and} \quad \tilde{a}^e(z) := \sum_{k\in\mathbb{Z}^d} \tilde{a}_{e,k} z^k, \ z \in \mathbb{C}^d,$$

associated with the two-scale relations (4.2.1) and (4.2.2). Following [21, 25], we define for $1 \le \nu \le d$ the modified symbols by

$$a^{e,\nu}(z) \quad := \quad a^e(z) \begin{cases} \frac{2}{1+z_\nu} & , \text{ if } e_\nu = 0 \\ \frac{1-z_\nu}{2} & , \text{ if } e_\nu = 1, \end{cases} \tag{4.3.1}$$

$$\tilde{a}^{e,\nu}(z) \quad := \quad a^e(z) \begin{cases} \frac{1+\overline{z_\nu}}{2} & , \text{ if } e_\nu = 0, \\ \frac{2}{1-\overline{z_\nu}} & , \text{ if } e_\nu = 1. \end{cases} \tag{4.3.2}$$

Moreover, let ∂_ν denote the weak partial derivative in a fixed direction $\mathbf{e}_\nu = (\delta_{1,\nu}\ , \ldots,\ \delta_{d,\nu}) \in \mathbb{R}^d$. Then the following holds [25].

Proposition 4.2 *The wavelets constructed in the previous section have the following properties.*

1. *For all $e \in E\backslash\{\mathbf{0}\}$ one has*

$$\partial_\nu \psi_e(x) = \begin{cases} \psi_e^{(\nu)}(x) - \psi_e^{(\nu)}(x - \mathbf{e}_\nu), & e_\nu = 0, \\ 4\psi_e^{(\nu)}(x), & e_\nu = 1, \end{cases} \tag{4.3.3}$$

where

$$\phi^{(\nu)}(x) = \prod_{i\neq\nu} \phi(x_i)\phi^*(x_\nu)$$

with $\phi^* = {}_{m-1}\phi$ being the B-spline of order $m-1$, and where $\psi_e^{(\nu)}$ is defined by (4.2.3) with respect to $\phi^{(\nu)}$ and $\mathbf{a}^{(e,\nu)}$.

2. *There exists a refinable function*

$$\tilde{\phi}^{(\nu)} = \prod_{i\neq\nu} \tilde{\phi}(x_i)\tilde{\phi}^*(x_\nu),$$

where $\tilde{\phi}^* = {}_{m-1,\tilde{m}+1}\tilde{\phi}$ is the dual scaling function for ϕ^*, such that the functions $\tilde{\psi}_e^{(\nu)}$ defined by (4.2.3) with respect to $\tilde{a}^{e,\nu}(z)$ satisfy

$$\partial_\nu \tilde{\psi}_e^{(\nu)}(x) = \begin{cases} \tilde{\psi}_e(x + \mathbf{e}_\nu) - \tilde{\psi}_e(x), & e_\nu = 0, \\ -4\tilde{\psi}_e(x), & e_\nu = 1. \end{cases} \tag{4.3.4}$$

3. Upon using periodization, the systems $\Psi^{(\nu)}, \tilde{\Psi}^{(\nu)}$ are also biorthogonal, i.e.,

$$\langle \psi^{(\nu)}_{j,e,k}, \tilde{\psi}^{(\nu)}_{j',e',k'} \rangle = \delta_{j,j'}\delta_{e,e'}\delta_{k,k'}, \tag{4.3.5}$$

and form Riesz bases for L_2.

It is clear that under suitable (regularity) assumptions on the initial wavelet bases analogs to the above result hold for higher order partial derivatives. To keep matters technically as simple as possible we confine the discussion to first order derivatives though which suffices for second order problems. It will be clear from the subsequent discussion how to treat $2n$th order boundary value problems.

For an index $I = (j, k, e) \in \mathcal{J}$, let us introduce next the notation

$$I_{\pm,\nu} := (j, k \pm \mathbf{e}_\nu, e).$$

The relations (4.3.3) and (4.3.4) can be reformulated for the periodic versions as follows

$$\partial_\nu \psi_I(x) = \begin{cases} 2^{|I|}(\psi^{(\nu)}_I(x) - \psi^{(\nu)}_{I_{+,\nu}}(x)), & e(I)_\nu = 0, \\ 2^{|I|+2}\psi^{(\nu)}_I(x), & e(I)_\nu = 1, \end{cases} \tag{4.3.6}$$

and

$$\partial_\nu \tilde{\psi}^{(\nu)}_I(x) = \begin{cases} 2^{|I|}(\tilde{\psi}_{I_{-,\nu}}(x) - \tilde{\psi}_I), & e(I)_\nu = 0, \\ -2^{|I|+2}\tilde{\psi}_I(x), & e(I)_\nu = 1. \end{cases} \tag{4.3.7}$$

4.4 The Structure of Stiffness Matrices

Suppose throughout the following that $\Psi, \tilde{\Psi}$ is the dual pair of biorthogonal one-periodic wavelet bases on $\Box$ given by (4.2.5) for a fixed dual pair of univariate generators ${}_m\phi, {}_{m,\tilde{m}}\tilde{\phi}$. Recall from (2.1.2), that $\mathbf{L}$ defined by (2.2.3), has for $t = 1$ the form

$$\mathbf{L} = \sum_{\nu,\mu=1}^{d} \mathbf{L}^{\mu,\nu}, \tag{4.4.1}$$

where for $I, I' \in \mathcal{J}$

$$(\mathbf{L}^{\nu,\mu})_{I,I'} \quad = \quad \int_\Box a_{\nu,\mu}(x) 2^{-|I|}\partial_\nu \psi_I(x) 2^{-|I'|}\partial_\mu \psi_{I'}(x)dx. \tag{4.4.2}$$

According the the discussion at the end of Section 4.1, our objective is to interpret $\mathbf{Lv}$ as a combination of arrays of dual wavelet coefficients with respect to an L_2-Riesz basis, in order to take advantage of the easy localization of approximation in L_2. To this end, introducing the shorthand notation

$$\diamondsuit_{e,\nu}\psi^{(\nu)}_I(x) = \begin{cases} \psi^{(\nu)}_I(x) - \psi^{(\nu)}_{I_{+,\nu}}(x), & \text{if} \quad e(I)_\nu = 0, \\ 4\psi^{(\nu)}_I(x), & \text{if} \quad e(I)_\nu = 1, \end{cases} \tag{4.4.3}$$

we can rewrite (4.4.2) in the following form

$$(\mathbf{L}^{\nu,\mu})_{I,I'} \quad = \quad \int_\Box \diamondsuit_{e,\nu}\psi^{(\nu)}_I(x) a_{\nu,\mu}(x) \diamondsuit_{e,\mu}\psi^{(\mu)}_{I'}(x)dx. \tag{4.4.4}$$

Thus, by Proposition 3.4, each matrix $\mathbf{L}^{\nu,\mu}$ is s^*-compressible where s^* depends on the regularity of the wavelets in $\Psi^{(\nu)}$.

We wish to approximate $\mathbf{L}\hat{\mathbf{v}}_{[l]}$, i.e. we aim at realizing approximate schemes APPR as in Proposition 4.1.1, where $\hat{\mathbf{v}}_{[l]}$ is the (normalized) second half of a 2^l-term approximation to $\mathbf{v}$, see (3.7). In order to do so, we will approximate first $\mathbf{L}^\nu\hat{\mathbf{v}}_{[l]}$ for fixed $\nu \in \{1, \dots, d\}$, where

$$\mathbf{L}^\nu := \sum_{\mu=1}^{d} \mathbf{L}^{\nu,\mu}.$$

Note that $\mathbf{L}^{\nu,\mu}$ can be split into

$$\mathbf{L}^{\nu,\mu} = \mathbf{B}^{\nu,\mu} - \mathbf{C}^{\nu,\mu}, \tag{4.4.5}$$

where

$$(\mathbf{C}^{\nu,\mu})_{I,I'} = \begin{cases} \langle \psi^{(\nu)}_{I_{+,\nu}}, a_{\nu,\mu}\Diamond_{e(I'),\mu}\psi^{(\mu)}_{I'}\rangle_\Box, & e(I)_\nu = 0, \\ 0, & e(I)_\nu = 1 \end{cases}$$

and

$$(\mathbf{B}^{\nu.\mu})_{I,I'} = \langle \xi_I, a_{\nu,\mu}\Diamond_{e(I'),\mu}\psi^{(\mu)}_{I'}\rangle_\Box,$$

where we made use of the rescaled basis functions

$$\xi_I := \begin{cases} \psi^{(\nu)}_I & \text{if} \quad e(I)_\nu = 0, \\ 4\psi^{(\nu)}_I & \text{if} \quad e(I)_\nu = 1, \end{cases} \tag{4.4.6}$$

which, by Proposition 4.2 3), form a Riesz basis $\Xi := \{\xi_I : I \in \mathcal{I}\}$ for L_2. The dual basis $\tilde{\Xi} = \{\tilde{\xi}_I : I \in \mathcal{I}\}$ is obtained from $\tilde{\Psi}^{(\nu)}$ by the corresponding inverse scaling. This will allow us to carry all approximation estimates over to L_2.

Because of (4.4.5), both matrices $\mathbf{B}^{\nu,\mu}$ and $\mathbf{C}^{\nu,\mu}$ are again s^*-compressible. Hence

$$\mathbf{B} = \mathbf{B}^\nu := \sum_{\mu=1}^{d} \mathbf{B}^{\nu,\mu} \quad \text{and} \quad \mathbf{C} = \mathbf{C}^\nu := \sum_{\mu=1}^{d} \mathbf{C}^{\nu,\mu}$$

are also s^*-compressible and we have $\mathbf{L}^\nu = \mathbf{B} - \mathbf{C}$.

It will be convenient, to define to (4.4.3) the analogous operator also for sequences. Therefore, for any array $\mathbf{d} = (d_I)_{I\in\mathcal{J}}$ let in agreement with (4.4.3)

$$(\Diamond_\nu \mathbf{d})_I = \begin{cases} d_I - d_{I_{+,\nu}}, & \text{if} \quad e(I)_\nu = 0, \\ 4d_I, & \text{if} \quad e(I)_\nu = 1. \end{cases} \tag{4.4.7}$$

Moreover, let us introduce the notation

$$\vartheta_l := \operatorname{supp} \mathbf{v}_{[l]},$$

for a given $\mathbf{v}$ and $l \in \mathbb{N}_0$. We are now prepared to interpret the array $\mathbf{L}^\nu\hat{\mathbf{v}}_{[l]}$ as dual wavelet coefficients of a function.

Remark 4.3 *The array* $\mathbf{L}^\nu \hat{\mathbf{v}}_{[l]}$ *can be written in the form*

$$\mathbf{L}^\nu \hat{\mathbf{v}}_{[l]} = \Diamond_\nu \mathbf{g}_{l,\nu}, \tag{4.4.8}$$

where

$$\mathbf{g}_{l,\nu} := \left\langle \Psi^{(\nu)}, g_l(\cdot|\mathbf{v}) \right\rangle_\square, \tag{4.4.9}$$

is the array of wavelet coefficients of the function

$$g_l^\nu(x|\mathbf{v}) := \sum_{I\in\vartheta_l} (\hat{\mathbf{v}}_{[l]})_I \left(\sum_{\mu=1}^{d} a_{\nu,\mu}(x) \Diamond_{e,\mu} \psi_I^{(\mu)}(x) \right), \tag{4.4.10}$$

with respect to the dual basis $\tilde{\Psi}^{(\nu)}$ *from Proposition 4.2.*

Proof: Indeed

$$\begin{aligned}
(\mathbf{L}^\nu \hat{\mathbf{v}}_{[l]})_I &= \sum_{I'\in\vartheta_l} (\hat{\mathbf{v}}_{[l]})_{I'} \sum_{\mu=1}^{d} \left\langle \Diamond_{e,\nu}\psi_I^{(\nu)}, a_{\nu,\mu}\Diamond_{e,\mu}\psi_{I'}^{(\mu)} \right\rangle \\
&= \left\langle \Diamond_{e,\nu}\psi_I^{(\nu)}, g_l^\nu(\cdot|\mathbf{v}) \right\rangle \\
&= \begin{cases} \langle \psi_I^{(\nu)}, g_l^\nu(\cdot|\mathbf{v})\rangle - \langle \psi_{I_{+,\nu}}^{(\nu)}, g_l^\nu(\cdot|\mathbf{v})\rangle, & \text{if} \quad e(I)_\nu = 0, \\ 4\langle \psi_I^{(\nu)}, g_l^\nu(\cdot|\mathbf{v})\rangle, & \text{if} \quad e(I)_\nu = 1 \end{cases} \\
&= \begin{cases} (\mathbf{g}_{l,\nu})_I - (\mathbf{g}_{l,\nu})_{I_{+,\nu}}, & \text{if} \quad e(I)_\nu = 0, \\ 4(\mathbf{g}_{l,\nu})_I & \text{if} \quad e(I)_\nu = 1 \end{cases} \\
&= (\Diamond_\nu \mathbf{g}_{l,\nu})_I.
\end{aligned}$$

■

In order to approximate the result of $\mathbf{L}^\nu \hat{\mathbf{v}}_{[l]}$, we will employ the application scheme (3.8) i.e., we wish to compute

$$\mathbf{w}_K := \sum_{l=0}^{K} \mathbf{L}_{K-l}^\nu \hat{\mathbf{v}}_{[l]}$$

where $K = K(\eta) \in \mathbb{N}_0$ is a chosen accuracy parameter. Recall from Proposition 3.4 and Remark 3.5 that the support of the columns of the compressed versions $\mathbf{L}_{K-l}^\nu$ of $\mathbf{L}^\nu$ are finite and explicitely known, cf. [8]. Therefore one can compute

$$\mathcal{I}_{l,K}^\nu := \bigcup_{k\in\vartheta_l} \operatorname{supp}\,(\mathbf{L}_{K-l}^\nu)_{\cdot,k}, \tag{4.4.11}$$

where $(\mathbf{L}_{K-l}^\nu)_{\cdot,k}$ is the kth column of $\mathbf{L}_{K-l}^\nu$. Recall that since $\mathbf{L}^\nu$ is s^*-compressible and $\#\vartheta_l \lesssim 2^l$, one has for $s < s^*$

$$\#\mathcal{I}_{l,K}^\nu \leq \alpha_{K-l} 2^K, \tag{4.4.12}$$

where $(\alpha_i)_{i\in\mathbb{N}_0}$ is some summable positive sequence, cf. Definition 3.6. Because of

$$\mathcal{I}_{l,K,+}^\nu := \mathcal{I}_{l,K} \cup \{I_{+,\nu} : I \in \mathcal{I}_{l,K}\}$$

one still has

$$\#\mathcal{I}^{\nu}_{l,K,+} \leq 2\alpha_{K-l}2^{K}. \tag{4.4.13}$$

For convenience, we will sometimes drop the superscripts in writing briefly $g_l(x|\mathbf{v}) = g_l^{\nu}(x|\mathbf{v})$, $\mathcal{I}_{l,K} = \mathcal{I}^{\nu}_{l,K}$ and $\mathcal{I}_{l,K,+} = \mathcal{I}^{\nu}_{l,K,+}$.

Remark 4.4 *In the following let $s' < s^*$ be fixed. With the notation introduced above, one has for any $s \leq s'$*

$$\|g_l(\cdot|\mathbf{v}) - \sum_{I\in\mathcal{I}_{l,K,+}} \langle g_l(\cdot|\mathbf{v}), \psi_I^{(\nu)}\rangle_\Box \tilde{\psi}_I^{(\nu)}\|_{L_2} \lesssim \alpha_{K-l}2^{-s(K-l)}, \tag{4.4.14}$$

where the constant depends only on s'.

Proof: Expanding $g_l(\cdot|\mathbf{v})$ in the above (Riesz-) basis Ξ will lead to

$$g_l(\cdot|\mathbf{v}) = \sum_{I\in\mathcal{I}} \langle g_l(\cdot|\mathbf{v}), \xi_I\rangle_\Box \tilde{\xi}_I,$$

where $\xi_I \in \Xi, \tilde{\xi} \in \tilde{\Xi}$ are defined in (4.4.6). Since Ξ is a Riesz basis for L_2 we know hat

$$\begin{aligned}\|g_l(\cdot|\mathbf{v}) - \sum_{I\in\mathcal{I}_{l,K,+}} \langle g_l(\cdot|\mathbf{v}), \xi_I\rangle_\Box \tilde{\xi}_I\|_{L_2} \quad &= \quad \|\sum_{I\in\mathcal{I}\setminus\mathcal{I}_{l,K,+}} \langle g_l(\cdot|\mathbf{v}), \xi_I\rangle_\Box \tilde{\xi}_I\|_{L_2} \\ &\lesssim \quad \|\left(\langle g_l(\cdot|\mathbf{v}), \xi_I\rangle_\Box\right)_{I\in\mathcal{I}\setminus\mathcal{I}_{l,K,+}}\|_{\ell_2(\mathcal{I})}.\end{aligned}$$

Now, let $\mathbf{B}_{\mathcal{I}_{l,K},\vartheta_l}$ denote the $\mathcal{I}_{l,K,+} \times \vartheta_l$ block of $\mathbf{B}$. By Definition (4.4.10) of $g_l(\cdot|\mathbf{v})$ we obtain

$$\begin{aligned}&\|\left(\langle g_l(\cdot|\mathbf{v}), \xi_I\rangle_\Box\right)_{I\in\mathcal{I}\setminus\mathcal{I}_{l,K,+}}\|_{\ell_2(\mathcal{I})} \\ &\quad = \|\Big(\langle \sum_{I'\in\vartheta_l} (\hat{\mathbf{v}}_{[l]})_{I'} [\sum_{\mu=1}^{d} a_{\nu,\mu}(x) \Diamond_{e,\mu} \psi_{I'}^{(\mu)}(x)], \xi_I\rangle_\Box\Big)_{I\in\mathcal{I}\setminus\mathcal{I}_{l,K,+}}\|_{\ell_2(\mathcal{I})} \\ &\quad \leq \|(\mathbf{B} - \mathbf{B}_{\mathcal{I}_{l,K},\vartheta_l})\hat{\mathbf{v}}_{[l]}\|_{\ell_2(\mathcal{I})} \lesssim \|\mathbf{B} - \mathbf{B}_{K-l}\|_{\ell_2\to\ell_2} \\ &\quad \lesssim \alpha_{K-l}2^{-s(K-l)}.\end{aligned} \tag{4.4.15}$$

This proves the claim. ■

It will not surprise that approximating $g_l(\cdot|\mathbf{v})$ well with terms in $\mathcal{I}_{l,K,+}$, yields good approximations to $\mathbf{L}^{\nu}\hat{\mathbf{v}}_{[l]}$.

Proposition 4.5 *Suppose, for $\eta > 0$, we can find a finitely supported array $\mathbf{d}$ such that*

$$\|\mathbf{d}^T\tilde{\Psi}^{(\nu)} - \sum_{I\in\mathcal{I}_{l,K,+}} \langle g_l(\cdot|\mathbf{v}), \psi_I^{(\nu)}\rangle_\Box \tilde{\psi}_I^{(\nu)}\|_{L_2} \leq \eta, \tag{4.4.16}$$

then one has

$$\|\mathbf{L}^{\nu}\hat{\mathbf{v}}_{[l]} - \Diamond_\nu \mathbf{d}\|_{\ell_2(\mathcal{I})} \lesssim \eta + \alpha_{K-l}2^{-s(K-l)}. \tag{4.4.17}$$

Proof: Let us split the operator $\Diamond_\nu$ as follows

$$\Diamond_\nu = \Diamond_{\nu,1} - \Diamond_{\nu,2},$$

with

$$(\Diamond_{\nu,1}\mathbf{d})_I = \begin{cases} d_I & \text{if} \quad e(I)_\nu = 0, \\ 4d_I & \text{if} \quad e(I)_\nu = 1, \end{cases} \qquad \text{and} \qquad (\Diamond_{\nu,2}\mathbf{d})_I = \begin{cases} d_{I_{+,\nu}} & \text{if} \quad e(I)_\nu = 0, \\ 0 & \text{if} \quad e(I)_\nu = 1. \end{cases}$$

Note that, due to the Riesz basis property, we have

$$\begin{aligned} \|\mathbf{d}^T\tilde{\Psi}^{(\nu)} - \sum_{I\in\mathcal{I}_{l,K,+}} \langle g_l(\cdot|\mathbf{v}), \psi_I^{(\nu)}\rangle_\Box \tilde{\psi}_I^{(\nu)}\|_{L_2} \quad &\sim \quad \|\mathbf{d} - \big(\langle g_l(\cdot|\mathbf{v}), \psi_I^{(\nu)}\rangle_\Box\big)_{I\in\mathcal{I}_{l,K,+}}\|_{\ell_2(\mathcal{I})} \\ = \|\mathbf{d}_{\mathcal{I}\setminus\mathcal{I}_{l,K,+}}\| \quad &+ \quad \|\big(\langle g_l(\cdot|\mathbf{v}), \psi_I^{(\nu)}\rangle_\Box - d_I\big)_{I\in\mathcal{I}_{l,K,+})}\|_{\ell_2(\mathcal{I})} \\ &\geq \quad \|\big(\langle g_l(\cdot|\mathbf{v}), \psi_I^{(\nu)}\rangle_\Box - d_I\big)_{I\in\mathcal{I}_{l,K,+})}\|_{\ell_2(\mathcal{I})}. \end{aligned}$$

Therefore, without loss of generality we can assume that $\mathbf{d}$ in (4.4.16) is supported in $\mathcal{I}_{l,K,+}$. Recall now that by definition, we have

$$\mathbf{B}\hat{\mathbf{v}}_{[l]} = \Big(\langle g_l(\cdot|\mathbf{v}), \xi_I\rangle_\Box\Big)_{I\in\mathcal{I}} = \Diamond_{\nu,1}\Big(\langle g_l(\cdot|\mathbf{v}), \psi_I^{(\nu)}\rangle_\Box\Big)_{I\in\mathcal{I}}.$$

This and the splitting $\mathbf{L}^\nu = \mathbf{B} - \mathbf{C}$ lead to

$$\begin{aligned} \|\mathbf{L}^\nu\hat{\mathbf{v}}_{[l]} - \Diamond_\nu\mathbf{d}\|_{\ell_2(\mathcal{I})} \quad &= \quad \|(\mathbf{B}-\mathbf{C})\hat{\mathbf{v}}_{[l]} - (\Diamond_{\nu,1} - \Diamond_{\nu,2})\mathbf{d}\|_{\ell_2(\mathcal{I})} \\ &\leq \quad \|\mathbf{B}\hat{\mathbf{v}}_{[l]} - \Big(\langle g_l(\cdot|\mathbf{v}), \xi_I\rangle_\Box\Big)_{I\in\mathcal{I}_{l,K,+}}\|_{\ell_2(\mathcal{I})} \\ &+ \quad \|\Diamond_{\nu,1}\Big(\langle g_l(\cdot|\mathbf{v}), \psi_I^{(\nu)}\rangle_\Box - d_I\Big)_{I\in\mathcal{I}_{l,K,+}}\|_{\ell_2(\mathcal{I})} + \|\Diamond_{\nu,2}\mathbf{d} - \mathbf{C}\hat{\mathbf{v}}_{[l]}\|_{\ell_2(\mathcal{I})} \\ &\leq \quad \|\Big(\langle g_l(\cdot|\mathbf{v}), \xi_I\rangle_\Box\Big)_{I\in\mathcal{I}\setminus\mathcal{I}_{l,K}}\|_{\ell_2(\mathcal{I})} \\ &+ \quad 4\|\Big(\langle g_l(\cdot|\mathbf{v}), \psi_I^{(\nu)}\rangle_\Box\Big)_{I\in\mathcal{I}_{l,K}} - \mathbf{d}\|_{\ell_2(\mathcal{I})} \\ &+ \quad \|\Big(d_{I_{+,\nu}} - \langle g_l(\cdot|\mathbf{v}), \psi_{I_{+,\nu}}^{(\nu)}\rangle_\Box\Big)_{I\in\mathcal{I}_{l,K}, e(I)_\nu=0}\|_{\ell_2(\mathcal{I})} \\ &\leq \quad \|\Big(\langle g_l(\cdot|\mathbf{v}), \xi_I\rangle_\Box\Big)_{I\in\mathcal{I}\setminus\mathcal{I}_{l,K}}\|_{\ell_2(\mathcal{I})} \\ &+ \quad 5\|\Big(\langle g_l(\cdot|\mathbf{v}), \psi_I^{(\nu)}\rangle_\Box\Big)_{I\in\mathcal{I}_{l,K}} - \mathbf{d}\|_{\ell_2(\mathcal{I})} \\ &\lesssim \quad \alpha_{K-l}2^{-s(K-l)} + \|\mathbf{d}^T\tilde{\Psi}^{(\nu)} - \sum_{I\in\mathcal{I}_{l,K,+}} \langle g_l(\cdot|\mathbf{v}), \psi_I^{(\nu)}\rangle_\Box \tilde{\psi}_I^{(\nu)}\|_{L_2}, \end{aligned}$$

where we have used (4.4.15) in the last step. Now, the assertion follows from (4.4.16). ∎

Thus, the problem of approximating $\mathbf{L}^\nu\hat{\mathbf{v}}_{[j]}$ and hence $\mathbf{L}\hat{\mathbf{v}}$ can be solved by approximating the functions $g_j^\nu(\cdot|\mathbf{v})$ in the dual systems $\tilde{\Psi}^{(\nu)}$ but with respect to the L_2-norm. This, in turn, will be done using the recovery scheme from [19] which we briefly recall next.

4.5 A Recovery Scheme

We adhere to the notion of the previous section. In view of Remark 4.4 and Proposition 4.5, our objective is to approximate the function $g_j(\cdot|\mathbf{v}) = g_j^\nu(\cdot|\mathbf{v})$ from (4.4.10) by linear combinations of the wavelets $\tilde{\psi}_I^{(\nu)}$. Moreover, we know beforehand that the significant coefficients will be found in $\mathcal{I}_{j,k,+}$ which is explicitly known beforehand from the compressibility properties of the wavelet representation, see Remark 3.5. Moreover, the cardinality of this set at most the order $\alpha_{k-j}2^k$, see (4.4.13). This is exactly the kind of task the schemes in [19, 1] have been developed for. We briefly recall here only the main features and refer the reader for details to [19, 1]. The core routine

$\textsc{Rec}\,[\eta, g, \tilde{\Psi}, \mathcal{J}] \to \mathbf{d}$

operates on a *prediction set* $\mathcal{J}$ (with certain properties to be specified later) identifying significant coefficients for g such that

$$\|g - \sum_{I\in\mathcal{J}} \langle g, \psi_I\rangle_\Box \tilde{\psi}_I\|_{L_2} \le \delta, \tag{4.5.1}$$

and determines for a given function g, a wavelet basis $\tilde{\Psi}$ and a finitely supported array $\mathbf{d}$ such that

$$\|\mathbf{d}^T\tilde{\Psi} - g\|_{L_2} \lesssim \eta + \delta, \tag{4.5.2}$$

where η is some additional perturbation depending on g and $\mathcal{J}$, caused e.g. by quadrature. Here, of course, the issue is to obtain the array $\mathbf{d}$ in an efficient way *without* approximating *directly* the wavelet coefficients $\langle g, \tilde{\psi}_I\rangle_\Box$ which are, in fact, the quantities we really want to recover.

Moreover, the objective is to ensure that, at least under certain circumstances, η will remain proportional to δ. This requires, in particular, the set $\mathcal{J}$ to be a *graded* tree in the following sense. The set of indices $I = (j, e, k), e \in E$ can be associated with the scaled cube $\Box_{(j,k)} := 2^{-j}(k + [0, 1]^d)$. We will sometimes write in slight abuse of notation $\Box_I$ although the scaled cube does not depend on $e(I)$. $\mathcal{J} \subset \mathcal{I}$ is called a *tree* if $I \in \mathcal{J}$ implies that every $I' \in \mathcal{J}$ which is associated with any dyadic cube containing $\Box_I$ (including $\Box_I$ itself). In particular, either all or no index associated with some dyadic cube belong to a tree. Those cubes which have no children in the tree are called *leaves*. The set $\partial\mathcal{J}$ of leaves forms a partition of $\Box$. The tree is called *graded* if any two neighboring leaves differ by at most one dyadic generation. The tree is *M-graded* if, when stepping down on the leaves by one dyadic level, one has to march through at least M cubes before reaching the next location where one can step further down, see [19].

The structure of the scheme $\textsc{Rec}$ can be described as follows.

1) Given a *prediction* set $\mathcal{J}$ identifying significant coefficients in the wavelet expansion $\mathbf{g}^T\tilde{\Psi}$ of g, expand the set $\mathcal{J}$ (if necessary) to an *M-graded tree $\mathcal{T}$*.

2) Compute an approximation

$$R_{\mathcal{T}}g = \sum_{l=l_0}^{L} \sum_{k\in\mathcal{T}^l} q_{l.k}\tilde{\phi}_{l,k}, \tag{4.5.3}$$

where the meaning of $\mathcal{T}^l$ will be explained below.

3) Apply a local multiscale transform that turns the local scaling function representation (4.5.3) into the wavelet representation

$$R_{\mathcal{T}} g = \sum_{I \in \mathcal{I}(\mathcal{T})} \hat{\mathbf{d}}_I \tilde{\psi}_I, \tag{4.5.4}$$

where the finite sparse index set $\mathcal{I}(\mathcal{T}) \subseteq \mathcal{J}$ will also be commented on below.

We will now comment briefly on these steps and refer for a detailed analysis to [19, 1]. As for step 1), since in the present situation $g = g_l^\nu(\cdot|\mathbf{v})$, the role of $\mathcal{J}$ will be played by the sets $\mathcal{I}_{j,k}$ which usually already have tree structure due to the properties of $\mathbf{L}$. An additional refinement may be necessary to realize M-gradedness, where M will depend on the supports of $\tilde{\psi}^{(\nu)}$ and hence of the orders $m, \tilde{m}$ of the univariate initial scaling functions. Thus one still has

$$\#\mathcal{T} \lesssim \mathcal{J} \tag{4.5.5}$$

uniformly in the target accuracy δ.

Step 2) is motivated by the telescoping expansion of g

$$g = \tilde{Q}_{j_0} g + \sum_{j=j_0}^{\infty} (\tilde{Q}_{j+1} - \tilde{Q}_j) g, \tag{4.5.6}$$

in terms of the canonical projectors

$$\tilde{Q}_j g := \sum_{k \in Z^d/2^j \mathbb{Z}^d} \langle g, \phi_{j,k} \rangle_\Box \tilde{\phi}_{j,k},$$

with suitable pairs of dual generators $\phi, \tilde{\phi}$. In our present situation the modified pairs $\phi^{(\nu)}, \tilde{\phi}^{(\nu)}$ will be used. Clearly each dyadic difference can be written as $(\tilde{Q}_{j+1} - \tilde{Q}_j)g = \sum_{|I|=j} \langle g, \psi_I \rangle_\Box \tilde{\psi}_I$. Confining the indices I on level j to those carrying significant coefficients, i.e., to those of $\mathcal{J}$ or more generously in $\mathcal{T}$, the telescoping expansion ends locally on regions determined by the leaves. On the other hand, M-gradedness ensures that locally at most two different levels interact, when using the scaling function representation of the $\tilde{Q}_j$. The key is then to approximate the inner products $\langle g, \phi_{j,k} \rangle_\Box$ by a quadrature rule of fixed polynomial exactness $\geq m$, say *only* on regions corresponding to the leaves in $\mathcal{T}$. For level l these indices are identified by the sets $\mathcal{T}^l$ in (4.5.3), again see [19, 1] for details. Assuming the availability of an error estimate for the quadrature, those quadrature domains for which the error bound is maximal are successively subdivided until the the ℓ_2-norm of the array of error bounds is less than or equal to the target accuracy η.

We know that we have only to look for coefficients in the set $\mathcal{J}$. However, this prediction is often pessimistic and a final coarsening step could very well produce a smaller set $\mathcal{I}(\mathcal{T})$ which already warrants the desired accuracy. We wish to stress that the precise realization of Rec is not essential in this context. More efficient specifications which, for instance, in contrast to [19] avoid the generation of the full expanded tree $\mathcal{T}$ working in a single top-to-bottom sweep with incorporated coarsening is given in [1].

The scheme Rec (or any of its specifications) provides a realization of the scheme Appr in Proposition 4.1 which is to generate approximations to the terms $\mathbf{L}\mathbf{v}_{[j]}$. Our version of the scheme Appr $(\mathbf{C}, \mathbf{v}, j, k)$ in the current setting reads as follows.

Appr $(\mathbf{L}, \mathbf{v}, j, k) \to \mathbf{d}_{k-j}$

(i) For each ν determine the sets $\mathcal{I}_{j,k} = \mathcal{I}^{\nu}_{j,k}$ according to (4.4.1), (4.4.11) as well as the support ϑ_j of $\mathbf{v}_{[j]}$, $j \le k$ along with the function $g_j(\cdot|\mathbf{v}) = g^{\nu}_j(\cdot,|\mathbf{v})$ from (4.4.10).

(ii) Set $\eta_{j,k} = \alpha_{k-j} 2^{-s'(k-j)}$, with $s' < s^*$ from Remark 4.4 fixed, and apply

$$\textsc{Rec}\,[\eta_{j,k}, g^{\nu}_j, \tilde{\Psi}^{(\nu)}, \mathcal{I}^{\nu}_{j,k}] \to \mathbf{d}^{\nu}.$$

(iii) Set

$$\mathbf{d}_{k-j} := \sum_{\nu} \Diamond_{\nu} \mathbf{d}^{\nu}.$$

A word on step (ii) is in order. Recall from (4.4.10) that

$$g^{\nu}_l(x|\mathbf{v}) := \sum_{I \in \vartheta_l} (\hat{\mathbf{v}}_{[l]})_I \left(\sum_{\mu=1}^{d} a_{\nu,\mu}(x) \Diamond_{e,\mu} \psi^{(\mu)}_I(x) \right).$$

For each pair of indices (μ, ν) we can transform $\sum_{I\in\vartheta_l} (\hat{\mathbf{v}}_{[l]})_I \Diamond_{e,\mu} \psi^{(\mu)}_I(x)$ into local scaling function representation at $\mathcal{O}(\#\mathcal{T}(\vartheta_l))$ cost, where $\mathcal{T}(\vartheta_l)$ is the smallest tree containing ϑ_l with those elements of $\mathbf{v}_{[l]}$ serving as roots which have no ancestor in ϑ_l. Thus, we have to approximate inner products of the form $\langle a_{\nu,\mu} \phi^{\mu}_{j,k}, \phi^{\nu}_{j',k'} \rangle_{\Box}$ by means of a sufficiently accurate quadrature. Here is a possible strategy for adapting the accuracy of the quadrature routine. Suppose first that $a_{\nu,\mu}$ is smooth on the support of $\phi^{\mu}_{j,k}$, say. We can expand locally $a_{\nu,\mu}$ into orthogonal polynomials up some initial order $\ge m$ on the smallest dyadic cube $\Box'$ containing $\operatorname{supp} \phi^{\mu}_{j,k}$. Using the schemes from [16], the resulting expression $\langle P\phi^{\mu}_{j,k}, \phi^{\mu}_{j',k'} \rangle_{\Box}$ can be computed exactly at $\mathcal{O}(1)$ cost. One can successively raise the order of the polynomial expansion until the target accuracy is met in the sense explained above. If $a_{\nu,\mu}$ (which is explicitly given) is not smooth on $\operatorname{supp} \phi^{\mu}_{j,k}$, we can approximate it locally by a tensor product B-spline based quasi-interpolant, arriving at an integral that can again be computed exactly at unit cost with the aid of the schemes from [16] since one always obtains integrals of products of refinable functions. In this case accuracy is increased by subdividing the mesh for the B-spline approximation. Of course, the cost increases with the number of such subdivisions. As an immediate consequence we can state the following fact.

Remark 4.6 *When the coefficients $a_{\nu,\mu}(x)$ are constant or piecewise polynomial on a fixed partition into dyadic cubes the scheme is exact, i.e.,* $\mathbf{L}_{k-j} \hat{\mathbf{v}}_{[j]} = \mathbf{d}_{k-j}$.

In general, we can state the following immediate consequence of Proposition 4.5.

Remark 4.7 *The output* $\mathbf{d}_{k-j}$ *of* Appr *satisfies*

$$\|\mathbf{L}_{k-j}\hat{\mathbf{v}}_{[j]} - \mathbf{d}_{k-j}\|_{\ell_2(\mathcal{I})} \lesssim \alpha_{k-j}2^{-s(k-j)}, \tag{4.5.7}$$

so that, in particular, Proposition 4.1 applies.

We have now collected all ingredients of the

Main Algorithm: Given a finitely supported $\mathbf{v} \in \ell_2(\mathcal{I})$ and a target accuracy η, the scheme

Mult $[\eta, \mathbf{L}, \mathbf{v}] \to \mathbf{w}_\eta$ produces a finitely supported vector $\mathbf{w}_\eta$ such that

$$\|\mathbf{L}\mathbf{v} - \mathbf{w}_\eta\|_{\ell_2(\mathcal{I})} \leq C\eta, \tag{4.5.8}$$

for some constant independent of η as follows:

- Determine K such that with the a-priori estimates of the form (3.6) for the compressible matrix $\mathbf{L}$ the corresponding upper bound of (3.9) is at most $\eta/2$, see [8].
- Apply Appr $(\mathbf{L}, \mathbf{v}, j, K) \to \mathbf{d}_{K-j}$ and set

$$\hat{\mathbf{w}}_\eta = \sum_{j=0}^{K} \|\mathbf{v}_{[j]}\|_{\ell_2(\mathcal{I})}\mathbf{d}_{K-j}.$$

- The resulting finitely supported coefficient vector $\hat{\mathbf{w}}_\eta$ is subjected to a coarsening step, i.e., one determines $\mathbf{w}_\eta$ with smallest possible support (by discarding the smallest entries of $\hat{\mathbf{w}}_\eta$) such that

$$\|\hat{\mathbf{w}}_\eta - \mathbf{w}_\eta\|_{\ell_2(\mathcal{I})} \leq \eta. \tag{4.5.9}$$

Note that the estimate (4.5.8) follows from Proposition 4.1 combined with Remark 4.7.

4.6 A Possible Alternative

The following alternative to the above scheme may come to mind. The union $\mathcal{J}_k$ of the sets $\mathcal{I}^\nu_{j,k}$ for $j \leq k$ from (4.4.11) forms a set of cardinality $\lesssim 2^k$ which, in view of (3.10), is the optimal bound when $\mathbf{v} \in \mathcal{A}^s$. Therefore, one could define also

$$\text{Mult}\,[\eta, \mathbf{L}, \mathbf{v}] := \sum_{\nu=1}^{d} \text{Rec}\,[\eta, g^\nu, \tilde{\Psi}, \mathcal{J}], \tag{4.6.1}$$

where

$$g^\nu(x|\mathbf{v}) := \sum_{I\in\mathcal{I}} v_I \left(\sum_{\mu=1}^{d} a_{\nu,\mu}(x)\Diamond_{e,\mu}\psi_I^{(\mu)}(x)\right), \tag{4.6.2}$$

so that now for the array $\mathbf{g}_\nu := \left\langle \Psi^{(\nu)}, g^\nu(\cdot|\mathbf{v}) \right\rangle_\Box$, one has $\mathbf{L}^\nu \mathbf{v} = \lozenge_\nu \mathbf{g}_\nu$, see (4.4.8). Thus the recovery scheme applies only once to the full image of $\mathbf{L}^\nu$. Whenever the support of $\mathbf{v}$ has essentially already tree structure, this persists to be the case for the prediction set $\mathcal{J} = \mathcal{J}_K$ for the accuracy parameter $K = K(\eta)$ so that the cardinality of the expanded tree stays controlled by the size of support of $\mathbf{v}$ and the computational work required by REC stays proportional to $\#\mathcal{J}$ provided that the quadrature relative to the partition induced by the expanded tree already provides sufficient accuracy. In this case this version would have the desired asymptotically optimal computational complexity. We shall return to this point in somewhat more detail in the following section.

The reason for discussing the somewhat more involved version presented in the previous section is the following: The prediction sets $\mathcal{I}^\nu_{j,k}$ used in step (ii) of APPR are often too pessimistic. So the output of each application of APPR to the typically smaller problems is conveniently coarsened before assembling the approximation to $\mathbf{L}\mathbf{v}$. One may therefore expect a quantitatively better performance with regard to avoiding unnecessarily large prediction sets.

4.7 Remarks on Computational Complexity

It remains to discuss the computational work required by the scheme MULT. The following comments aim only at outlining an assessment of the *asympotics* of the computational complexity. To this end, note first that the comments following the description of REC combined with Proposition 4.1 can be summarized as follows.

Remark 4.8 *If $\#\mathcal{T}(\vartheta_l) \lesssim \#\vartheta_l$ and if only a finite uniformly bounded number of quadrature subdivisions are needed to meet the target accuracies $\eta_{j,k}$ in* REC*, the total computational complexity remains proportional to $\#\mathrm{supp}\, \hat{\mathbf{w}}_\eta$ as $\eta \to 0$.*

We shall indicate next circumstances under which the assumptions of Remark 4.8 can be expected to hold. This requires recalling a few facts from [19]. The tree $\mathcal{J}$ is called *δ-balanced* (with respect to g) if the *local errors* satisfy for each $J \in \partial\mathcal{J}$

$$\|g - \sum_{I \in \mathcal{J}} \langle g, \psi_I \rangle_\Box \tilde{\psi}_I \|_{L_2(\Box_J)} \le \delta \tag{4.7.1}$$

while for any leave of some subtree $\mathcal{J}'$ with $\#\mathcal{J}' \ge c\#\mathcal{J}$ for some fixed $c > 0$, this estimate fails to hold. The significance of this latter fact can be explained as follows. When g belongs to a Besov space $B^r_q(L_p)$ with

$$\frac{1}{p} < \frac{r}{d} + \frac{1}{2} \tag{4.7.2}$$

then balancing local errors is known to produce asymptotically best possible approximation errors in the following sense. Let Π_m denote the space of polynomials of degree at most m in d variables and let $m \ge s$.

Remark 4.9 *It has been shown that, even when instead of $\|g - \sum_{I\in\mathcal{J}} \langle g, \psi_I \rangle_\Box \tilde{\psi}_I \|_{L_2(\Box_J)}$ one balances some local* upper bound *of these local errors such as $\inf_{P\in\Pi_m} \|g - P\|_{L_2(\Box^*_J)}$*

for a somewhat larger domain $\Box_J^* \supset \Box_J$ (of comparable diameter), one can show the following fact [19], see also [7]. For every $N \in \mathbb{N}$ and fixed $M \in \mathbb{N}$ there exists an M-graded tree $\mathcal{T}_N$ such that $\#\mathcal{T}_N \lesssim N$ and

$$\|g - \sum_{I \in \mathcal{T}_N} \langle g, \psi_I \rangle_\Box \tilde{\psi}_I\|_{L_2} \lesssim N^{-r/d} \|g\|_{B_q^r(L_p)}, \quad N \to \infty. \tag{4.7.3}$$

It is known that the rate of best N-term approximation in L_2 over the unit ball in $B_q^r(L_p)$ for r, p related by (4.7.2) is indeed $N^{-r/d}$. Moreover, it is asymptotically achieved even when constraining the choice of terms to form trees and in fact graded trees (which are automatically produced through the somewhat larger sets $\Box_J^*$). Note also that, once a tree is graded it can be made M-graded for any fixed M by dyadic subdivision while preserving the asymptotic growth rate. This accounts for the first assumption in Remark 4.8 when $\mathbf{v}$ is an approximation to the wavelet coefficients of an element in $B_q^{sd}(L_p)$ for some $s \le s'$.

Moreover, the quadrature causes an error that can be bounded by local best polynomial approximation of the type mentioned in Remark 4.9. Thus when the initial set $\mathcal{J}$ is balanced one expects error bounds of the form

$$\|g - R_{\mathcal{T}} g\|_{L_2} \lesssim (\#\mathcal{J})^{-s} \|g\|_{B_q^{sd}(L_p)}, \tag{4.7.4}$$

for p related to s by $p^{-1} < s + 1/2$.

Let us point out next why these concepts are relevant in the present context. We shall concentrate here only on the conceptually simpler version from Section 4.6 because the effect of the quadrature should overall be the same. In the current situation the role of g will be played by the function $g^\nu(\cdot|\mathbf{v})$ from (4.6.2) and $\tilde{\Psi}$ is here $\tilde{\Psi}^{(\nu)}$. Recall that the matrices $\mathbf{L}^\nu$ are s^*-compressible for some $s^* > 0$ depending on the order of smoothness and vanishing moments of the wavelets Ψ. In the following we shall always assume as above that $s \le s'$ where s' is a fixed number strictly less than s^*. Recall that any finitely supported $\mathbf{v}$ belongs to $\mathcal{A}^s$ for any $s > 0$. Thus, setting $\mathbf{w}_\eta = \text{MULT}[\eta, \mathbf{L}, \mathbf{v}]$ and $\mathbf{z}_\eta := \text{APPLY}[\eta, \mathbf{L}, \mathbf{v}]$, we have by (3.10) for $K = K(\eta)$ as above

$$\|\mathbf{L}\mathbf{v} - \mathbf{w}_\eta\|_{\ell_2(\mathcal{I})} \lesssim 2^{-Ks} \|\mathbf{v}\|_{\mathcal{A}^s} + \|\mathbf{z}_\eta - \mathbf{w}_\eta\|_{\ell_2(\mathcal{I})}, \tag{4.7.5}$$

where the second summand accounts for the quadrature error caused by REC.

Now recall from [20] that the space $\mathcal{A}^s(\mathbf{D}^{-1}\Psi) := \{\mathbf{v}^T \mathbf{D}^{-1}\Psi : \mathbf{v} \in \mathcal{A}^s\} \supset B_\tau^{t+sd}(L_\tau)$ where $\tau^{-1} = s + 1/2$, where the difference between these spaces is rather small. Moreover, both are embedded in H^t. Note that whenever $v = \sum_{I \in \mathcal{I}} v_I 2^{-t|I|} \psi_I$ belongs to the Besov space $B_q^{t+sd}(L_p)$ then $v^\nu = \sum_{I \in \mathcal{I}} v_I \psi_I^{(\nu)}$ belongs to $B_q^{sd}(L_p)$ and so does $\mathbf{g}^\nu(\cdot|\mathbf{v})$, by the assumptions on the coefficient matrix $\mathbf{a}(x)$. Now suppose that the inputs $\mathbf{v}$ of MULT are even uniformly bounded in the smaller space $B_q^{sd}(L_p)$, for $p^{-1} < s + 1/2$ and $s \le s'$. The fact that this space is continuously embedded in $\mathcal{A}^s(\Psi^{(\nu)})$ and that, by the above remarks, the supports of the best N-term approximations to $\mathbf{v}$ have nearly tree structure, estimates of the type (4.7.4) allow us to conclude (under the assumption that the trees are balanced with respect to local polynomial approximation) that the second summand on the right hand side of (4.7.5) can be bounded also by $2^{-Ks}\|\mathbf{v}\|_{B_q^{sd}(L_p)}$ and hence is of the same order as the first part. Thus, we conclude an overall error bound of the form

$$\|\mathbf{L}\mathbf{v} - \mathbf{w}_\eta\|_{\ell_2(\mathcal{I})} \lesssim 2^{-Ks}\|\mathbf{v}\|_{B_q^{sd}(L_p)}, \tag{4.7.6}$$

with $K = K(\eta)$ given by estimates of the form (3.9). By the properties of REC the computational work remains proportional to $2^{K(\eta)}$. Hence, under such circumstances we would have an asymptotically optimal work/accuracy balance for the scheme MULT. However, one should note that in the context of the adaptive schemes mentioned in the introduction, the finitely supported iterates $\mathbf{u}^n$ are generally not guaranteed to give rise to approximants that are stable in spaces $B_q^{t+sd}(L_p)$, with s, p related as above, but only on the somewhat larger spaces determined by *best tree N-term approximation rates* where the N terms are subjected to a tree structure, see [10]. In this sense the above assumption leading to (4.7.6) will, in general, be slightly too strong.

5 Numerical Examples

The accuracy of the scheme APPLY, given the exact entries of $\mathbf{L}$ required at each stage, has been already tested and discussed in [2]. So it remains to focus here on the effect of the further perturbations caused by the recovery scheme REC. Therefore it suffices to consider the application of $\mathbf{L}^{\mu,\nu}$ for fixed μ, ν. As pointed out in Section 4.4 this amounts essentially to applying the (weighted) Gramian matrix $\mathbf{A}$ with respect to L_2-Riesz bases. Here we have chosen the biorthogonal wavelet basis generated by the dual generator pair $({}_2\phi, {}_{2,2}\tilde{\phi})$, where ${}_2\phi$ is the second order B-spline. $\mathbf{A}$ is applied to a randomly generated input vector $\mathbf{v}$. The reference solution is computed by assembling the entire matrix up to the dyadic level 12 and executing a standard matrix vector multiplication.

We consider first a single summand produced by the scheme APPR appearing in MULT. Specifically, the solid line refers to the exact application of the compressed version $\mathbf{A}_k$ to $\mathbf{v}_{[3]}$, while the dashed line represents APPR$(\mathbf{A}, \mathbf{v}, 3, k)$. For varying k, the first plot in Figure 1 shows the behavior of the corresponding error while the second plot in Figure 2 monitors the cardinality of the corresponding arrays. We observe that, at least for a reasonable size of the compressed matrix, namely for $k \geq 5$, the accuracy offered by APPR matches that of the standard multiplication very well, while the number of coefficients in the output is significantly smaller. One could, of course, replace the matrices $\mathbf{A}_k$ by $\mathbf{A}_{k+5}$ in (3.8) to avoid the above effect for small k which would be damped for real inputs with corresponding decay anyway.

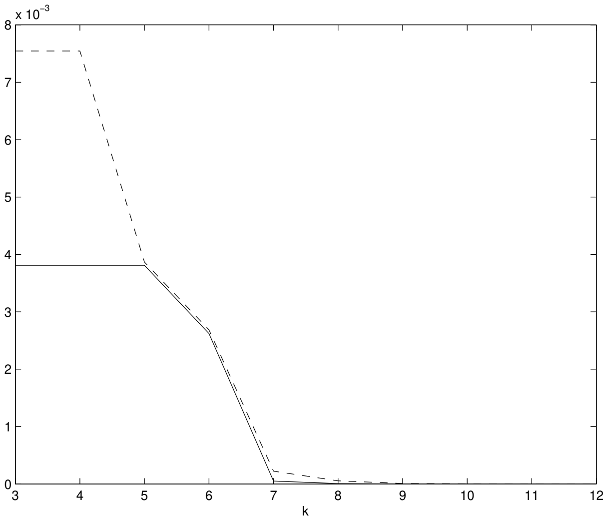

Figure 1: Error for APPR for a random vector

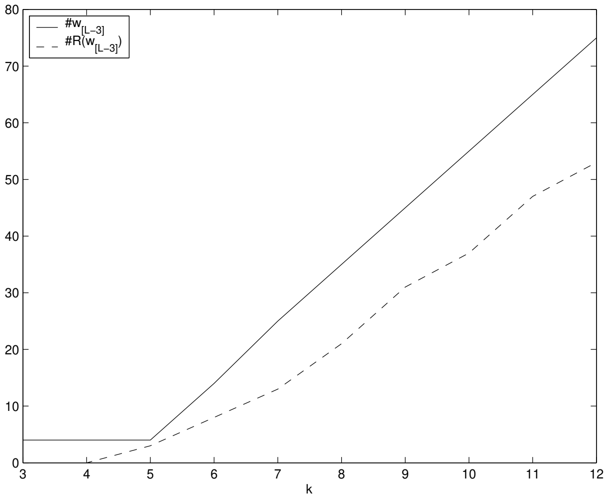

Figure 2: Size of the output of APPR for a random vector

Figure 3 shows the result of MULT$(2^{-0.8k}, \mathbf{A}, \mathbf{v})$ for a sample input $\mathbf{v}$ and a varying compression parameter k. We have displayed the log-log diagram of the cardinality $N = N(k)$ of the support and the achieved accuracy. Following the lines of the proof of Lemma 4.1 in [2], it can be shown, that $\mathbf{A}$ is s^*-compressible for $s^* < 3/2$. This is confirmed by the numerical experiments, showing an approximate error decay of order 1.6.

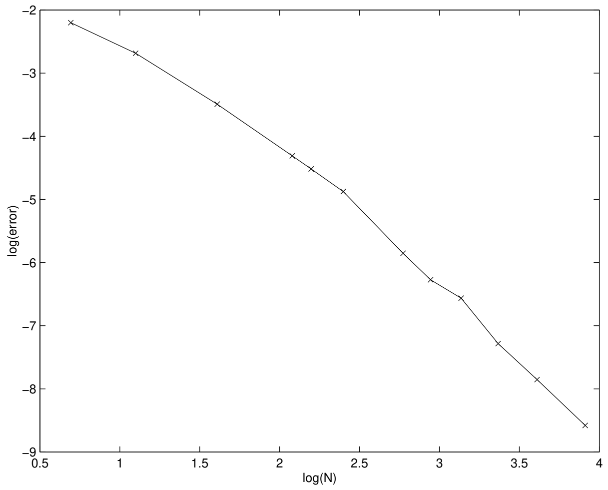

Figure 3: Performance of MULT

References

[1] A. Barinka, Fast Evaluation Tools for Adaptive Wavelet Schemes. PHD-thesis (in preparation).

[2] A. Barinka, T. Barsch, P. Charton, A. Cohen, S. Dahlke, W. Dahmen, K. Urban, Adaptive wavelet schemes for elliptic problems – Implementation and numerical experiments, SIAM J. Sci. Comp., 23, No. 3 (2001), 910–939.

[3] S. Bertoluzza, C. Canuto, K. Urban, On the adaptive computation of integrals of wavelets, Appl. Numer. Math., 34(1) (2000), 13–38.

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

[5] G. Beylkin, J. M. Keiser, An adaptive pseudo-wavelet approach for solving nonlinear partial differential equations, in: *Multiscale Wavelet Methods for PDEs*, W. Dahmen, A. J. Kurdila, P. Oswald (eds.) Academic Press, 1997, 137–197.

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

[7] A. Cohen, W. Dahmen, I. Daubechies, R. DeVore, *Tree approximation and optimal encoding*, Applied and Computational Harmonic Analysis, 11 (2001), 192–226.

[8] A. Cohen, W. Dahmen, and R.A. DeVore, *Adaptive Wavelet Methods for Elliptic Operator Equations — Convergence Rates*, Math. Comp., 70 (2001), 27–75.

[9] A. Cohen, W. Dahmen, and R.A. DeVore, *Adaptive Wavelet Methods II – Beyond the Elliptic Case*, Foundations of Computational Mathematics, 2 (2002), 203–245.

[10] A. Cohen, W. Dahmen, and R.A. DeVore, *Adaptive wavelet schemes for nonlinear variational problems*, IGPM Report # 221, RWTH Aachen, July 2002, SIAM J. Numer. Anal. in press.

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

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

[13] W. Dahmen, *Wavelet and multiscale methods for operator equations*, Acta Numerica, Vol. 6, 1997, 55–228.

[14] W. Dahmen, Wavelet Methods for PDEs – Some Recent Developments, J. Comp. Appl. Math., 128 (2001), 133–185.

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

[16] W. Dahmen, C.A. Micchelli, *Using the refinement equation for evaluating integrals of wavelets*, SIAM J. Numer. Anal., 30(1993), 507-537.

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

[18] W. Dahmen and R. Schneider, *Wavelets with Complementary Boundary Conditions - Function Spaces on the Cube*, Results in Mathematics, 34 (1998), 255–293.

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

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

[21] P.G. Lemarié–Rieusset, Analyses multi-résolutions non orthogonales, commutation entre projecteurs et derivation et ondelettes vecteurs à divergence nulle, Rev. Mat. Iberoamericana, 8 (1992), 221–236.

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

[23] R. Stevenson, Adaptive solution of operator equations using wavelet frames, preprint no. 1239, Dept. of Math., Univ. of Utrecht, May 2002 (revised version Nov. 2002).

[24] P. Tchamitchian, Wavelets, Functions, and Operators, in: *Wavelets: Theory and Applications*, G. Erlebacher, M.Y. Hussaini, L. Jameson (eds.), ICASE/LaRC Series in Computational Science and Engineering, Oxford University Press, 1996, 83–181.

[25] K. Urban, On divergence free wavelets, Advances in Computational Mathematics, 4 (1995), 51–82.

Wolfgang Dahmen, Arne Barinka
Institut für Geometrie und Praktische Mathematik
RWTH Aachen
Templergraben 55
52056 Aachen
Germany
e–mail: {dahmen, barinka}@igpm.rwth-aachen.de
WWW: http://www.igpm.rwth-aachen.de/∼dahmen/
Tel: 49-241-803950

Stephan Dahlke
Philipps-Universität Marburg
FB12 Mathematik und Informatik
Hans-Meerwein Straße
Lahnberge
35032 Marburg
e–mail: dahlke@mathematik.uni-marburg.de
WWW: http://www.mathematik.uni-marburg.de/∼dahlke/
Tel: 49-6421-2825474