ADAPTIVE WAVELET METHODS – BASIC CONCEPTS AND APPLICATIONS TO THE STOKES PROBLEM

WOLFGANG DAHMEN, KARSTEN URBAN, JÜRGEN VORLOEPER

RWTH Aachen, Institut für Geometrie und Praktische Mathematik
Templergraben 55, 52056 Aachen, Germany
Email: {dahmen,urban,jvor}@igpm.rwth-aachen.de

This paper is concerned with recent developments of adaptive wavelet schemes. Central issues are the design of such algorithms and concepts for proving their asymptotically optimal complexity properties when compared with best N–term approximation. After describing the scope of variational problems to be treated, the main concepts of adaptive strategies are briefly reviewed. This is subsequently applied to the important class of saddle point problems. A new convergence proof exemplifies the basic ingredients of the complexity analysis. Finally, the theoretical results are applied to the Stokes problem as a representative example of saddle point problems. In particular, we propose a new variant of an Uzawa iteration based on a different treatment of the divergence operator. We conclude with some numerical comparisons of the different versions of the adaptive saddle point schemes.

Key Words: Variational problems, saddle point problems, wavelet bases, adaptive application of operators, Stokes equations, convergence estimates.
AMS MSC (2000): 41A25, 42C40, 65N12, 65T99

1. Introduction

The attraction of wavelet bases is mainly due to the following essential features. They offer nearly sparse representations of functions. These representations convey quantitative information about the components of such functions living on different length scales. This has led to important applications in signal/image analysis and compression. While in these applications the objects to be analyzed or compressed are given explicitly, it is natural to ask to what extent the above features of wavelet bases can also be exploited for the efficient approximation of objects that are only given *implicitly*, such as the solutions to integral or differential equations. This issue has triggered numerous studies during the past ten years and the reader is referred to survey articles like [17,28,29,47] and further references cited there. First natural steps were to use multiresolution spaces spanned by wavelets (or corresponding scaling

functions) as test and trial spaces for Galerkin methods. In connection with elliptic boundary value problems suitable wavelet bases lead to *asymptotically optimal preconditioners* in terms of simple diagonal scalings in wavelet coordinates. Asymptotically optimal means here that the resulting linear systems can be solved within discretization error accuracy at a computational expense that stays proportional to the problem size. In that sense such schemes are, in principle, comparable with multigrid methods. For certain singular integral equations arising, for instance, as boundary integral equations, in addition to preconditioning, wavelet representations of these operators turn out to be nearly sparse [11]. This has led to matrix compression schemes in connection with collocation or Galerkin schemes for boundary integral equations that again lead to numerical schemes that are asymptotically optimal in the above sense, even for such global operators [30,45]. The key features of wavelet bases that are responsible for these effects are: (I) they induce norm equivalences between sequence spaces and a range of function spaces; (II) they have cancellation properties typically entailed by vanishing polynomial moments, that allow one to tell regions where a function is smooth from those exhibiting more irregular behavior.

These facts and some obvious analogies to nonlinear wavelet compression schemes in image processing have motivated researchers early on to go a step beyond the efficient solution of a linear system. Instead of asking how to efficiently deal with a *given* discretization, one tries to find a possibly economical discretization which then is to be treated efficiently. This leads to the concept of *adaptivity* where the necessary information about how and where to spend the degrees of freedom has to be acquired during the solution process, in contrast to image compression where full information about the object is available from the beginning. Adaptive solution concepts for the numerical treatment of operator equations have been studied intensely in different communities. In the finite element context *a-posteriori error estimators* have proven to yield powerful adaptation strategies, see e.g. [2,3,4,13,36,37]. In the wavelet context it is natural to derive information from the size of the wavelet coefficients of current approximations, see e.g. [1,9,12,21,23,27]. However, neither of these techniques has lead to rigorous *convergence and complexity estimates* that allow one to relate the target accuracy ϵ to the computational work and the adaptively generated number $N(\epsilon)$ of degrees of freedom needed to achieve that target accuracy. In fact, in the context of information based complexity one considers computational models where adaptivity is shown not to pay off [43].

Only recently, rigorous complexity estimates in the above sense have been obtained for certain adaptive wavelet schemes. First in [18] asymptotic optimality has been proven for elliptic problems, see [7] for corresponding data

structures, implementations and first numerical experiments. The extension of these results to more general indefinite problems in [19] culminated in a new conceptual framework. The original problem is first transformed with the aid of suitable wavelet bases into an equivalent problem that is well posed in Euclidean metric. Then one seeks an iteration scheme for the infinite dimensional problem for which the error is reduced at each step by at least a fixed ratio. Then, at last, the application of the involved (infinite dimensional) operators is carried out approximately in an adaptive way within suitable dynamically updated accuracy tolerances. Similar ideas are pursued in [25,27] to deal with the important class of *saddle point problems*. An Uzawa iteration for the infinite dimensional transformed problem is shown to lead to asymptotically optimal performance. Moreover, this approach allows one to recycle the ingredients for the elliptic case. A similar idea is used in [3] in a finite element context, where however, as in [27], no complexity estimates are given.

The objective of this paper is twofold. First we briefly summarize the paradigm proposed in [19]. Second we show how to interpret the Uzawa approach from [25] for saddle point problems as a special realization of the framework in [19]. The corresponding new convergence proof also serves to exemplify this framework and its main ingredients from nonlinear approximation theory. We then revisit the Stokes problem and discuss a new algorithmic realization that offers several quantitative improvements over [25]. It concerns the application of the divergence operator in wavelet coordinates. With the aid of judiciously chosen wavelet bases it can be carried out exactly and simplifies the Uzawa updates in several ways. Some numerical experiments are used to compare the different versions.

The layout of the paper is as follows. In Section 2 we describe the class of variational problems of interest with special emphasis on saddle point problems and the Stokes problem as a representative example. Section 3 contains a brief discussion of our basic questions and objectives. Section 4 begins with an outline of the basic algorithmic paradigm, identifies the key ingredients of a general adaptive solution scheme (GASS), reviews some relevant concepts of nonlinear approximation and best N-term approximation and formulates a general complexity estimate from [19]. In Section 5 we describe an equivalent ℓ_2 formulation of saddle point problems, based on suitable wavelet bases for the relevant function spaces, and motivate the Uzawa scheme. The main algorithmic ingredients of GASS for the class of saddle point problems are realized in Section 6. It is interesting to note that the GASS is used on two levels, namely for the outer Uzawa iteration as well as for the interior elliptic subproblems. Section 7 is devoted to the main complexity estimate for this scheme followed in Section 8 by a brief discussion of the question under which circumstances

adaptive concepts outperform simpler schemes based on uniform refinements by an asymptotically better rate. In Section 9 we focus on the Stokes problem. After briefly describing the wavelet bases we address an alternative application of the divergence operator in wavelet coordinates, which is based on a special choice of biorthogonal wavelet bases. Finally, we present in Section 10 some numerical experiments comparing the different versions of the adaptive wavelet scheme.

Throughout the remainder of the paper we shall adopt the following notational convention. Whenever the specific value of a constant in an inequality does not matter we shall write $a \lesssim b$ to express that a can be bounded by some constant multiple of b, uniformly with respect to any parameters on which a and b may depend. Likewise $a \sim b$ means that $a \lesssim b$ and $b \lesssim a$ hold.

2. Variational Problems

We shall be concerned with the following problem setting. Let $\mathcal{H}$ be a Hilbert space and $A(\cdot,\cdot) : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$ a continuous bilinear form, i.e.,

$$|A(V,U)| \lesssim \|V\|_{\mathcal{H}}\|U\|_{\mathcal{H}}, \quad V,U \in \mathcal{H}. \tag{1}$$

We shall consider the variational problem: Given $F \in \mathcal{H}'$ find $U \in \mathcal{H}$ such that

$$A(V,U) = \langle V,F\rangle, \quad V \in \mathcal{H}. \tag{2}$$

We explain first what we mean when saying that (2) is *well-posed*. To this end, define the operator $\mathcal{L} : \mathcal{H} \to \mathcal{H}'$ by

$$\langle V,\mathcal{L}U\rangle = A(V,U), \quad V \in \mathcal{H}, \tag{3}$$

so that (2) is equivalent to

$$\mathcal{L}U = F. \tag{4}$$

Then (2) is called well-posed (on $\mathcal{H}$) if there exist positive finite constants $c_{\mathcal{L}}, C_{\mathcal{L}}$ such that

$$c_{\mathcal{L}}\|V\|_{\mathcal{H}} \le \|\mathcal{L}V\|_{\mathcal{H}'} \le C_{\mathcal{L}}\|V\|_{\mathcal{H}}, \quad V \in \mathcal{H}. \tag{5}$$

We will refer to (5) as *mapping property*. Clearly (5) implies the existence of a unique solution for any $F \in \mathcal{H}'$ which depends continuously on the data F (with respect to the topology of $\mathcal{H}$).

Remark 2.1. It should be noted that in many cases, given $A(\cdot,\cdot)$ the first (and often most crucial) task is to *identify* a suitable Hilbert space $\mathcal{H}$ such that the mapping property (5) holds.

The simplest example that fits into this framework is the following classical *scalar elliptic boundary value problem.* Suppose that $\Omega \subset \mathbb{R}^d$ is a bounded (Lipschitz) domain and $a(x)$ is a symmetric (bounded) matrix that is uniformly positive definite on Ω. The classical boundary value problem associated with this second order partial differential equation reads

$$-\mathrm{div}\big(a(\cdot)\nabla u\big) + k(\cdot)\, u = f \text{ on } \Omega, \quad u = 0 \text{ on } \partial\Omega. \tag{6}$$

We reserve lower case letters for this problem for later purposes. Its weak formulation has the form (2) with

$$a(v,w) := \int_\Omega (a\nabla v^T \nabla w + k\, v\, w)dx, \quad \mathcal{H} = H_0^1(\Omega), \quad \mathcal{H}' = H^{-1}(\Omega). \tag{7}$$

Classical finite difference or finite element discretizations turn (7) into a *finite dimensional* linear system of equations. When solving these systems, one encounters the following *obstructions:*

- The systems are symmetric, positive definite and sparse but in realistic cases often very large so that the use of direct solvers based on elimination techniques, is excluded. In fact, the *fill-in* caused by elimination would result in prohibitive storage and CPU demands.
- Hence one has to resort to iterative solvers whose efficiency depends on the *condition numbers* of the systems. Unfortunately, the systems are increasingly *ill-conditioned.* When the mesh size h decreases, $\mathrm{cond}_2(a(\Phi_h,\Phi_h)) \sim h^{-2}$, where $a(\Phi_h,\Phi_h)$ denotes the stiffness matrix with respect to an L_2-stable single scale basis, such as a standard nodal finite element basis.

An important class of problems that are no longer coercive are *saddle point problems.* A detailed treatment of this type of problems can be found in [14,38]. Suppose X, M are Hilbert spaces and that $a(\cdot,\cdot)$, $b(\cdot,\cdot)$ are bilinear forms on $X \times X$, respectively $X \times M$ which are continuous

$$|a(v,w)| \lesssim \|v\|_X \|w\|_X, \quad |b(q,v)| \lesssim \|v\|_X \|q\|_M. \tag{8}$$

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

$$A(U,V) := \begin{cases} a(u,v) + b(p,v) = \langle f, v\rangle, \\ b(q,u) \qquad\quad\; = \langle q, g\rangle. \end{cases} \tag{9}$$

Note that when $a(\cdot,\cdot)$ is symmetric positive definite, the solution component u minimizes the quadratic functional $J(w) := \frac{1}{2}a(w,w) - \langle f, w\rangle$ subject to the

constraint $b(u,q) = \langle q, g\rangle$, for all $q \in M$, i.e., one seeks a pair (u,p) solving

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

This accounts for the term saddle point problem (even under more general assumptions on $a(\cdot,\cdot)$) [14].

In order to write (9) as an operator equation, define the operators A, B by

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

so that (9) becomes

$$\mathcal{L}U := \begin{pmatrix} A & B' \\ B & 0 \end{pmatrix}\begin{pmatrix} u \\ p \end{pmatrix} = \begin{pmatrix} f \\ g \end{pmatrix} =: F. \tag{10}$$

As for the mapping property (5), a simple (sufficient) condition reads as follows [14,38]. If $a(\cdot,\cdot)$ is *elliptic* on

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

i.e.,

$$a(v,v) \sim \|v\|_X^2, \quad v \in \ker B, \tag{11}$$

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

$$\inf_{q\in M}\sup_{v\in X}\frac{b(v,q)}{\|v\|_X\|q\|_M} > \beta \tag{12}$$

for some positive β, then (8) is well-posed, i.e., $\mathcal{L}$ defined by (10) satisfies

$$c_{\mathcal{L}}\left(\|v\|_X^2 + \|q\|_M^2\right)^{1/2} \le \|\mathcal{L}\begin{pmatrix} v \\ q \end{pmatrix}\|_{X'\times M'} \le C_{\mathcal{L}}\left(\|v\|_X^2 + \|q\|_M^2\right)^{1/2}. \tag{13}$$

Condition (12) means that B is surjective (and thus has closed range). Condition (11) is actually too strong. It can be replaced by requiring bijectivity of A on $\ker B$, see [14]. We pause indicating the obstructions posed by this class of problems.

- As in the previous case discretizations lead usually to large linear systems that become more and more ill-conditioned when the resolution of the discretization increases.
- An additional difficulty is caused by the fact that the form (9) is *indefinite*, so that more care has to be taken when devising an iterative scheme.

- An important point is that the well-posedness of the infinite dimensional problem (13) is *not* automatically inherited by a finite dimensional Galerkin discretization. In fact, the trial spaces in X and M have to be compatible in the sense that they satisfy the inf-sup condition (12) *uniformly* with respect to the resolution of the chosen discretizations. This is called the *Ladyshenskaya-Babuška-Brezzi-condition* (LBB). The construction of trial spaces that satisfy LBB may, depending on the problem, be a delicate task.

A prominent example is the inhomogeneous *Stokes system* as the simplest model for viscous incompressible fluid flow

$$\begin{aligned} -\nu\Delta u + \nabla p &= f, \qquad \text{in } \Omega, \\ \operatorname{div} u &= g, \qquad \text{in } \Omega, \\ u|_{\partial\Omega} &= 0, \end{aligned} \tag{14}$$

where u and p are the velocity, respectively pressure, see [14,38]. The relevant function spaces are

$$X = \mathbf{H}_0^1(\Omega) := (H_0^1(\Omega))^d, \quad M = L_{2,0}(\Omega) := \left\{ q \in L_2(\Omega) : \int_\Omega q = 0 \right\}, \tag{15}$$

where $H_0^1(\Omega)$ consists of those L_2-functions with first order weak derivatives in L_2 whose trace vanishes on the boundary $\partial\Omega$. Moreover, one can show that the range of the divergence operator is $L_{2,0}(\Omega)$. The weak formulation of (14) is

$$\begin{aligned} \nu\langle \nabla v, \nabla u\rangle_{\mathbf{L}_2(\Omega)} + \langle \operatorname{div} v, p\rangle_{L_2(\Omega)} &= \langle f, v\rangle_{\mathbf{L}_2(\Omega)}, \quad v \in \mathbf{H}_0^1(\Omega) \\ \langle \operatorname{div} u, q\rangle_{L_2(\Omega)} &= \langle g, q\rangle_{L_2(\Omega)}, \quad q \in L_{2,0}(\Omega), \end{aligned} \tag{16}$$

i.e., one seeks a solution $U = (u, p)$ in the energy space

$$\mathcal{H} = X \times M = \mathbf{H}_0^1(\Omega) \times L_{2,0}(\Omega), \tag{17}$$

for which the mapping property (5) can be shown to hold, see e.g. [38].

3. Objectives

A common approach to problems of the form (2) is to choose some finite dimensional trial space $S_h \subset \mathcal{H}$ and seek $U_h \in S_h$ satisfying

$$\langle V_h, U_h\rangle = \langle V_h, F\rangle, \quad V_h \in S_h. \tag{18}$$

We have used here for simplicity the same test space as the trial space which is often referred to as *Galerkin* discretization. (Of course, there are cases where different test spaces would be preferable). Determining U_h from such an a-priori choice of a trial space is a *linear* process. The subscript h represents

typically a *mesh size* which appears in *a-priori error estimates*. Specifically, when $\mathcal{H} = H^t$ is (a closed subspace of) a Sobolev space of order t, such estimates take the form

$$\|U - U_h\|_{H^t} \lesssim h^{s-t}\|U\|_{H^s}, \quad h \to 0, \tag{19}$$

provided that the test spaces have sufficiently high order and that the solution belongs to H^s for some $s > t$. Note that in these error estimates the error is estimated in the same metric (here L_2) as the regularity of the solution. Thus the more regular the solution is, the larger h can be kept in order to achieve some *target accuracy* ϵ. Now, when the solution happens to have singularities so that the global Sobolev regularity s exceeds the order t of the energy norm by only a little, the mesh size h has to be small, to meet the target accuracy, which means that the linear systems resulting from (18) become large.

Since away from singularities a coarser mesh size would suffice to resolve the solution adequately, a natural idea for reducing the computational complexity is to *locally adapt* the trial space to the structure of the solution. For instance, a-priori knowledge about singularities may result in either adding singularity functions or using locally refined meshes. In case of insufficient knowledge about the solution, as is usually the case in 3D problems, it would be particularly desirable, if the solution process can be arranged to be "intelligent" enough so as to acquire during the solution process the necessary information for placing the degrees of freedom in such a way that the target accuracy is met at the expense of possibly few degrees of freedom. Since this may result in highly nonuniform meshes it makes no longer sense to measure the error in terms of mesh size. The cost of the numerical task and the corresponding storage demands are better reflected by the size of the resulting linear systems and hence by the number N of degrees of freedom.

However, to appraise the merit of an adaptive solution concept, that will naturally require some computational overhead and possibly much more complicated data structures, one would have to ask under which circumstances the cost N is related in which way to the target accuracy ϵ. To make this more precise, note first that also the above estimates (19) can be restated in terms of the *number of degrees of freedom* (d.o.f.). In fact, since on bounded domains in $\mathbb{R}^d$ the d.o.f. N of a regular grid behaves like h^{-d}, we have in the case (19) the familiar relation $\epsilon \sim h^{s-t} \sim N^{-(s-t)/d}$ between accuracy and computational work. We shall refer to this relation as the *work/accuracy rate* in this case of the *linear* solution process. Suppose now that some adaptive, and hence *nonlinear* solution process provides an approximate solution $U_{(N)}$, involving N degrees of freedom, such that

$$\|U - U_{(N)}\|_{H^t} \leq \epsilon. \tag{20}$$

Then the question arises how is N related to ϵ under which assumptions on U? To our knowledge, nothing is known about this question for classical discretization environments. Determining and realizing work/accuracy rates of the form $\epsilon \lesssim N^{-\alpha/d}$ or equivalently $N \lesssim \epsilon^{-d/\alpha}$ also for adaptive schemes is therefore a central issue in the subsequent discussion and has driven the developments in [18,19,25].

4. The Basic Paradigm

The *classical approach* to the numerical treatment of operator equations may be summarized as follows. Starting with a variational formulation, the choice of finite dimensional trial and test spaces determines a discretization of the continuous problem which eventually leads to a *finite dimensional problem*. The issue then is to develop efficient solvers for such problems.

As indicated in Section 2, typical obstructions are then the size of the systems, ill-conditioning, as well as compatibility constraints like the *LBB condition*.

In connection with wavelet based schemes a different paradigm has been proposed and analyzed in [19].

(I) One starts again with a variational formulation but puts first most emphasis on the *mapping property* (cf. (5)).
(II) Instead of turning to a finite dimensional approximation, the continuous problem is transformed into an *equivalent* ∞*-dimensional* ℓ_2-problem which is well-conditioned.
(III) One then tries to devise a *convergent iteration* for the ∞*-dimensional* ℓ_2-problem.
(IV) This iteration is, of course, only conceptual. Its numerical realization relies on the *adaptive application* of the involved operators.

In this framework adaptivity enters only through ways of applying infinite dimensional operators within some accuracy tolerance. Moreover, all algorithmic steps take place in ℓ_2. The choice of the wavelet basis fully encodes the original problem, in particular, the underlying geometry. Of course, the realization of appropriate bases by itself may be a difficult problem depending on the case at hand. We shall briefly address this issue later in connection with numerical applications.

The meaning of well-posedness in step (I) has been already clarified in (5).

The objective addressed in step (II) above is to transform the original variational problem (2), or equivalently the operator equation (4), into an

equivalent infinite dimensional system

$$\mathbf{Mp} = \mathbf{G}, \tag{21}$$

whose solution $\mathbf{p}$ uniquely determines the solution U of the original problem. The transformed problem is to be well posed in the sense that $\mathbf{M}$ is an automorphism on ℓ_2, i.e., there exist positive constants c_M, C_M such that

$$c_M\|\mathbf{q}\|_{\ell_2} \leq \|\mathbf{Mq}\|_{\ell_2} \leq C_M\|\mathbf{q}\|_{\ell_2}, \quad \mathbf{q} \in \ell_2. \tag{22}$$

Moreover, we require here that $\mathbf{M}$ is even *symmetric positive definite*. Before explaining possible realizations of (II), we follow [19] and outline first its basic consequences. Note that, on account of (22), there exists some relaxation weight α with

$$\|\mathbf{I} - \alpha\mathbf{M}\|_{\ell_2 \to \ell_2} \leq \rho < 1. \tag{23}$$

Thus simple iterations of the form

$$\mathbf{p}^{n+1} = \mathbf{p}^n + \alpha(\mathbf{G} - \mathbf{Mp}^n) \tag{24}$$

would then converge with a *fixed* reduction rate at most ρ per step.

4.1. *An Adaptive Iteration*

Of course, the scheme (24) is *idealized* since it cannot be executed in practice. All one can hope for is to *approximate* the *exact residual* $\mathbf{G} - \mathbf{Mp}^n$ in each step within appropriate accuracy tolerances. Thus one needs a routine

$$\mathbf{RES}\,[\eta, \mathbf{M}, \mathbf{G}, \mathbf{q}] \to \mathbf{r}_\eta,$$

which for any finitely supported $\mathbf{q}$ produces a finitely supported output $\mathbf{r}_\eta$ such that

$$\|\mathbf{G} - \mathbf{Mq} - \mathbf{r}_\eta\|_{\ell_2} \leq \eta. \tag{25}$$

The realization of **RES** depends on the concrete application and will be discussed later. An essential ingredient will always be a way of applying an infinite matrix to a finitely supported sequence within some accuracy tolerance.

A second basic ingredient is the following routine which has been introduced and analyzed in [18]:

$$\mathbf{COARSE}\,[\eta, \mathbf{w}] \to \bar{\mathbf{w}}_\eta,$$

which for any finitely supported input vector $\mathbf{w}$ produces another vector $\bar{\mathbf{w}}_\eta$ with smallest possible support such that

$$\|\mathbf{w} - \bar{\mathbf{w}}_\eta\|_{\ell_2} \leq \eta. \tag{26}$$

The need for the routine **COARSE** may not be clear yet at this point so that a few comments are in order. It will be seen later that in the simplest case the right hand side $\mathbf{G}$ in (21) is comprised of the (scaled dual) wavelet coefficients of a given right hand side f, and thus is in principle accessible. One should then think of computing in a preprocessing step a highly accurate approximation to f in the dual basis along with the corresponding coefficients collected in a finitely supported array $\bar{\mathbf{G}}$. The necessary accuracy can be easily related to the target accuracy of the adaptive process. So part of the routine **RES** would in this case be an application of **COARSE** to this finitely supported array $\bar{\mathbf{G}}$.

The routine **COARSE** can then be realized by (quasi-)sorting the coefficients and by adding successively the squares of the entries from small to large, until the target threshold is reached, see [18] for details.

It will also be seen later that **COARSE** will actually play an essential role in estimating the complexity of approximate iterations based on the above routines.

We will *assume* for the moment that we have the above routine **RES** at hand and wish to determine first for which tolerances η the corresponding perturbation of the ideal scheme (24) converges. To this end, let

$$K := \min\{l \in \mathbb{N} : \rho^{l-1}(\alpha l + \rho) \leq 1/10\}, \tag{27}$$

where α, ρ are the relaxation parameter from (24) and the error reduction rate from (23), respectively. We consider then the following perturbed version of (24) [19]:

SOLVE $[\epsilon, \mathbf{M}, \mathbf{G}] \to \bar{\mathbf{p}}(\epsilon)$

(i) Set $\bar{\mathbf{p}}^0 = \mathbf{0}$, $\epsilon_0 := c_M^{-1}\|\mathbf{G}\|_{\ell_2}$, $j = 0$.
(ii) If $\epsilon_j \leq \epsilon$, stop $\bar{\mathbf{p}}^j \to \mathbf{p}(\epsilon)$. Else $\mathbf{q}^0 := \bar{\mathbf{p}}^j$.

(ii.1) For $l = 0, \ldots, K-1$:
RES $[\rho^l \epsilon_j, \mathbf{M}, \mathbf{G}, \mathbf{q}^l] \to \mathbf{r}_l$; and set $\mathbf{q}^{l+1} := \mathbf{q}^l + \alpha \mathbf{r}_l$.
(ii.2) **COARSE** $[\mathbf{q}^K, 2\epsilon_j/5] \to \bar{\mathbf{p}}^{j+1}$,
$\epsilon_{j+1} := \epsilon_j/2$, $j+1 \to j$ go to **(ii)**.

Given the above basic routines as ingredients, the scheme **SOLVE** can be shown to reach the target accuracy after finitely many steps [19]. In fact, when

in the jth iteration block (ii.1) $\mathbf{p}^l$ is the exact iteration $\mathbf{p}^{l+1} = \mathbf{p}^l + \alpha(\mathbf{G} - \mathbf{M}\mathbf{p}^l)$ with starting value $\mathbf{p}^0 = \mathbf{q}^0 = \bar{\mathbf{p}}^j$, one has

$$\mathbf{q}^{l+1} - \mathbf{p}^{l+1} = (\mathbf{I} - \alpha\mathbf{M})(\mathbf{q}^l - \mathbf{p}^l) + \alpha\left((\mathbf{G} - \mathbf{M}\mathbf{q}^l) - \mathbf{r}_l\right),$$

so that

$$\|\mathbf{q}^{l+1} - \mathbf{p}^{l+1}\|_{\ell_2} \leq \rho\|\mathbf{q}^l - \mathbf{p}^l\|_{\ell_2} + \alpha\rho^l\epsilon_j \leq \alpha(l+1)\rho^l\epsilon_j, \tag{28}$$

and the arguments in [19] provide the following result.

Prop 4.1. The approximations $\bar{\mathbf{U}}^j$ satisfy

$$\|\mathbf{p} - \bar{\mathbf{p}}^j\|_{\ell_2} \leq \epsilon_j, \quad j \in \mathbb{N}. \tag{29}$$

Note that the accuracy tolerances are at each stage comparable to the current accuracy, which will be important for later complexity estimates.

In some cases a better initial guess is known. If one has for some finitely supported $\tilde{\mathbf{p}}$ that $\|\mathbf{p} - \tilde{\mathbf{p}}\|_{\ell_2} \leq \delta$, step (i) can be replaced by

(i)' Set $\bar{\mathbf{p}}^0 = \mathbf{q}$, $\epsilon_0 := \delta$.

If this matters we shall write $\mathbf{SOLVE}\,[\epsilon, \mathbf{M}, \mathbf{G}, \tilde{\mathbf{p}}] \to \bar{\mathbf{p}}(\epsilon)$ in order to indicate the particular initial guess.

The above scheme should be viewed as the simplest example of a perturbed representation of an iteration for the infinite dimensional problem. Several alternatives come to mind. Instead of applying always K steps in (ii.1), one can monitor the approximate residual for possible earlier termination. Furthermore, the fixed relaxation parameter α can be replaced by a stage dependent parameter α_l resulting from a line search in a *gradient iteration*. Finally, one could resort to (approximate) conjugate gradient iterations. We postpone discussing these issues to later numerical realizations.

4.2. *Best N-Term Approximation*

Before specializing the above concepts to concrete cases we shall lay out the basic principles for estimating the complexity of schemes like **SOLVE**. We wish to compare the performance of the above algorithm with what could be achieved *ideally*, namely with the work/accuracy balance of the *best N-term approximation* in ℓ_2, whose error is given by

$$\sigma_{N,\ell_2}(\mathbf{q}) := \|\mathbf{q} - \mathbf{q}_N\|_{\ell_2} = \min_{\#\mathrm{supp}\,\mathbf{z} \leq N} \|\mathbf{q} - \mathbf{z}\|_{\ell_2}. \tag{30}$$

Obviously, $\mathbf{q}_N$ consists of the N largest terms (in modulus) in the sequence $\mathbf{q}$. It will be important to understand the class of those sequences for which σ_{N,ℓ_2} decays at a given rate.

One way to describe sequences that are *sparse* in the sense that $\sigma_{N,\ell_2}(\mathbf{v})$ decays fast, is to control the number of terms exceeding any given threshold. In fact, set for some $\tau < 2$

$$\ell_\tau^w := \left\{\mathbf{v} \in \ell_2 : \#\{\lambda \in \mathcal{J} : |v_\lambda| > \eta\} \leq C_{\mathbf{v}} \eta^{-\tau}\right\}, \tag{31}$$

noting that, for $\tau \geq 2$, there is no constraint imposed by the above condition. When for a given $\mathbf{v} \in \ell_\tau^w$, $C_{\mathbf{v}}$ is the smallest constant in (31), one has

$$C_{\mathbf{v}}^{1/\tau} = \sup_{n \in \mathbb{N}} n^{1/\tau} v_{1+n}^* =: |\mathbf{v}|_{\ell_\tau^w}, \tag{32}$$

where $\{v_n^*\}_{n \in \mathbb{N}_0}$ is a non-decreasing rearrangement of $\mathbf{v}$. Thus

$$\|\mathbf{v}\|_{\ell_\tau^w} := \|\mathbf{v}\|_{\ell_2} + |\mathbf{v}|_{\ell_\tau^w} \tag{33}$$

is a (quasi-) norm for ℓ_τ^w [18]. It is easy to see that ℓ_τ^w is very close to ℓ_τ as reflected by the following continuous embeddings

$$\ell_\tau \subset \ell_\tau^w \subset \ell_{\tau+\epsilon} \subset \ell_2, \quad \tau < \tau + \epsilon < 2. \tag{34}$$

The following characterization of sequences with polynomial best N-term approximation rates from [18] will be needed later.

Prop 4.2. Let

$$\frac{1}{\tau} = s + \frac{1}{2}. \tag{35}$$

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

$$\|\mathbf{v} - \mathbf{v}_N\|_{\ell_2} \lesssim N^{-s} \|\mathbf{v}\|_{\ell_\tau^w}.$$

Moreover, the following result from [18] indicates the role of the coarsening step (ii.2) in **SOLVE**.

Prop 4.3. If $\mathbf{v} \in \ell_\tau^w$ and $\|\mathbf{v} - \mathbf{w}\|_{\ell_2} \leq \eta/5$ with $\#\operatorname{supp} \mathbf{w} < \infty$. Then $\bar{\mathbf{w}}_\eta := \mathbf{COARSE}\,[\mathbf{w}, 4\eta/5]$ satisfies

$$\#\operatorname{supp} \bar{\mathbf{w}}_\eta \lesssim \|\mathbf{v}\|_{\ell_\tau^w}^{1/s} \eta^{-1/s}, \quad \|\mathbf{v} - \bar{\mathbf{w}}_\eta\|_{\ell_2} \leq \eta, \tag{36}$$

and

$$\|\bar{\mathbf{w}}_\eta\|_{\ell_\tau^w} \lesssim \|\mathbf{v}\|_{\ell_\tau^w}. \tag{37}$$

Thus the coarsening step controls the ℓ_τ^w-norm of the current approximants which are thereby pulled towards the best N-term approximation. To explain its role in step (ii.2) of **SOLVE**, note that the choice of the constant K in step (ii.1) guarantees that in the $(j+1)$st iteration block of **SOLVE** the iterate $\mathbf{q}^K$ satisfies $\|\mathbf{p} - \mathbf{q}^K\|_{\ell_2} \leq \epsilon_j/10$. Thus, by Proposition 4.3, the threshold

parameter $2\epsilon_j/5$ in step (ii.2) ensures that the ℓ_2-error after coarsening is still at most $\epsilon_j/2$, while, when $\mathbf{p} \in \ell_\tau^w$, the ℓ_τ^w-norm of $\bar{\mathbf{p}}^j$ remains bounded and the support size of $\bar{\mathbf{p}}^j$ grows at most like $\epsilon_j^{-1/s}$. This is the best N-term rate of elements in ℓ_τ^w.

4.3. *The Main Complexity Estimate*

In order to analyze the complexity of the full scheme **SOLVE**, it remains, in view of the above remarks, to control the complexity of the iteration blocks between coarsening steps. This will rely on properties of the routine **RES**. The following result, whose formulation is tailored to the present needs, can be easily extracted from the analysis in [19].

Theorem 4.1. *Assume that for some fixed $\tau^* > 0$ and any $\tau^* < \tau < 2$ the routine **RES** has the following properties:*

(i) For any finitely supported $\mathbf{q}$ the output $\mathbf{r}_\eta$ of $\mathbf{RES}[\eta, \mathbf{M}, \mathbf{G}, \mathbf{q}]$ satisfies

$$\begin{aligned} \|\mathbf{r}_\eta\|_{\ell_\tau^w} &\lesssim \max\{\|\mathbf{q}\|_{\ell_\tau^w}, \|\mathbf{G}\|_{\ell_\tau^w}\}, \\ \#\mathrm{supp}\,\mathbf{r}_\eta &\lesssim \max\{\|\mathbf{q}\|_{\ell_\tau^w}^{1/s}, \|\mathbf{G}\|_{\ell_\tau^w}^{1/s}\}\eta^{-1/s}. \end{aligned} \tag{38}$$

where as before $\tau^{-1} = s + 1/2$, see (35).

(ii) Moreover, the number of floating point operations stays proportional to $\#\mathrm{supp}\,\mathbf{r}_\eta$. where in (i) and (ii) the constants depend only on τ when τ tends to τ^*.

Then the output $\mathbf{p}(\epsilon)$ of $\mathbf{SOLVE}[\epsilon, \mathbf{M}, \mathbf{G}]$ satisfies the following properties. For every $\epsilon > 0$ it produces after finitely many steps a finitely supported solution $\mathbf{p}(\epsilon)$ such that $\|\mathbf{p} - \mathbf{p}(\epsilon)\|_{\ell_2} \le \epsilon$. Moreover, when $\mathbf{p} \in \ell_\tau^w$ for $\tau^* < \tau < 2$, one has for s related to τ according to (35),

$$\#\mathrm{supp}\,\mathbf{p}(\epsilon) \lesssim \|\mathbf{p}\|_{\ell_\tau^w}^{1/s}\epsilon^{-1/s}, \quad \|\mathbf{p}(\epsilon)\|_{\ell_\tau^w} \lesssim \|\mathbf{p}\|_{\ell_\tau^w}, \tag{39}$$

and the number of floating point operations stays proportional to $\#\mathrm{supp}\,\mathbf{p}(\epsilon)$.

This result will be applied later on several levels.

5. An Equivalent ℓ_2-Problem

In order to apply the above concepts we need to specify the transformation to the equivalent ℓ_2-problem in step (II) and to realize the corresponding basic routines. The general idea proposed in [19] is to base the transformation of (2) to a problem of the form (21) on appropriate *wavelet bases* for the energy space $\mathcal{H}$.

We postpone discussing the concrete construction of such bases but are content for the moment with describing their general format. We consider collections $\Psi = \{\psi_\lambda : \lambda \in \mathcal{J}\} \subset L_2(\Omega)$ of functions – wavelets – that are normalized in L_2, i.e. $\|\psi_\lambda\|_{L_2} = 1$, $\lambda \in \mathcal{J}$, where $\dim \Omega = d$. Here $\mathcal{J} = \mathcal{J}_\phi \cup \mathcal{J}_\psi$ is an infinite index set where: $\#\mathcal{J}_\phi < \infty$ representing the "scaling functions", living on the coarsest scale. For Euclidean domains these functions will span polynomials up to some order which will be called the order of the basis Ψ. The indices in $\mathcal{J}_\psi$ represent the "true" wavelets spanning complements between refinement levels. Each index $\lambda \in \mathcal{J}$ encodes different types of information, namely the *scale* $j = j(\lambda) = |\lambda|$, the *spatial location* $k = k(\lambda)$ and the *type* $e = e(\lambda)$ of the wavelet. Recall that e.g. for tensor product constructions one has $2^d - 1$ different types of wavelets associated with each spatial index k.

As an example, for $d = 2$ one has three different kind of functions, namely

$$\begin{aligned}
2^j\psi^{1,0}(2^j(x,y) - (k,l)) &= 2^{j/2}\psi(2^j x - k)\ 2^{j/2}\phi(2^j y - l),\\
2^j\psi^{0,1}(2^j(x,y) - (k,l)) &= 2^{j/2}\phi(2^j x - k)\ 2^{j/2}\psi(2^j y - l),\\
2^j\psi^{1,1}(2^j(x,y) - (k,l)) &= 2^{j/2}\psi(2^j x - k)\ 2^{j/2}\psi(2^j y - l),
\end{aligned}$$

for $\lambda = (j,(k,l),(1,0))$, $\lambda = (j,(k,l),(0,1))$ and $\lambda = (j,(k,l),(1,1))$, respectively. To simplify notation we shall formally view Ψ as an infinite column vector (whose components ψ_λ are ordered in some fixed but unspecified way) an denote an expansion briefly as $\mathbf{d}^T\Psi = \sum_{\lambda\in\mathcal{J}} d_\lambda\psi_\lambda$.

We will explain next the first feature that qualifies Ψ as a *wavelet basis* in our context. The key is that suitable wavelet bases induce an *isomorphism* between the energy space $\mathcal{H}$ and ℓ_2. We shall exemplify this for the case of saddle point problems. Recall that in the saddle point case we have $\mathcal{H} = X \times M$. We need a wavelet basis for each component space

$$X \leftrightarrow \Psi_X, \qquad M \leftrightarrow \Psi_M.$$

By this we mean that there exist diagonal scaling matrices $\mathbf{D}_X, \mathbf{D}_M$ such that the norm equivalences

$$c_X\|\mathbf{v}\|_{\ell_2(\mathcal{J}_X)} \le \|\mathbf{v}^T\mathbf{D}_X^{-1}\Psi_X\|_X \le C_X\|\mathbf{v}\|_{\ell_2(\mathcal{J}_X)}, \tag{40}$$

and

$$c_M\|\mathbf{q}\|_{\ell_2(\mathcal{J}_M)} \le \|\mathbf{q}^T\mathbf{D}_M^{-1}\Psi_M\|_M \le C_M\|\mathbf{q}\|_{\ell_2(\mathcal{J}_M)}, \tag{41}$$

hold. The latter relation deserves a few more comments. To this end, recall that in the case of the Stokes problem we have $X = \mathbf{H}_0^1(\Omega), M = L_{2,0}(\Omega)$, see (17). Suitable scaling weights are then $(\mathbf{D}_X)_{\lambda,\lambda} = 2^{|\lambda|}$, $(\mathbf{D}_M)_{\lambda,\lambda} = 1$ [28]. However, M is in this case not the full space $L_2(\Omega)$, usually characterized by wavelets, but a closed subspace of codimension one. In general, when M is a

closed subspace of finite codimension in some larger Hilbert space $\hat{M}$ for which (41) holds, the arrays of wavelet coefficients of elements in M will in general form a closed subspace $\ell_{2,0}(\mathcal{J}_M)$ of finite codimension in $\ell_2(\mathcal{J}_M)$. In the case of the Stokes problem $\ell_{2,0}(\mathcal{J}_M)$ is determined by a linear constraint only on the scaling function coefficients. We refer to [25] for more details and will, for simplicity, suppress here this distinction in the following.

Setting,

$$\mathbf{A} := \mathbf{D}_X^{-1} a(\Psi_X, \Psi_X) \mathbf{D}_X^{-1}, \quad \mathbf{B} := \mathbf{D}_M^{-1} b(\Psi_M, \Psi_X) \mathbf{D}_X^{-1}, \tag{42}$$

and

$$\mathbf{f} := \mathbf{D}_X^{-1} \langle \Psi_X, f \rangle, \quad \mathbf{g} := \mathbf{D}_M^{-1} \langle \Psi_M, g \rangle, \tag{43}$$

(9) is equivalent to

$$\mathbf{LU} = \mathbf{F} \iff \underbrace{\begin{pmatrix} \mathbf{A} & \mathbf{B}^T \\ \mathbf{B} & \mathbf{0} \end{pmatrix}}_{\mathbf{L}} \underbrace{\begin{pmatrix} \mathbf{u} \\ \mathbf{p} \end{pmatrix}}_{\mathbf{U}} = \underbrace{\begin{pmatrix} \mathbf{f} \\ \mathbf{g} \end{pmatrix}}_{\mathbf{F}}, \tag{44}$$

which is (4) in wavelet coordinates. Moreover, under the assumptions (11), (12) together with (40), (41) one can show that the mapping property

$$c_L \|\mathbf{V}\|_{\ell_2} \leq \|\mathbf{LV}\|_{\ell_2} \leq C_L \|\mathbf{V}\|_{\ell_2}, \quad \mathbf{V} \in \ell_2, \tag{45}$$

holds, see e.g. [29,28].

Thus, in principle, we have found an equivalent formulation of (2) that is well-posed in ℓ_2. However, $\mathbf{L}$ is an indefinite operator, so that our preference of having a positive definite formulation is not met yet and hence simple iterations of the form (24) cannot be expected to work. One option is to use a least squares formulation $\mathbf{L}^T\mathbf{LU} = \mathbf{L}^T\mathbf{F}$, noting that the resulting operator $\mathbf{M} := \mathbf{L}^T\mathbf{L}$ is still boundedly invertible on ℓ_2, see [19]. However, this would entail squaring the condition number, which could raise the error reduction rate ρ in (24). Therefore, we shall pursue here a different line closely related to [25], see also earlier work in [27]. However, taking here a different point of view in the light of the above framework, the idea is to choose (21) as a Schur complement problem. To this end, recall from (11) that for (45) to hold, A need not be invertible on all of X but only on $\ker B$. Hence $\mathbf{A}$ need not be invertible on $\ell_2(\mathcal{J}_X)$. However, whenever the saddle point problem (9) is well-posed, one can show that for some suitable $c > 0$ the matrix

$$\hat{\mathbf{A}} := \mathbf{A} + c\mathbf{B}^T\mathbf{B} \tag{46}$$

is invertible on all of $\ell_2(\mathcal{J}_X)$. One can then replace (44) by an equivalent system (with adjusted right hand side data) so that (45) is valid. Under these

premises the combination of norm equivalences with the mapping properties of the operators A, B again ensures the existence of positive constants c_A, C_B, C_B such that

$$c_A\|\mathbf{v}\|_{\ell_2(\mathcal{J}_X)} \leq \|\mathbf{A}\mathbf{v}\|_{\ell_2(\mathcal{J}_X)} \leq C_A\|\mathbf{v}\|_{\ell_2(\mathcal{J}_X)}, \quad \mathbf{v} \in \ell_2(\mathcal{J}_X), \tag{47}$$

and

$$\|\mathbf{B}\mathbf{v}\|_{\ell_2(\mathcal{J}_M)} \leq C_B\|\mathbf{v}\|_{\ell_2(\mathcal{J}_X)}, \quad \mathbf{v} \in \ell_2(\mathcal{J}_X). \tag{48}$$

In particular, the Schur complement of $\mathbf{L}$ is well defined. Therefore we will assume in the following that either A is invertible on all of X, or that the above precaution has been been used, leading to a matrix $\hat{\mathbf{A}}$, henceforth again denoted by $\mathbf{A}$.

Thus we can use block elimination and observe that (44) is equivalent to

$$\begin{cases} \mathbf{M}\mathbf{p} := \mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T\mathbf{p} = \mathbf{B}\mathbf{A}^{-1}\mathbf{f} - \mathbf{g} =: \mathbf{G}, \\ \mathbf{A}\mathbf{u} \qquad\qquad\qquad\quad = \mathbf{f} - \mathbf{B}^T\mathbf{p}. \end{cases} \tag{49}$$

On account of (45), we know that

$$\mathbf{M} := \mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T : \ell_2(\mathcal{J}_M) \to \ell_2(\mathcal{J}_M), \quad \|\mathbf{M}\mathbf{q}\|_{\ell_2(\mathcal{J}_M)} \sim \|\mathbf{q}\|_{\ell_2(\mathcal{J}_M)}. \tag{50}$$

Thus, our choice of (21), in the case of saddle point problems, is the first system in (49) with $\mathbf{M}$ being the Schur complement, noting that once $\mathbf{p}$ has been found, $\mathbf{u}$ can, in principle, be determined from the second system in (49). Of course, the obvious obstruction along this line is that $\mathbf{M}$ involves the *inverse* of $\mathbf{A}$ which might make it hard to realize a **RES** scheme for $\mathbf{M}$. To overcome this obstruction note first that, in view of the second relation in (50), there exists a positive relaxation parameter α such that a fixed point iteration (or a gradient iteration) based on the identity

$$\begin{aligned} \mathbf{p} &= \mathbf{p} + \omega\mathbf{R}\left((\mathbf{B}\mathbf{A}^{-1}\mathbf{f} - \mathbf{g}) - \mathbf{M}\mathbf{p}\right) \\ &= \mathbf{p} + \omega\mathbf{R}\left(\mathbf{B}\underbrace{\mathbf{A}^{-1}(\mathbf{f} - \mathbf{B}^T\mathbf{p})}_{=\mathbf{u}} - \mathbf{g}\right) = \mathbf{p} + \omega\mathbf{R}(\mathbf{B}\mathbf{u} - \mathbf{g}) \end{aligned}$$

converges with a fixed reduction rate $\rho < 1$. Here $\mathbf{R}$ is any ℓ_2-isomorphism. It is explained in [25] under which circumstances a nontrivial $\mathbf{R}$ is actually necessary. We shall return to this issue later in connection with the Stokes problem. But since in the present study the simple choice $\mathbf{R} = \mathbf{I}$ will suffice and in order to keep the exposition as simple as possible we set $\mathbf{R} = \mathbf{I}$ throughout the rest of this section and refer to [25] for a detailed discussion of this issue. Thus, replacing for some iterate $\mathbf{p}^n$ the expression $\mathbf{A}^{-1}(\mathbf{f} - \mathbf{B}^T\mathbf{p}^n)$ in this iteration by the solution $\mathbf{u}^n$ of

$$\mathbf{A}\mathbf{u}^n = \mathbf{f} - \mathbf{B}^T\mathbf{p}^n, \tag{51}$$

the iteration (24) for $\mathbf{M}$, given by (50), reduces to the simple update

$$\mathbf{p}^{n+1} = \mathbf{p}^n + \omega(\mathbf{B}\mathbf{u}^n - \mathbf{g}), \tag{52}$$

with $\mathbf{u}^n$ from (51). The iteration (52) is *equivalent* to a gradient iteration (24) for the Schur complement problem with $\mathbf{M}, \mathbf{G}$ from (49) and (50). This is the idea of the *Uzawa scheme* here formulated for the original infinite dimensional problem in wavelet coordinates.

We are now prepared to discuss the last step (IV). In fact, by the above remarks, a possible **RES** scheme could be based here on two major ingredients, namely

a) an approximate solution of (51),
b) an approximate application of $\mathbf{B}$.

The natural idea is now to perform a) by applying **SOLVE** at the nth stage of the outer iteration (52) with $\mathbf{M} = \mathbf{A}$, since $\mathbf{A}$ is already symmetric positive definite, and $\mathbf{G} := \mathbf{f} - \mathbf{B}^T\mathbf{p}^n$. The realization of **RES**, in turn, requires then the following ingredients: 1) the approximate application of the wavelet representation $\mathbf{A}$, coarsening the preprocessed $\mathbf{f}$ and the approximate application of $\mathbf{B}^T$. In fact, we could invoke the results from [18] for that purpose. The only wrinkle is that the right hand side $\mathbf{f}-\mathbf{B}^T\mathbf{p}^n$ is stage dependent and involves the application of $\mathbf{B}^T$ and coarsening $\mathbf{g}$. Similarly b) requires applying $\mathbf{B}$. Hence, in summary, the essential ingredients are therefore the scheme **COARSE** for treating preprocessed versions of the data $\mathbf{f}, \mathbf{g}$ and schemes for the approximate application of wavelet representations of operators like $\mathbf{A}, \mathbf{B}, \mathbf{B}^T$, to come up with computable versions of $\mathbf{RES}\,[\eta, \mathbf{M}, \mathbf{G}, \mathbf{q}]$, with $\mathbf{M}, \mathbf{G}$ from (49). The approach in [25] can be interpreted in this way where the approximate application of $\mathbf{A}, \mathbf{B}, \mathbf{B}^T$ was based on the adaptive application scheme from [18]. This latter scheme uses the concept of *compressible matrices*, noting that wavelet representations of differential operators are compressible in a sense to be explained in more detail later below. Our present approach differs in two ways, namely we apply the concepts from Section 4 in a more direct way to the Schur complement problem, and secondly, we use in the special case of the Stokes problem a different scheme for applying $\mathbf{B}$ that takes advantage of special problem adapted wavelet bases.

6. Realization of RES

This section is devoted to describing the routine **RES** for the numerical realization of the iteration (52). As explained above one needs, in particular, a way of applying the wavelet representation of an operator to a finitely supported

vector. The relevant notion in this context is the notion of *compressibility* introduced in [18]. A matrix $\mathbf{C}$ is said to be s^*-compressible – $\mathbf{C} \in \mathcal{C}_{s^*}$ – if for any $0 < s < s^*$ and every $j \in \mathbb{N}$, there exists a matrix $\mathbf{C}_j$ obtained by replacing all but the order of $\alpha_j 2^j$ ($\sum_j \alpha_j < \infty$) entries per row and column in $\mathbf{C}$ by zero, while still

$$\|\mathbf{C} - \mathbf{C}_j\|_{\ell_2 \to \ell_2} \leq \alpha_j 2^{-js}, \quad j \in \mathbb{N}, \quad \sum_j \alpha_j < \infty. \tag{53}$$

One can use the cancellation properties (CP) to confirm the following claim.

Remark 6.1. The scaled wavelet representations $\mathbf{L}$ for a wide class of differential and integral operators belong to $\mathcal{C}_{s^*}$ for some $s^* = s^*(\mathcal{L}, \Psi) > 0$. In particular, this covers the matrices $\mathbf{A}, \mathbf{B}$ appearing in the Stokes problem when the wavelet bases are chosen appropriately.

In fact, for compressible matrices one can devise approximate application schemes that exhibit in a certain range the asymptotically optimal work/accuracy balance required in Theorem 4.1, see (38). To this end, abbreviate for any finitely supported $\mathbf{v}$ the best 2^j-term approximations as $\mathbf{v}_{[j]} := \mathbf{v}_{2^j}$ and define

$$\mathbf{w}_j := \mathbf{A}_j \mathbf{v}_{[0]} + \mathbf{A}_{j-1}(\mathbf{v}_{[1]} - \mathbf{v}_{[0]}) + \cdots + \mathbf{A}_0(\mathbf{v}_{[j]} - \mathbf{v}_{[j-1]}), \tag{54}$$

as an approximation to $\mathbf{A}\mathbf{v}$. In fact, triangle inequality together with the above compression estimates yield

$$\|\mathbf{A}\mathbf{v} - \mathbf{w}_j\|_{\ell_2} \leq c \underbrace{\|\mathbf{v} - \mathbf{v}_{[j]}\|_{\ell_2}}_{\sigma_{2^j,\ell_2}(\mathbf{v})} + \sum_{l=0}^{j} \alpha_l 2^{-ls} \underbrace{\|\mathbf{v}_{[j-l]} - \mathbf{v}_{[j-l-1]}\|_{\ell_2}}_{\lesssim\ \sigma_{2^{j-l-1},\ell_2}(\mathbf{v})}. \tag{55}$$

One can now exploit the *a-posteriori* information offered by the quantities $\sigma_{2^{j-l-1},\ell_2}(\mathbf{v})$ to choose the smallest j for which the right hand side of (55) is smaller than a given target accuracy η. Since the sum is finite for each finitely supported input $\mathbf{v}$ such a j does indeed exist. This leads to a concrete multiplication scheme

APPLY $[\eta, \mathbf{A}, \mathbf{v}] \to \mathbf{w}_\eta$

which for any finitely supported input vector $\mathbf{v}$ produces a finitely supported output $\mathbf{w}_\eta$, satisfying

$$\|\mathbf{A}\mathbf{v} - \mathbf{w}_\eta\| \leq \eta. \tag{56}$$

This scheme has been developed and analyzed in [18] and implemented in [7]. The main result can be formulated as follows [18].

Theorem 6.1. *If* $\mathbf{A} \in \mathcal{C}_{s^*}$ *then* $\mathbf{A}$ *is bounded on* ℓ_τ^w *for* $\frac{1}{\tau} = s + \frac{1}{2}$ *and* $s < s^*$. *Moreover, for any finitely supported* $\mathbf{v}$ *the output* $\mathbf{w}_\eta$ *of* **APPLY** $[\eta, \mathbf{A}, \mathbf{v}]$ *satisfies* $\|\mathbf{w}_\eta\|_{\ell_\tau^w} \lesssim \|\mathbf{v}\|_{\ell_\tau^w}$ *and one has*

$$\#\mathrm{supp}\,\mathbf{w}_\eta, \quad \#\mathbf{flops} \lesssim \eta^{-1/s}\|\mathbf{v}\|_{\ell_\tau^w}^{1/s}.$$

Thus, **APPLY** has in some range, depending on the compressibility index s^*, an asymptotically optimal work/accuracy balance. In fact, it is pointed out in [6] that an original logarithmic factor, due to sorting operations can be avoided.

In principle, **APPLY** can be used to apply the matrix $\mathbf{A}$ as well as $\mathbf{B}$ e.g. in the Stokes problem as was done in [25]. Later we will introduce, however, an alternative option for $\mathbf{B}$ that will in this case be even more economical.

We now have all the ingredients at hand needed to define a **RES** scheme for the Schur complement $\mathbf{M}$. According to a) above, the first one is a scheme

$$\mathbf{SOLVE}_{\mathrm{ell}}\,[\eta, \mathbf{A}, \mathbf{B}, \mathbf{f}, \mathbf{q}] \to \mathbf{u}(\eta)$$

that produces for a given target accuracy η, the matrix $\mathbf{A}$ from (42), respectively (46), the matrix $\mathbf{B}$ from (43), some $\mathbf{q} \in \ell_2(\mathcal{J}_M)$ and the right hand side data $\mathbf{f}$, an approximate solution $\mathbf{u}(\eta)$ of the problem

$$\mathbf{A}\mathbf{u}_G = \mathbf{G}, \quad \text{where} \quad \mathbf{G} = \mathbf{f} - \mathbf{B}^T\mathbf{q}, \tag{57}$$

satisfying

$$\|\mathbf{u}_G - \mathbf{u}(\eta)\|_{\ell_2} \le \eta. \tag{58}$$

In fact, we simply take

$$\mathbf{SOLVE}_{\mathrm{ell}}[\eta, \mathbf{A}, \mathbf{B}, \mathbf{f}, \mathbf{q}] := \mathbf{SOLVE}\,[\eta, \mathbf{A}, \mathbf{G}], \tag{59}$$

where **SOLVE** is the scheme from Section 4.1 with the routine $\mathbf{RES}_{\mathrm{ell}}$, which for $\mathbf{G}$, given by (57), is defined as follows:

$\mathbf{RES}_{\mathrm{ell}}[\eta, \mathbf{A}, \mathbf{G}, \mathbf{q}] \to \mathbf{r}_\eta$

(i) **APPLY** $[\eta/2, \mathbf{A}, \mathbf{q}] \to \mathbf{w}_\eta$;
(ii) **COARSE** $[\eta/6, \mathbf{f}] \to \mathbf{f}_\eta$;
COARSE $[\eta/(6C_B), \mathbf{q}] \to \mathbf{q}_\eta$;
APPLY $[\eta/(6C_B), \mathbf{B}^T, \mathbf{q}_\eta] \to \mathbf{z}_\eta$, where C_B is the constant from (48).
(iii) Set $\mathbf{r}_\eta := \mathbf{f}_\eta - \mathbf{z}_\eta - \mathbf{w}_\eta$.

One readily checks that (25) holds for $\mathbf{M} = \mathbf{A}$, $\mathbf{G} = \mathbf{f} - \mathbf{B}^T\mathbf{q}$.

In the case that $\mathbf{A}$ is not just the wavelet representation of A but has the form (46), (i) has to be modified as shown in [25]. In fact, the modified matrix

is not assembled but the application involves now the additional applications of $\mathbf{B}$ and $\mathbf{B}^T$.

Remark 6.2. We shall actually apply the routine $\mathbf{SOLVE}_{\mathrm{ell}}$ always with a specific initial guess $\hat{\mathbf{u}}$ along with an error bound for $\|\mathbf{u}-\tilde{\mathbf{u}}\|_{\ell_2(\mathcal{J}_X)}$, which will be indicated as before by writing $\mathbf{SOLVE}_{\mathrm{ell}}[\eta,\mathbf{A},\mathbf{B},\mathbf{f},\mathbf{q},\tilde{\mathbf{u}}]$.

We are now in a position to employ the scheme **SOLVE** also for the Uzawa iteration (52), which only requires specifying a corresponding $\mathbf{RES}_{\mathrm{sc}}$ scheme where the subscript stands for Schur complement. In fact, for

$$\mathbf{M} := \mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T, \quad \mathbf{G} := \mathbf{B}\mathbf{A}^{-1}\mathbf{f}-\mathbf{g}$$

we set

$$\mathbf{UZAWA}\,[\epsilon,\mathbf{A},\mathbf{B},\mathbf{f},\mathbf{g}] := \mathbf{SOLVE}\,[\epsilon,\mathbf{M},\mathbf{G}], \tag{60}$$

where the corresponding **RES** scheme is defined as follows: Given an approximation $\tilde{\mathbf{u}}$ to the exact solution component $\mathbf{u}$ in (44) and error estimates

$$\|\mathbf{u}-\tilde{\mathbf{u}}\|_{\ell_2(\mathcal{J}_X)} \le \delta_u, \quad \|\mathbf{p}-\tilde{\mathbf{p}}\|_{\ell_2(\mathcal{J}_M)} \le \delta_p, \tag{61}$$

proceed as follows:

$\mathbf{RES}_{\mathrm{sc}}[\eta,\mathbf{M},\mathbf{G},\tilde{\mathbf{p}},\tilde{\mathbf{u}},\delta_u,\delta_p] \to (\mathbf{r}_\eta,\mathbf{u}_\eta)$

(i) $\mathbf{COARSE}\,[4\delta_u,\tilde{\mathbf{u}}] \to \bar{\mathbf{u}}_\delta$;
(ii) $\mathbf{COARSE}\,[\eta/3,\mathbf{g}] \to \mathbf{g}_\eta$;
(iii) $\mathbf{SOLVE}_{\mathrm{ell}}[\eta/(3C_B),\mathbf{A},\mathbf{B},\mathbf{f},\tilde{\mathbf{p}},\bar{\mathbf{u}}_\delta] \to \mathbf{u}_\eta$;
(iv) $\mathbf{APPLY}\,[\eta/3,\mathbf{B},\mathbf{u}_\eta] \to \mathbf{w}_\eta$; and set $\mathbf{r}_\eta := \mathbf{w}_\eta - \mathbf{g}_\eta$.

It follows from the choice of the tolerances and the properties of the ingredients that

$$\begin{aligned}\|\mathbf{r}_\eta-(\mathbf{G}-\mathbf{M}\tilde{\mathbf{p}})\|_{\ell_2(\mathcal{J}_M)} &= \|\mathbf{w}_\eta-\mathbf{g}_\eta-\mathbf{B}\mathbf{A}^{-1}(\mathbf{f}-\mathbf{B}^T\tilde{\mathbf{p}})+\mathbf{g}\|_{\ell_2(\mathcal{J}_M)}\\ &\le \|\mathbf{w}_\eta-\mathbf{B}\mathbf{u}_\eta\|_{\ell_2(\mathcal{J}_M)}+\|\mathbf{B}\mathbf{u}_\eta-\mathbf{B}\mathbf{u}_G\|_{\ell_2(\mathcal{J}_M)}\\ &\quad +\|\mathbf{g}-\mathbf{g}_\eta\|_{\ell_2(\mathcal{J}_M)}\\ &\le \eta/3+C_B\|\mathbf{u}_\eta-\mathbf{u}_G\|_{\ell_2(\mathcal{J}_X)}+\eta/3 \le \eta\end{aligned}$$

so that indeed (25) holds.

Note that the output of $\mathbf{RES}_{\mathrm{sc}}$ is not only an approximation to the residual of the Schur complement for a current approximation $\tilde{\mathbf{p}}$ of the exact solution component $\mathbf{p}$, but also a new approximation $\mathbf{u}_\eta$ of $\mathbf{u}$. Since

$$\mathbf{A}(\mathbf{u}-\mathbf{u}_G) = \mathbf{f}-\mathbf{B}^T\mathbf{p}-\mathbf{f}+\mathbf{B}^T\tilde{\mathbf{p}} = \mathbf{B}^T(\tilde{\mathbf{p}}-\mathbf{p}),$$

so that $\|\mathbf{u}-\mathbf{u}_G\|_{\ell_2(\mathcal{J}_X)} \le c_A^{-1}C_B\|\mathbf{p}-\tilde{\mathbf{p}}\|_{\ell_2(\mathcal{J}_M)}$, we conclude that

$$\|\mathbf{u}-\mathbf{u}_\eta\|_{\ell_2(\mathcal{J}_X)} \le \|\mathbf{u}-\mathbf{u}_G\|_{\ell_2(\mathcal{J}_X)} + \|\mathbf{u}_G-\mathbf{u}_\eta\|_{\ell_2(\mathcal{J}_X)} \le c_A^{-1}C_B\delta_p + \eta. \quad (62)$$

The application of $\mathbf{RES}_{\mathrm{sc}}$ in **UZAWA** is then to be understood as follows. In the first call of $\mathbf{RES}_{\mathrm{sc}}$ in the $(j+1)$st iteration block (ii) of $\mathbf{UZAWA}\,[\epsilon,\mathbf{A},\mathbf{B},\mathbf{f},\mathbf{g}] := \mathbf{SOLVE}\,[\epsilon,\mathbf{M},\mathbf{G}]$, $\tilde{\mathbf{u}} = \bar{\mathbf{u}}^j$ is the last output of $\mathbf{RES}_{\mathrm{sc}}$ in the previous jth block (respectively some initial guess when $j=0$). Likewise $\tilde{\mathbf{p}} = \bar{\mathbf{p}}^j$ is the output of (ii.2) in the preceeding iteration block. Hence, since in the first call of the $(j+1)$st block the tolerance is $\eta = \epsilon_j$, the error bounds δ_u, δ_p have, in view of (62) and Proposition 4.1, the form

$$\delta_u = (c_A^{-1}C_B + 1)\epsilon_j, \quad \delta_p = \epsilon_j. \quad (63)$$

In the following $K-1$ updates in step (ii) of **UZAWA** these bounds get tightened. In fact, denoting by $\mathbf{p}^l$ the exact Uzawa updates with $\mathbf{p}^0 = \mathbf{q}^0 := \bar{\mathbf{p}}^j$, one obviously has $\|\mathbf{p}-\mathbf{p}^l\|_{\ell_2(\mathcal{J}_M)} \le \rho^l\|\mathbf{p}-\mathbf{p}^0\| = \rho^l\|\mathbf{p}-\bar{\mathbf{p}}^j\| \le \rho^l\epsilon_j$. Hence, one concludes from (28) that

$$\|\mathbf{p}-\mathbf{q}^l\|_{\ell_2(\mathcal{J}_M)} \le \rho^l\epsilon_j + \alpha l\rho^{l-1}\epsilon_j + (\rho+\alpha l)\rho^{l-1}\epsilon_j. \quad (64)$$

Thus, for $l \ge 1$ in the $(l+1)$st call of $\mathbf{RES}_{\mathrm{sc}}$ of the $(j+1)$st iteration block, the input parameters η, δ_u, δ_p are given by

$$\begin{aligned} \eta := \rho^l\epsilon_j, \quad \delta_p := (\rho+\alpha l)\rho^{l-1}\epsilon_j, \\ \delta_u := c_A^{-1}C_B\delta_p + \rho^l\epsilon_j = \rho^{l-1}\epsilon_j(\rho + c_A^{-1}C_B(\rho+\alpha l)). \end{aligned} \quad (65)$$

A further specification is necessary when the multiplier space M is characterized by wavelet coefficient arrays in a closed subspace $\ell_{2,0}(\mathcal{J}_M) \subset \ell_2(\mathcal{J}_M)$ of finite codimension. In fact, the coarsening step in (ii.2) of **UZAWA** should preserve these linear constraints. How to modify **COARSE** to an appropriate *constrained* scheme **CCOARSE** will be exemplified later for the Stokes problem.

7. Convergence Estimates

We can now formulate the main result of this paper, see also [25].

Theorem 7.1. *Assume that the matrices* $\mathbf{A},\mathbf{B}$ *belong to* $\mathcal{C}_{s^*}$ *for some* $s^* > 0$ *and satisfy (47), (48). Then* $\mathbf{UZAWA}\,[\epsilon,\mathbf{A},\mathbf{B},\mathbf{f},\mathbf{g}]$ *produces after a finite number of steps approximations* $\bar{\mathbf{p}}(\epsilon), \bar{\mathbf{u}}(\epsilon)$, *where* $\bar{\mathbf{u}}(\epsilon)$ *is the output of step (iii) in the last call of* $\mathbf{RES}_{\mathrm{sc}}$, *satisfying for any target accuracy* $\epsilon > 0$

$$\|\mathbf{p}-\bar{\mathbf{p}}(\epsilon)\|_{\ell_2(\mathcal{J}_M)} \le \epsilon, \quad \|\mathbf{u}-\bar{\mathbf{u}}(\epsilon)\|_{\ell_2(\mathcal{J}_X)} \le C_u\epsilon. \quad (66)$$

Here $(\mathbf{u},\mathbf{p})$ *is the exact solution of (44) and* $C_u := \frac{2C_B}{5\rho c_A}\epsilon$.

Moreover, when $\mathbf{u} \in \ell^w_\tau(\mathcal{J}_X)$, $\mathbf{p} \in \ell^w_\tau(\mathcal{J}_M)$ *for* $\tau > \tau^*$ *and* $\tau^* := (s^* + 1/2)^{-1}$ *(or equivalently when* $\sigma_{N,\ell_2(\mathcal{J}_X)}(\mathbf{u}), \sigma_{N,\ell_2(\mathcal{J}_M)}(\mathbf{p})$ *decay like* N^{-s} *for* s, τ *related by (35)), then the following properties hold:*

(i) One has

$$\|\bar{\mathbf{u}}(\epsilon)\|_{\ell^w_\tau(\mathcal{J}_X)} \lesssim \|\mathbf{u}\|_{\ell^w_\tau(\mathcal{J}_X)}, \quad \|\bar{\mathbf{p}}(\epsilon)\|_{\ell^w_\tau(\mathcal{J}_M)} \lesssim \|\mathbf{p}\|_{\ell^w_\tau(\mathcal{J}_M)}.$$

(ii) The number of floating point operations needed to compute $\bar{\mathbf{u}}(\epsilon), \bar{\mathbf{p}}(\epsilon)$ *stays proportional to the supports of* $\bar{\mathbf{u}}(\epsilon), \bar{\mathbf{p}}(\epsilon)$*, i.e., for* $\epsilon \to 0$

$$\#\mathrm{supp}\,\bar{\mathbf{u}}(\epsilon) \lesssim \|\mathbf{u}\|^{1/s}_{\ell^w_\tau(\mathcal{J}_X)} \epsilon^{-1/s}, \quad \#\mathrm{supp}\,\bar{\mathbf{p}}(\epsilon) \lesssim \|\mathbf{p}\|^{1/s}_{\ell^w_\tau(\mathcal{J}_M)} \epsilon^{-1/s}.$$

Thus, in the range $0 < s < s^*$ *the asymptotic work/accuracy rate of best*N*-term approximation is recovered.*

An immediate consequence of the norm equivalences (40) and (41) are the following error estimates in the energy norm.

$$\begin{aligned} \Big\| u - \sum_{\lambda \in \mathrm{supp}\,\bar{\mathbf{u}}(\epsilon)} \bar{u}(\epsilon)_\lambda D^{-1}_{X,\lambda} \psi_{X,\lambda} \Big\|_X &\le C_X \epsilon, \\ \Big\| p - \sum_{\lambda \in \mathrm{supp}\,\bar{\mathbf{p}}(\epsilon)} \bar{p}(\epsilon)_\lambda D^{-1}_{M,\lambda} \psi_{M,\lambda} \Big\|_M &\le C_M C_u \epsilon. \end{aligned} \tag{67}$$

Proof of Theorem 7.1. By definition of **UZAWA** (60) and the previous remark that (25) holds for $\mathbf{RES}_{\mathrm{sc}}$, Proposition 4.1 already implies the first estimate in (66). As for the second estimate, $\bar{\mathbf{u}}(\epsilon) = \mathbf{u}_\eta$ is the output of step (iii) in $\mathbf{RES}_{\mathrm{sc}}$ with $\eta = \rho^{K-1}\epsilon_J$ and $\epsilon_J \le 2\epsilon$. Moreover, the accuracy of $\mathbf{q}^{K-1}$ at that stage is, on account of (64) or (65), $\delta_p = (\rho + (K-1)\alpha)\rho^{K-2}\epsilon_j \le 2(\rho + (K-1)\alpha)\rho^{K-2}\epsilon$. Therefore, the error incurred by $\bar{\mathbf{u}}(\epsilon)$ is, in view of (62), bounded by

$$c_A^{-1} C_B 2(\rho + \alpha(K-1))\rho^{K-2}\epsilon + 2\rho^{K-1}\epsilon = 2\left(c_A^{-1} C_B K + 1\right)\rho^{K-1}\epsilon \le \frac{2C_B}{5\rho c_A},$$

where we have used (27). This is the second part of (66).

As for the remaining part of the assertion, again by (60) and Theorem 4.1, the main task is to verify the validity of the properties (i) and (ii) in Theorem 4.1 for the specific residual routine $\mathbf{RES}_{\mathrm{sc}}$ defined above.

To this end, recall from [19] that the constant K is chosen so that at the jth stage the output $\mathbf{p}^K$ of step (ii.1) of **SOLVE** satisfies $\|\mathbf{p} - \mathbf{q}^K\|_{\ell_2(\mathcal{J}_M)} \le \epsilon_j/10$. Thus, in view of Proposition 4.3, the coarsening step (ii.2) in **SOLVE** ensures that $\|\mathbf{p} - \bar{\mathbf{p}}^{j+1}\|_{\ell_2(\mathcal{J}_M)} \le \epsilon_j/2$ while, in addition

$$\|\bar{\mathbf{p}}^{j+1}\|_{\ell^w_\tau(\mathcal{J}_M)} \lesssim \|\mathbf{p}\|_{\ell^w_\tau(\mathcal{J}_M)}, \tag{68}$$

with a constant independent of j. Now in the next finite block of perturbed gradient iterations, $\bar{\mathbf{p}}^{j+1}$ is the initial input in $\mathbf{RES}_{\mathrm{sc}}$. Moreover, by Proposition 4.3, step (i) of $\mathbf{RES}_{\mathrm{sc}}$ ensures that the result $\bar{\mathbf{u}}_\delta$ has a controlled ℓ_τ^w-norm, i.e. $\|\bar{\mathbf{u}}_\delta\|_{\ell_\tau^w(\mathcal{J}_X)} \lesssim \|\mathbf{u}\|_{\ell_\tau^w(\mathcal{J}_X)}$ with a constant independent of j. Since at the $(j+1)$st stage the accuracy δ_u of the input $\tilde{\mathbf{u}}$ is, by (63), $(c_A^{-1}C_B+1)\epsilon_j$ and the coarsening step increases this bound by at most a fixed constant while the target accuracy of the $(j+1)$st iteration block is, by (65), still a fixed fraction of ϵ_j, only a uniformly bounded finite number of iterations in $\mathbf{SOLVE}_{\mathrm{ell}}$ (depending on K) are required to reach that target accuracy. By definition of $\mathbf{RES}_{\mathrm{ell}}$ above, each iteration involves coarsening and applications of **APPLY** with respect to $\mathbf{A}$ and $\mathbf{B}^T$. Under the above assumptions, Theorem 6.1 ensures that the intermediate outputs $\|\mathbf{u}_\eta\|_{\ell_\tau^w(\mathcal{J}_X)}$, $\eta = \rho^l\epsilon_j$, and $\|\mathbf{q}^l\|_{\ell_\tau^w(\mathcal{J}_M)}$ stay uniformly bounded by $\|\mathbf{u}\|_{\ell_\tau^w(\mathcal{J}_X)}, \|\mathbf{p}\|_{\ell_\tau^w(\mathcal{J}_M)}$, respectively, with constants depending only on the problem parameters ρ, K. (Note that steps (ii.2) in $\mathbf{SOLVE}_{\mathrm{sc}}$ and step (i) in $\mathbf{RES}_{\mathrm{sc}}$ set these constants back to one which is independent of ρ, K when proceeding to the next iteration block). Now observe that the assumption $\mathbf{u} \in \ell_\tau^w(\mathcal{J}_X), \mathbf{p} \in \ell_\tau^w(\mathcal{J}_M)$ on the exact solutions imply, in view of (44), that $\mathbf{f} \in \ell_\tau^w(\mathcal{J}_X)$ and $\mathbf{g} \in \ell_\tau^w(\mathcal{J}_M)$. Since coarsening also does not increase ℓ_τ^w-norms, the output of $\mathbf{RES}_{\mathrm{sc}}$ stays uniformly bounded by fixed constant multiples of $\|\mathbf{u}\|_{\ell_\tau^w(\mathcal{J}_X)}, \|\mathbf{p}\|_{\ell_\tau^w(\mathcal{J}_M)}$, which confirms the validity of requirement (i) in Theorem 4.1. Requirement (ii) is a consequence of Theorem 6.1. Hence Theorem 4.1 can be applied and finishes the proof. □

The above version of **UZAWA** can be varied in several ways. Instead of coarsening the $\mathbf{u}$-component in each call of $\mathbf{RES}_{\mathrm{sc}}$, one could coarsen only the last $\mathbf{u}$-output after the Kth call of $\mathbf{RES}_{\mathrm{sc}}$ along with the $\mathbf{p}$-component in step (ii.2) of **UZAWA**. We shall briefly discuss further variations later in connection with the Stokes problem.

8. Approximation Properties and Regularity – When does Adaptivity pay?

A fundamental theme in approximation theory is to relate approximation properties to the *regularity* of the approximated function. In order to judge the performance of an adaptive scheme versus a much simpler scheme based on uniform refinements say, requires, in view of Theorem 7.1, comparing the approximation power of best N-term approximation versus approximations based on uniform refinements. As mentioned in Section 3, the latter ones are governed essentially by Sobolev regularity, i.e., regularity in L_2. Convergence rates N^{-s} on uniform grids with respect to some energy norm $\|\cdot\|_{H^t}$ are obtained

essentially if and only if the approximated function belongs to H^{t+ds}. On the other hand, the same rate can still be achieved by best N-term approximation as long as the approximated functions belong to the (much larger) *Besov* space $B_\tau^{t+sd}(L_\tau)$ where $\tau^{-1} = s + 1/2$, i.e. weakening the smoothness measure can be compensated by a nonlinear choice of degrees of freedom.

The connection of these facts with the present adaptive schemes can be summarized as follows. In fact, Proposition 4.2 and (34) say that those sequences for which $\sigma_{N,\ell_2}(\mathbf{v})$ decays like N^{-s} are (almost) those in ℓ_τ. On the other hand, ℓ_τ is directly related to regularity. Here the following version is relevant which refers to measuring the error in $\mathcal{H} = H^t$, say, see [24]. In fact, when $\mathcal{H} = H^t$, $\mathbf{D} = \mathbf{D}^t$ and $\mathbf{D}^{-t}\Psi$ is a Riesz basis for H^t, one has

$$\mathbf{u} \in \ell_\tau \quad \Longleftrightarrow \quad u = \sum_\lambda u_\lambda 2^{-t|\lambda|}\psi_\lambda \in B_\tau^{t+sd}(L_\tau(\Omega)),$$

where again $\frac{1}{\tau} = s + \frac{1}{2}$. Thus, functions in the latter space, that do not belong to H^{t+ds} can be recovered by the adaptive scheme at an *asymptotically* better rate when compared with uniform refinements. The situation is illustrated by the "DeVore diagram" in Figure 1, which indicates the topography of function spaces. It shows embedding in H^t. The larger $r = t + sd$ the bigger the gap between H^r and $B_\tau^r(L_\tau)$. The loss of regularity when moving to the right from H^r at height r is compensated by judiciously placing the degrees of freedom through nonlinear approximation. Moreover, Theorem 7.1 says that this is preserved by the adaptive scheme.

Now one might wonder whether and under which circumstances the solutions to (2) have higher *Besov-regularity* than *Sobolev-regularity* in which case the adaptive scheme would perform asymptotically better. We stress though that this kind of knowledge is not required by the adaptive scheme. For scalar elliptic problems Besov regularity has been investigated e.g. in [22,26]. The result essentially says that for rough boundaries such as Lipschitz or even polygonal boundaries the Besov-regularity of solutions is indeed higher than the relevant Sobolev-regularity, which indicates the effective use of adaptive techniques.

The analysis also shows that the quantitative performance of the adaptive scheme, compared with one based on uniform refinements, is the better the larger the H^r-norm of the solution is compared with its $B_\tau^r(L_\tau)$-norm.

Similar results can be found for the *Stokes System* [22,25,26]. We know that, if the solution $U = (u, p)$ of (14) satisfies

$$u \in B_\tau^{1+sd}(L_\tau(\Omega)), \quad p \in B_\tau^{sd}(L_\tau(\Omega)), \quad \frac{1}{\tau} = s + \frac{1}{2}, \tag{69}$$

the solution components satisfy

$$\sigma_{N,H_0^1(\Omega)}(u) \lesssim N^{-s}, \quad \sigma_{N,L_2(\Omega)}(q) \lesssim N^{-s}.$$

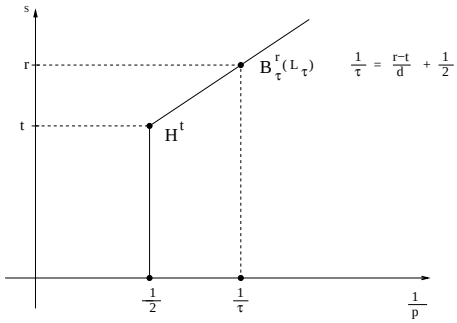

Figure 1. Embedding in H^t

Thus, again the question arises under which circumstances has the solution to the Stokes problem a high Besov regularity, which according to Theorem 7.1 and (69), would result in correspondingly high convergence rates provided by the adaptive scheme. The following result has been derived in [25].

Theorem 8.1. *For $d = 2$ the strongest singularity solutions (u_S, p_S) of the Stokes problem on an L-shaped domain in $\mathbb{R}^2$ belong to the above scale of Besov spaces for any $s > 0$. The Sobolev regularity is limited by* 1.5445, *resp.* 0.5445. *Thus arbitrarily high asymptotic rates can be obtained by adaptive schemes of correspondingly high order.*

9. Application to the Stokes Problem

We shall apply next the above concepts to the Stokes problem (14). We begin with some remarks on the choice of wavelet bases.

9.1. *B-spline Wavelets and the Exact Application of the Divergence*

As mentioned already in Section 5, the iteration (52) will generally take the form

$$\mathbf{p}^{i+1} = \mathbf{p}^i + \alpha_{i+1}\mathbf{R}(\mathbf{B}\mathbf{u}^{i+1} - \mathbf{g}).$$

The necessity of a nontrivial $\mathbf{R}$ arose in [25] for the following reason. The multiplier space $M = L_{2,0}(\Omega)$ in the case of the Stokes problem is a subspace of $L_2(\Omega)$ of codimension one. Realizing the zero mean amounts to a linear constraint on the scaling function coefficients only. However, when using the scheme **APPLY** for the wavelet representation $\mathbf{B}$ of the divergence, this linear relation is violated since the evaluation is only approximate. But since the

image of $\mathbf{B}$, say, are wavelet coefficients with respect to the dual wavelet basis $\tilde{\Psi}_M$, a different projection would be needed. In this case $\mathbf{R}$ plays the role of a Riesz-operator that maps the image under $\mathbf{B}^T$ back into coefficients of the primal basis. Since $\mathbf{R}$, as a mass matrix of the dual wavelet basis $\tilde{\Psi}_M$, is also compressible this does not jeopardize the asymptotic behavior of the scheme but requires dual bases of sufficiently high order to ensure good compressibility and introduces an additional perturbation that has to be kept track of. We are therefore interested here in alternatives that allow us to take simply $\mathbf{R} = \mathbf{I}$. We shall show that this is indeed possible by an *exact* evaluation of $\mathbf{Bv}$ based on suitable wavelet bases for velocity and pressure. In addition this helps diminishing the perturbation incurred by the approximate application of the operator $\mathbf{B}$ and thus saves the additional tolerances in step (iv) in $\mathbf{RES}_{\mathrm{sc}}$ related to the application of $\mathbf{B}$. The key point is the following theorem from [41] which we state first for wavelets on the whole real line.

Theorem 9.1. *Let $\psi, \tilde{\psi}$ be compactly supported biorthogonal wavelets on $\mathbb{R}$ such that $\psi \in H^1(\mathbb{R})$. Then, there exists another pair of biorthogonal wavelets $\psi^-, \tilde{\psi}^-$ such that*

$$\frac{d}{dx}\psi(x) = 4\psi^-(x), \qquad (-4)\tilde{\psi}(x) = \frac{d}{dx}\tilde{\psi}^-(x).$$

We shall exploit this fact in connection with spline wavelets generated by cardinal B-splines ${}_m\varphi$ of order m together with the compactly supported dual generators ${}_{m,\tilde{m}}\tilde{\varphi}$ of order $\tilde{m}$ constructed in [20] for $\tilde{m} \geq m$, $m + \tilde{m}$ even. The corresponding primal and dual wavelets are denoted by ${}_{m,\tilde{m}}\psi$, ${}_{m,\tilde{m}}\tilde{\psi}$, respectively, that satisfy $\langle {}_{m,\tilde{m}}\psi, {}_{m,\tilde{m}}\tilde{\psi}(\cdot - k)\rangle_{L_2(\mathbb{R})} = \delta_{0,k}$, $k \in \mathbb{Z}$. In these terms the above relation can be rephrased as

$$\frac{d}{dx}{}_{m,\tilde{m}}\psi(x) = 4\,{}_{m-1,\tilde{m}+1}\psi(x), \tag{70}$$

i.e. here we have $\psi = {}_{m,\tilde{m}}\psi$ and $\psi^- = {}_{m-1,\tilde{m}+1}\psi$. Moreover, we recall the following relation for the scaling functions

$$\frac{d}{dx}{}_m\varphi(x) = {}_{m-1}\varphi(x) - {}_{m-1}\varphi(x-1), \quad m > 0. \tag{71}$$

In the bivariate case we choose

$$\mathbf{\Psi}_X = \begin{pmatrix} {}_{m,\tilde{m}}\mathbf{\Psi} \otimes {}_{m-1,\tilde{m}+1}\mathbf{\Psi} \\ {}_{m-1,\tilde{m}+1}\mathbf{\Psi} \otimes {}_{m,\tilde{m}}\mathbf{\Psi} \end{pmatrix}$$

as a basis for the velocity space. To explain the notation, the first component basis ${}_{m,\tilde{m}}\mathbf{\Psi} \otimes {}_{m-1,\tilde{m}+1}\mathbf{\Psi}$ consists of the functions $\psi^1_{X,j,(k,l)}(x,y) = {}_{m,\tilde{m}}\psi^e_{j,k}(x)\,{}_{m-1,\tilde{m}+1}\psi^{e'}_{j,l}(y)$, $j \geq j_0 - 1$, $k, l \in \mathbb{Z}$, where we use the conventions $\psi^0 := \varphi, \psi^1 := \psi$, $\theta_{j,k} := 2^{j/2}\theta(2^j \cdot - k)$, $j \geq j_0$, while $\psi_{j_0-1,k} := \varphi_{j_0,k}$

and $e = e' = 0$ for $j = j_0 - 1$, $(e, e') \in \{0,1\}^2 \setminus \{(0,0)\}$, $j \geq j_0$. Likewise the basis for the pressure space has the form

$$\mathbf{\Psi}_M = {}_{m-1,\tilde{m}+1}\mathbf{\Psi}_M \otimes {}_{m-1,\tilde{m}+1}\mathbf{\Psi}_M.$$

One easily infers from (70) and (71) that the divergence operator maps any element of $\mathbf{\Psi}_X$ into a linear combination of elements of $\mathbf{\Psi}_M$. In fact, one has

$$\begin{aligned} &\tfrac{\partial}{\partial x}{}_{m,\tilde{m}}\psi^e_{j,k}(x)\,{}_{m-1,\tilde{m}+1}\psi^{e'}_{j,l}(y) = \\ &\quad = \begin{cases} 2^{j+2}\,{}_{m-1,\tilde{m}+1}\psi^e_{j,k}(x)\,{}_{m-1,\tilde{m}+1}\psi^{e'}_{j,l}(y), & \text{if } e = 1, \\ 2^j({}_{m-1}\varphi_{j,k} - {}_{m-1}\varphi_{j,k+1})\,{}_{m-1,\tilde{m}+1}\psi^{e'}_{j,l}(y), & \text{if } e = 0, \end{cases} \end{aligned} \tag{72}$$

and an analogous relation holds for the second component. Hence, abbreviating

$$\mathbf{v}^T\mathbf{\Psi}_X := \begin{pmatrix} \sum_{(k,l)\in\mathbb{Z}^2} v_{1,(k,l)}\psi^1_{X,(k,l)} \\ \sum_{(k,l)\in\mathbb{Z}^2} v_{2,(k,l)}\psi^2_{X,(k,l)} \end{pmatrix},$$

we see that

$$\nabla\cdot(\mathbf{v}^T\mathbf{\Psi}_X) = \mathbf{w}^T\mathbf{\Psi}_M, \tag{73}$$

where the scalar sequence $\mathbf{w}$ is related to $\mathbf{v}$ by (72).

9.2. *Bounded Domains – Discretization of Velocity and Pressure*

So far these relations hold on all of $\mathbb{R}^2$. Thus we need to construct bases with essentially the above properties that are adapted to bounded domains. Let us briefly recall the main ingredients for their construction that matter here. Note that for the discretization of the velocity we need vector fields with H^1-components which implies that each component has to be globally continuous. On the other hand, the pressure only belongs to L_2 so that no continuity conditions apply. This entails a somewhat different treatment of the two variables.

Let us review the relevant facts of globally continuous B-spline wavelets on general domains described e.g. in [15,16,21,34,40]. The first step is to adapt the univariate bases for $\mathbb{R}$ to the interval $[0,1]$ by appropriately modifying (in a scale invariant way) only finitely many basis functions near the end points of the interval, retaining essentially relations like (70) and (71), see [31,35]. Then tensor products as above yield bases on the unit square (or more generally on the unit d-cube). Then one considers a non-overlapping decomposition of the relevant domain Ω into M subdomains

$$\bar{\Omega} = \bigcup_{i=1}^{M} \bar{\Omega}_i, \quad \Omega_j \cap \Omega_j = \emptyset, \quad i \neq j,$$

where each subdomain Ω_i is the parametric image of the reference domain $\hat{\Omega} := (0,1)^d$ under certain smooth functions $\kappa_i : \hat{\Omega} \to \Omega_i$, i.e., $\Omega_i = \kappa_i(\hat{\Omega})$. The wavelet functions restricted to each Ω_i are the parametric liftings of the above mentioned wavelets on the parameter domain $\hat{\Omega}$ which, as explained before, are tensor products of univariate wavelets on the interval $(0,1)$ [31].

The wavelets for the velocity components should belong to $H_0^1(\Omega)$. Therefore the modifications near patch boundaries are chosen in such a way that the arising wavelets are globally continuous on Ω and that homogeneous Dirichlet boundary conditions are satisfied on $\partial\Omega$, see [15,34,40].

Note that the pressure does not need to be globally continuous so that we do not have to enforce continuity across the interelement boundaries. Neither do we have to impose any boundary conditions on the pressure.

However, as an element of $L_{2,0}(\Omega)$, the pressure has to have vanishing mean value. This can be realized by a projection of the form $P_0 : L_2(\Omega) \to L_{2,0}(\Omega)$. Since the wavelets have vanishing moments, the projector P_0 leaves the linear combination of all "true" wavelets (those basis functions having vanishing moments) in a given an expansion unchanged and affects only the scaling functions. A detailed description can be found in [25].

9.3. *Divergence Relations on Bounded Domains*

As we have seen, a wavelet basis on the parameter domain $\hat{\Omega}$ is a key ingredient for constructing bases on the union Ω of smooth parametric images of $\hat{\Omega}$. A basis on $\hat{\Omega}$ can be obtained by taking a tensor product of bases on $(0,1)$. Hence, generalizations of $(70,71)$ to $(0,1)$ are required, [46]. Slightly more general relations also hold for wavelet systems on the interval $(0,1)$ from [31], see [46]. Note that one cannot use the construction from [31] in a straightforward way but has to adjust certain parameters. We consider two bases on $(0,1)$, namely

$$\Psi_{j,e} := \{\psi_{j,e,k} : k \in \mathcal{J}_{j,e}\}, \qquad \Psi_{j,e}^- := \{\psi_{j,e,k}^- : k \in \mathcal{J}_{j,e}^-\},$$

where $e = 0$ refers to the scaling functions and $e = 1$ to the wavelets. The system $\Psi_{j,e}$ is derived from a biorthogonal system on $\mathbb{R}$ with orders $m, \tilde{m}$, as explained in Section 9.1, whereas $\Psi_{j,e}^-$ arises from a basis with the parameters $m-1$ and $\tilde{m}+1$. Here $\mathcal{J}_{j,e}$, $e = 0,1$, denotes the index set of the boundary adapted scaling functions, respectively wavelets of order $m, \tilde{m}$ on level j.

As was shown in [46], $\Psi_{j,e}$ and $\Psi_{j,e}^-$ can be constructed in such a way that $\mathcal{J}_{j,1} = \mathcal{J}_{j,1}^-$ and that there exist matrices $D_{j,e} \in \mathbb{R}^{|\mathcal{J}_{j,e}^-| \times |\mathcal{J}_{j,e}|}$, $e = 0,1$ such that

$$\frac{d}{dx}\Psi_{j,e} = D_{j,e}\,\Psi_{j,e}^-. \tag{74}$$

Equation (74) implies that $\frac{d}{dx}\Psi_{j,e} \subset S(\Psi^-_{j,e})$. The matrices are of the form

$$D_{j,e} = \begin{array}{|cc|c|}\hline D^L & & \\ \hline & D^I & \\ \hline & & D^R \\ \hline \end{array} \in \mathbb{R}^{|\mathcal{J}^-_{j,e}|\times|\mathcal{J}_{j,e}|},\ e = 0, 1.$$

Moreover, the matrices $D_{j,0}, D_{j,1}$ are sparse and depend only weakly on the level in the sense that they are just stretched and scaled when j grows. The size of the boundary blocks does not depend on j and only the interior matrix is stretched when j grows. For $e = 0$ the D^I is two-banded, for $e = 1$ it is a diagonal matrix. Denoting by $\mathcal{J}^I_{j,e}$ the indices of the 'interior' (i.e., unmodified) functions, we have

$$(D_{j,0})_{k,k'} = 2^j(\delta_{k,k'} - \delta_{k,k'-1}), \quad k \in \mathcal{J}^{I,-}_{j,0},\ \ k' \in \mathcal{J}^I_{j,0}$$

and

$$(D_{j,1})_{k,k'} = c\,2^{j+2}\delta_{k,k'}, \quad k \in \mathcal{J}^{I,-}_{j,1},\ \ k' \in \mathcal{J}^I_{j,1}.$$

The additional factor c occurs in the construction of the boundary adapted wavelets. The structure of these matrices is shown in Figure 2. The corresponding scaling functions and wavelets are displayed in Figures 3 and 4.

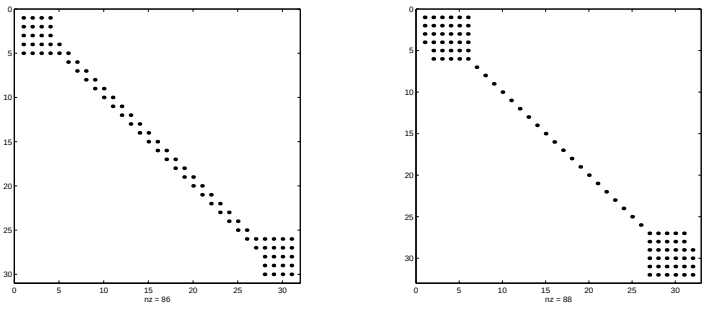

Figure 2. Matrices $D_{5,0}$ and $D_{5,1}$ for $m = \tilde{m} = 3$.

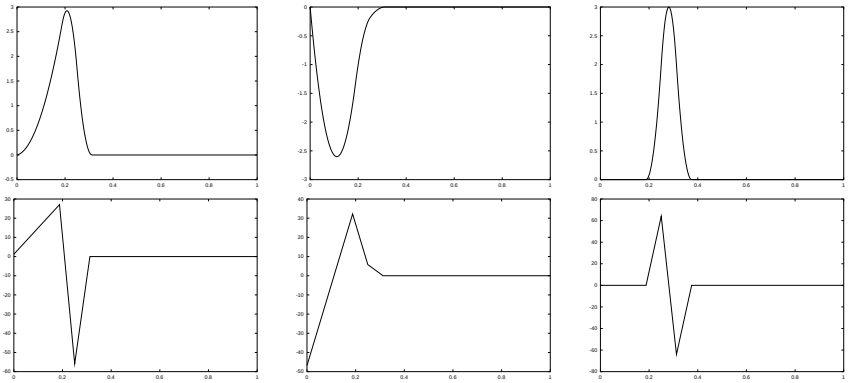

Figure 3. Scaling functions ${}_{3,3}\varphi_{4,2}$, ${}_{3,3}\varphi_{4,3}$, ${}_{3,3}\varphi_{4,4}$ (top) and their exactly computed derivatives being linear combinations of ${}_{2,4}\varphi_{4,k}$, for $k = 0, 1, 2$ (bottom).

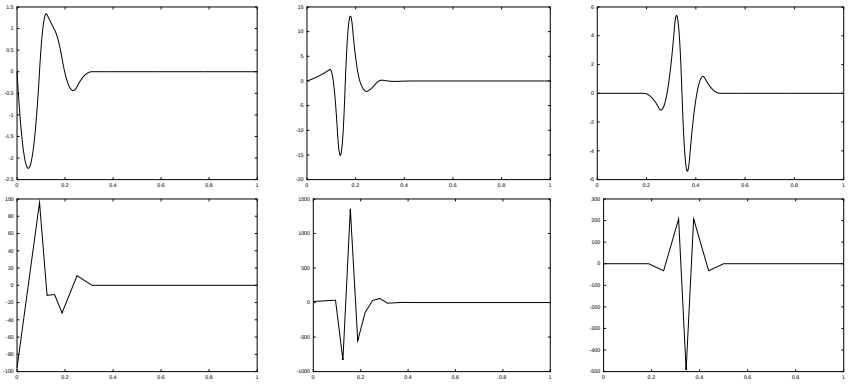

Figure 4. Scaling functions ${}_{3,3}\psi_{4,1}$, ${}_{3,3}\psi_{4,2}$, ${}_{3,3}\psi_{4,4}$ (top) and their exactly computed derivatives being linear combinations of ${}_{2,4}\psi_{4,k}$ for $k = 0, 1, 2$ (bottom).

9.4. *The Divergence Operator*

In our numerical examples Ω will be the standard L–shaped domain, obtained by removing $(-1, 0]^2$ from $(-1, 1)^2$. We continue denoting the resulting bases by $\mathbf{\Psi}_X, \mathbf{\Psi}_M$. Thus in summary, one still has the inclusion

$$(\nabla \cdot \mathbf{\Psi}_X) \subset \operatorname{span}(\Psi_M) \tag{75}$$

with an identity like (73), where the relation between $\mathbf{v}$ and $\mathbf{w}$ is, up to boundary modifications, the same as before. This leads to the following scheme for the exact evaluation of the divergence operator:

DIV $[\mathbf{v}] \to \overline{\mathbf{w}}, \Lambda$
determines for any finitely supported $\mathbf{v} \in \ell_2(\mathcal{J}_X)$ a finitely supported $\bar{\mathbf{w}} \in \ell_2(\mathcal{J}_M)$ such that $\nabla \cdot (\mathbf{v}^T \mathbf{D}_X^{-1} \mathbf{\Psi}_X) = \bar{\mathbf{w}}^T \mathbf{\Psi}_M$.

In view of the above relations and the scaling in (72), the following fact is immediate.

Remark 9.1. The output $\bar{\mathbf{w}}$ of **DIV** satisfies

$$\|\bar{\mathbf{w}}\|_{\ell_\tau^w(\mathcal{J}_M)} \lesssim \|\mathbf{v}\|_{\ell_\tau^w(\mathcal{J}_X)}, \quad \operatorname{supp} \bar{\mathbf{w}} \lesssim \operatorname{supp} \mathbf{v}.$$

Replacing $\mathbf{B}$ in step (iv) of $\mathbf{RES}_{\mathrm{sc}}$, by the routine **DIV**, gives rise to a variant of the adaptive Uzawa scheme that will be referred to as $\mathbf{UZAWA}_{\mathrm{new}}$. It will be tested below against the original version where the divergence operator is applied approximately with the aid of the **APPLY** scheme. One then has to use a nontrivial Riesz map $\mathbf{R}$ as in [25] which is the mass matrix of the dual basis $\tilde{\Psi}_M$. We shall refer to this version as $\mathbf{UZAWA}_{\mathrm{old}}$.

9.5. *Compressibility of* $\mathbf{A}$ *and* $\mathbf{B}^T$

Even when the application of $\mathbf{B}$ is replaced by **DIV** the scheme **APPLY** has to be used for $\mathbf{A}$ and $\mathbf{B}^T$. It is therefore important to know the compressibility ranges for these matrices.

Remark 9.2. By the same arguments as used in [25,7] one can show that the scaled wavelet representation $\mathbf{A}$ with respect to the above velocity basis belongs to $\mathcal{C}_{s^*}$ for $s^* = (m-5/2)/2$, noting that we are using here bases with anisotropic order and that the smaller order is $m-1$. Likewise one can show that also $\mathbf{B}^T \in \mathcal{C}_{s^*}$ with the same s^*.

In fact, for the above bases, $\mathbf{B}^T$ has essentially the same compressibility properties as the mass matrix for Ψ_M. The order of the dual basis for Ψ_M is, by construction $\tilde{m}+1$, so that $\tilde{m}_M = \tilde{m}+1 \geq m_X - 1 = m-1$ is indeed satisfied.

9.6. *Constrained Coarsening*

As has been mentioned earlier, the elements of the multiplier space M may be characterized by the fact that the arrays of wavelet coefficients belong to a space $\ell_{2,0}(\mathcal{J}_M)$ of finite codimension in $\ell_2(\mathcal{J}_M)$ and the coarsening procedure in step (ii.2) of **UZAWA** should preserve the corresponding linear constraints. As explained earlier, this is indeed the case for the Stokes problem when $\ell_{2,0}(\mathcal{J}_M)$ has codimension one corresponding to the vanishing mean of pressure functions. Since the wavelets have vanishing moments, this constraint

concerns only the scaling function coefficients. Therefore, they have to be exempt from coarsening. Thus it is to be understood throughout the following that coarsening in step (ii.2) of **UZAWA** applies only to the "true" wavelet coefficients. We shall refer to this version below as **CCOARSE**. Since there is only a finite number of basis functions on the coarsest level, this does not affect the asymptotic estimates and leaves the above analysis unchanged.

9.7. *Complexity of UZAWA*

As in [25] we shall always assume that the entries of $\mathbf{f}, \mathbf{g}$ are available and that the entries in $\mathbf{A}, \mathbf{B}^T$ are computable at unit cost, [8], see [10] for an alternative approach. By the above comments on the compressibility of $\mathbf{A}$ and $\mathbf{B}^T$ in Remark 9.2 and by Remark 9.1, we can apply Theorem 7.1 to arrive at the following conclusions, see also [25].

Corollary 9.3. *If the exact solution (u,p) of the Stokes problem (16) satisfies for some $s < (m - \frac{5}{2})/d$*

$$\inf_{\#\mathrm{supp}(\mathbf{v}) \le N} \|u - \mathbf{v}^T \mathbf{D}_{\mathbf{X}}^{-1} \mathbf{\Psi}_{\mathbf{X}}\|_{\mathbf{X}} \lesssim N^{-s}, \quad N \to \infty,$$

$$\inf_{\#\mathrm{supp}(\mathbf{q}) \le N} \|p - \mathbf{q}^T \mathbf{\Psi}_M\|_M \lesssim N^{-s}, \quad N \to \infty,$$

then the approximations $\bar{\mathbf{u}}(\varepsilon), \bar{\mathbf{p}}(\varepsilon)$, produced by by both versions $\mathbf{UZAWA}_{\mathrm{new}}$ *and* $\mathbf{UZAWA}_{\mathrm{old}}$*, satisfy*

$$\|u - \bar{\mathbf{u}}(\varepsilon)^T \mathbf{D}_{\mathbf{X}}^{-1} \mathbf{\Psi}_{\mathbf{X}}\|_{\mathbf{X}} \lesssim (\#\mathrm{supp}(\bar{\mathbf{u}}(\varepsilon)))^{-s},$$

$$\|p - \bar{\mathbf{p}}(\varepsilon)^T \mathbf{D}_M^{-1} \Psi_M\|_M \lesssim (\#\mathrm{supp}(\bar{\mathbf{p}}(\varepsilon)))^{-s}.$$

The computational work needed to compute $\bar{\mathbf{u}}(\varepsilon), \bar{\mathbf{p}}(\varepsilon)$ stays proportional to $\varepsilon^{-1/s}$.

9.8. *Some Variants of the Adaptive Scheme*

Further possible variants of the basic adaptive iteration scheme come to mind which we formulate first for the general case. For instance, the stationary Richardson iteration (R) used so far could be replaced by a a steepest descent (SD) or a conjugate gradient method (CG). Both (SD) and (CG) would relieve one from the task of obtaining a good guess for the damping parameter α in (23). The following well-known result tells us the optimal choice, see, e.g. [42].

Theorem 9.2. *Let $\mathbf{M}$ be a symmetric and positive definite operator and $\lambda_{\max}, \lambda_{\min}$ denote its largest, respectively smallest eigenvalue. Then the stationary Richardson method (24) converges if and only if $\alpha < \frac{2}{\lambda_{\max}}$. The optimal damping parameter α_{opt} is given by $\alpha_{\mathrm{opt}} = \frac{2}{\lambda_{\min} + \lambda_{\max}}$.*

It can be shown that $\rho(\mathbf{I}-\alpha_{\rm opt}\mathbf{M}) = \frac{{\rm cond}_2(\mathbf{M})-1}{{\rm cond}_2(\mathbf{M})+1}$. Hence using the optimal relaxation parameter α, the stationary Richardson iteration converges with the same rate as the steepest descent method. However, in general it is not a practically feasible task to determine $\alpha_{\rm opt}$. We therefore included the steepest descent method (SD), which is known to reduce the error for symmetric positive operators in each step also by $\rho = \frac{{\rm cond}_2(\mathbf{M})-1}{{\rm cond}_2(\mathbf{M})+1}$. Moreover, we include also the (CG) method whose exact version has an even better error reduction. We shall refer to the corresponding variants of the adaptive scheme as **SOLVE_Z** for $\mathbf{Z} \in \{\mathbf{R}, \mathbf{SD}, \mathbf{CG}\}$. However, (SD) and (CG) require an additional application of the Schur complement $\mathbf{M}$. One way of realizing such an application approximately is offered by the routine $\mathbf{APPLY}_{\rm sc}$. Given a finitely supported input $\mathbf{q}$ it determines a finitely supported output $\mathbf{z}_\eta$ satisfying $\|\mathbf{M}\mathbf{q}-\mathbf{z}_\eta\|_{\ell_2(\mathcal{J}_M)} \le \eta$ as follows:

$\mathbf{APPLY}_{\rm sc}[\eta, \mathbf{M}, \mathbf{q}] \to \mathbf{z}_\eta$
(i) $\mathbf{APPLY}\,[\eta/(3C_B), \mathbf{B}^T, \mathbf{q}] \to \mathbf{z}_1$;
(ii) $\mathbf{SOLVE}_{\rm ell}[c_A\eta/(3C_B), \mathbf{A}, \mathbf{z}_1] \to \mathbf{y}$;
(iii) $\mathbf{APPLY}\,[\eta/(3C_B), \mathbf{B}, \mathbf{q}] \to \mathbf{z}_\eta$.

As mentioned before, depending on the structure of the multiplier space M, one may have to apply in addition in (iii) a suitable Riesz map with a corresponding adjustment of accuracy tolerances. In the case of the Stokes problem we replace in our experiments the approximate application of $\mathbf{B}$ in step (iii) by the exact divergence evaluation **DIV**.

The (SD)-variant of **SOLVE** in the general case reads now as follows:

$\mathbf{SOLVE\_SD}\,[\epsilon, \mathbf{M}, \mathbf{G}] \to \bar{\mathbf{p}}(\epsilon)$

(i) Set $\bar{\mathbf{p}}^0 = \mathbf{0}$, $\epsilon_0 := c_M^{-1}\|\mathbf{G}\|_{\ell_2}$, $j = 0$.
(ii) If $\epsilon_j \le \epsilon$, stop $\bar{\mathbf{p}}^j \to \mathbf{p}(\epsilon)$. Else $\mathbf{q}^0 := \bar{\mathbf{p}}^j$.
(ii.1) For $l = 0, \ldots, \tilde{K}-1$:
$\mathbf{RES}\,[\rho^l\epsilon_j, \mathbf{M}, \mathbf{G}, \mathbf{q}^l] \to \mathbf{r}_l$;
$\mathbf{APPLY}\,[\rho^l\epsilon_j/5, \mathbf{M}, \mathbf{q}^l] \to \mathbf{w}^l$;$\alpha_l := \frac{(\mathbf{r}_l)^T\mathbf{r}_l}{(\mathbf{r}_l)^T\mathbf{w}_l}$; $\mathbf{q}^{l+1} := \mathbf{q}^l + \alpha_l\,\mathbf{r}_l$.
(ii.2) $\mathbf{COARSE}\,[\mathbf{q}^{\tilde{K}}, 2\epsilon_j/5] \to \bar{\mathbf{p}}^{j+1}$,
$\epsilon_{j+1} := \epsilon_j/2$, $j+1 \to j$ go to (ii).

Here and below $\tilde{K}$ stands for some fixed integer that can be determined from the problem constants in a similar way as before in (27), so as to ensure

that the error after $\tilde{K}$ steps is bounded by $\epsilon_j/10$.

We finally present also a prototype of a conjugate gradient iteration. Here we can expect an error reduction of $\rho = \frac{\sqrt{\kappa_2(\mathbf{M})}-1}{\sqrt{\kappa_2(\mathbf{M})}+1}$. In first numerical experiments it has performed better than **SOLVE_SD** for the solution of the elliptic problems.

SOLVE_CG $[\epsilon, \mathbf{M}, \mathbf{G}] \to \bar{\mathbf{p}}(\epsilon)$

(i) Set $\bar{\mathbf{p}}^0 = \mathbf{0}$, $\epsilon_0 := c_M^{-1}\|\mathbf{G}\|_{\ell_2}$, $j = 0$, $\xi > 1$.
(ii) If $\epsilon_j \le \epsilon$, stop $\bar{\mathbf{p}}^j \to \bar{\mathbf{p}}(\epsilon)$. Else $\mathbf{q}^0 := \bar{\mathbf{p}}^j$.

(ii.1) For $l = 0, \ldots, \tilde{K} - 1$:

(1) **RES** $[\rho^{\tilde{K}}\epsilon_j, \mathbf{M}, \mathbf{G}, \mathbf{q}^l] \to \mathbf{d}^{l,0}$;
(2) $\mathbf{r}^{l,0} := -\mathbf{d}^{l,0}$;
(3) $\mathbf{q}^{l,0} := \mathbf{q}^l$;
(4) For $i = 0, \ldots, J - 1$:
(a) **APPLY** $[\xi^i \rho^l \epsilon_j, \mathbf{M}, \mathbf{d}^{l,i}] \to \mathbf{w}^{l,i}$;
(b) $\lambda_{l,i} := \frac{(\mathbf{r}^{l,i})^T \mathbf{r}^{l,i}}{(\mathbf{r}^{l,i})^T \mathbf{w}^{l,i}}$;
(c) $\mathbf{q}^{l,i+1} := \mathbf{v}^{l,i} + \lambda_{l,i}\, \mathbf{d}^{l,i}$
(d) $\mathbf{r}^{l,i+1} := \mathbf{r}^{l,i} + \lambda_{l,i}\, \mathbf{w}^{l,i}$
(e) If $i < J$:
i. $\beta_{l,i} := \frac{(\mathbf{r}^{l,i+1})^T \mathbf{r}^{l,i+1}}{(\mathbf{r}^{l,i})^T \mathbf{r}^{l,i}}$;
ii. $\mathbf{d}^{l,i+1} := -\mathbf{r}^{l,i+1} + \beta_{l,i}\, \mathbf{d}^{l,i}$ go to (ii).1.
(5) $\mathbf{q}^{l+1} := \mathbf{q}^{l,J}$;

(ii.2) **COARSE** $[\mathbf{q}^{\tilde{K}}, 2\epsilon_j/5] \to \bar{\mathbf{p}}^{j+1}$,
$\epsilon_{j+1} := \epsilon_j/2$, $j + 1 \to j$ go to (ii).

Here we assume to work with the Euclidean inner product. One can think of a more general setting where the Euclidean inner product is replaced by another inner product $[\cdot, \cdot]$, which again may be evaluated only approximately. See [19] for an example. A complete error analysis of **SOLVE_CG** is more involved and goes beyond the scope of this paper since, in addition to approximating scalar quantities like the step size α in (SD), the directions are, of course, not exact. These issue will be addressed elsewhere in more detail.

We can now formulate, for the special case of the Stokes problem, corresponding residual approximation schemes $\mathbf{RES}_{\text{sc\_div\_Z}}[\eta, \mathbf{M}, \mathbf{G}, \tilde{\mathbf{p}}, \tilde{\mathbf{u}}, \delta_p, \delta_u] \to (\mathbf{r}_\eta, \mathbf{u}_\eta)$, employing the exact evaluation **DIV** of the divergence operator in

step (iv) of $\mathbf{RES}_{\mathrm{sc}}$ and using the variant **SOLVE_Z** for $\mathbf{Z} \in \{\mathbf{R}, \mathbf{SD}, \mathbf{CG}\}$ in step (ii) of $\mathbf{RES}_{\mathrm{sc}}$. Employing still the Richardson scheme as an outer Uzawa iteration, a typical variant reads as follows:

UZAWA_R_Z $[\epsilon, \mathbf{A}, \mathbf{B}, \mathbf{f}, \mathbf{g}] \to \big(\bar{\mathbf{u}}(\epsilon), \bar{\mathbf{p}}(\epsilon)\big)$

(i) Set $\bar{\mathbf{u}}^0 = \mathbf{0},\ \bar{\mathbf{p}}^0 = \mathbf{0},\ \ \epsilon_0 := c_M^{-1}\|\mathbf{G}\|_{\ell_2},\ \ j = 0.$

(ii) If $\epsilon_j \le \epsilon$, stop $(\bar{\mathbf{u}}^j, \bar{\mathbf{p}}^j) \to (\mathbf{u}(\epsilon), \mathbf{p}(\epsilon))$. Else $\mathbf{w}^0 := \bar{\mathbf{u}}^j,\ \mathbf{q}^0 := \bar{\mathbf{p}}^j.$

(ii.1) For $l = 0, \ldots, \tilde{K} - 1$:

If $l = 0$ set $\delta_p = \epsilon_j,\ \delta_u = (c_A^{-1} C_B + 1)\epsilon_j$;
If $l > 0$ set $\delta_u = (\rho + \alpha\, l)\rho^{l-1}\epsilon_j,\ \delta_p = c_A^{-1} C_B \delta_p + \rho^l \epsilon_j$;
$\mathbf{RES}_{\mathrm{uz\_div\_Z}}\,[\rho^l \epsilon_j, \mathbf{M}, \mathbf{G}, \mathbf{q}^l, \mathbf{w}^l, \delta_p, \delta_u] \to (\mathbf{r}_l, \mathbf{w}^{l+1})$;
$\mathbf{q}^{l+1} := \mathbf{q}^l + \alpha\, \mathbf{r}_l.$

(ii.2) $\mathbf{CCOARSE}\,[\mathbf{q}^{\tilde{K}}, 2\epsilon_j/5] \to \bar{\mathbf{p}}^{j+1}$;
$\mathbf{COARSE}\,[\mathbf{w}^{\tilde{K}}, 2\epsilon_j/5] \to \bar{\mathbf{u}}^{j+1}$;
$\epsilon_{j+1} := \epsilon_j/2,\ j+1 \to j$ go to (ii).

Here $\tilde{K}$ is chosen so that $\mathbf{q}^{\tilde{K}}$ deviates from $\mathbf{p}$ in the $(j+1)$st iteration block by at most $\epsilon_j/10$ and also $\|\mathbf{u} - \mathbf{w}^{\tilde{K}}\|_{\ell_2(\mathcal{J}_X)} \le \epsilon_j/10$. Moreover, α is the damping parameter from (23) giving rise to an error reduction $\rho < 1$ and c_M is the constant from (22). Note that we have included also the coarsening of the u component in step (ii.2), which means that step (i) in $\mathbf{RES}_{\mathrm{sc\_div\_Z}}$ can be removed. The arguments for the complexity analysis stay essentially the same but now the final accuracy bound for $\bar{\mathbf{u}}(\epsilon)$ is ϵ instead of $C_u\epsilon$, see (66). Of course, one could think of using (SD) or (CG) also for the outer Uzawa iteration.

10. Numerical Experiments

In this section, we present some numerical experiments for some of the variants presented above. We focus on two main objectives. First, we wish to compare the new algorithm $\mathbf{UZAWA}_{\mathrm{new}}$ (with exact application of $\mathbf{B}$ through the scheme **DIV**) with the previous method $\mathbf{UZAWA}_{\mathrm{old}}$ presented in [25], which involves the Riesz map $\mathbf{R}$. Since both versions have the same asymptotic behavior, we are in particular, interested in possible quantitative gains of one version against the other one. The second objective is to compare the quantitative efficiency of the Richardson iteration (R) with the steepest descent iteration (SD) described above. The numerical experiments have been carried

out using tools from [5]. A detailed description of routines concerning sparse vectors can be found in [7]. All tests have been performed with the choice $m = \tilde{m} = 3$.

10.1. *Experiment 1*

We wish to compare $\mathbf{UZAWA}_{\text{new}}$ with $\mathbf{UZAWA}_{\text{old}}$ from [25]. To this end, we use the same data as in [25], i.e., we consider the L–shaped domain $\Omega = (-1,1)^2 \backslash (-1,0]^2$. The results from regularity theory stated in Theorem 8.1 guarantee the existence of solutions to the Stokes problem that have higher Besov- than Sobolev-regularity, see also [25,39,44]. As in [25], we choose the data in such a way that the exact solution components displayed in Figure 5 are exactly those *singularity functions* mentioned in Theorem 8.1. For a detailed description of the data, we refer to [25]. In the following tables, we record the

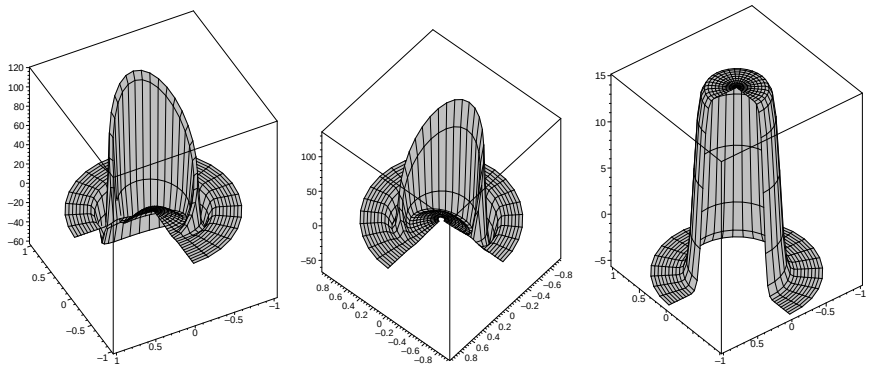

Figure 5. Exact solution for Experiment 1, first and second component of the velocity (left and center) and pressure (right).

following quantities for $\mathbf{x} \in \{\mathbf{u}^1, \mathbf{u}^2, \mathbf{p}\}$

$$\rho_{\mathbf{x}} := \frac{\|\mathbf{x} - \mathbf{x}_\Lambda\|_{\ell_2}}{\|\mathbf{x} - \mathbf{x}_{\#\Lambda}\|_{\ell_2}}, \quad r_{\mathbf{x}} := \frac{\|\mathbf{x} - \mathbf{x}_\Lambda\|_{\ell_2}}{\|\mathbf{x}\|_{\ell_2}},$$

i.e., $r_{\mathbf{x}}$ is the relative error and $\rho_{\mathbf{x}}$ is the ratio of the error of the current numerical approximation to the best N-term approximation. The corresponding results are listed in Table 1. In order to have a fair comparison, we used the Richardson iteration (R) with $\alpha = 1.3$ corresponding to the choice in [25]. The computed numerical approximations are displayed in Figure 7.

The comparison with the results reported in [25] is shown in Figure 6. We see a slightly different behavior of the velocity and the pressure. For the velocity,

Table 1. Numerical results for Example 1. Number of active coefficients, ratio of the error of the numerical approximation and the best N-term approximation and relative error for the first velocity component and the pressure. The results for the second velocity components are similar.

It	δ	$\#\Lambda_{\mathbf{u}^1}$	$\rho_{\mathbf{u}^1}$	$r_{\mathbf{u}^1}$	$\#\Lambda_{\mathbf{p}}$	$\rho_{\mathbf{p}}$	$r_{\mathbf{p}}$
0	26.1028	1	1	0.9704	868	1.15	0.1215
1	13.0514	1	1	0.9671	869	1.05	0.0540
2	6.5257	34	1	0.4900	868	1.02	0.0337
3	3.2629	58	1	0.2552	870	1.02	0.0279
4	1.6314	94	1	0.1302	869	1.02	0.0268
5	0.8157	152	1.01	0.0683	892	1.01	0.0210
6	0.4079	340	1.04	0.0352	995	1.00	0.0098

the new method always performs quantitatively better. The slope of the two curves showing the convergence history is the same for both methods after the first few iterations. This is expected since both methods are asymptotically optimal. For the pressure, the method from [25] is slightly better only for the first two iterations while from then on the new version appears to be superior.

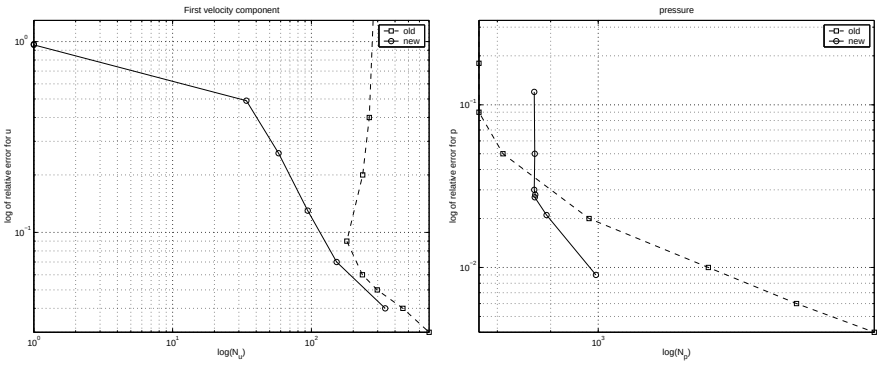

Figure 6. Comparison of the previous results (referred to as 'old') with the new ones for the first velocity component (left) and the pressure (right).

10.2. *Experiment 2*

In this experiment we compare the stationary Richardson iteration (R) with the steepest descent iteration (SD). To this end, we compare 10 successive iterations of both methods without any coarsening. In Table 2 we see that the convergence of the stationary Richardson iteration for small α is very poor.

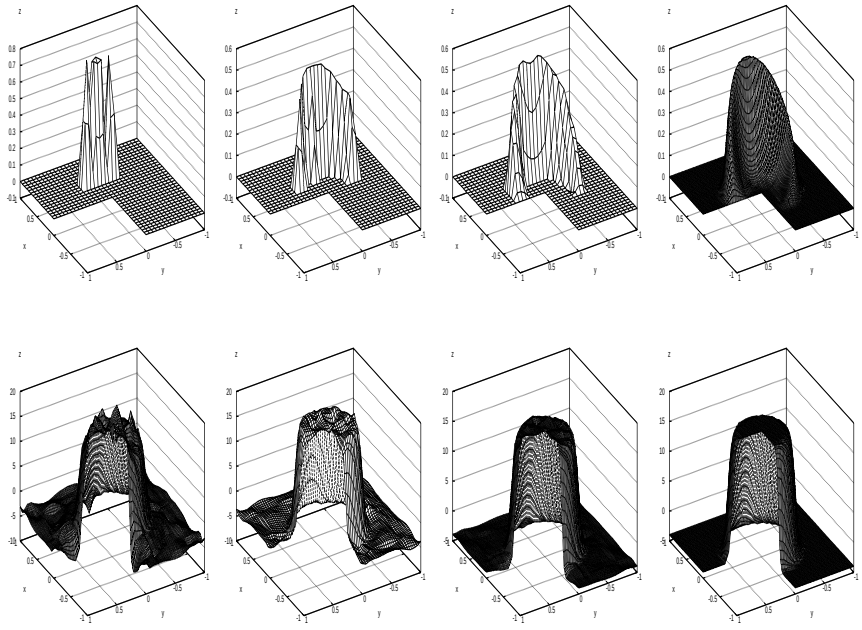

Figure 7. Numerical approximations in Example 1 for the first velocity component (top) and the pressure (bottom) for the iterations $i = 1, 2, 3, 6$.

For α too large, the iteration will not converge at all. The values indicate however that there is relatively large range for α for which the Richardson iteration converges approximatively as fast as the steepest descent method. So in practice, having a good guess for α at hand, there is no gain in using the steepest descent method since it requires an additional application of the Schur complement $\mathbf{M}$ per step.

Acknowledgments

This work has been supported in part by the SFB 401 funded by the German Research Foundation and by the TMR Network "Wavelets in Numerical Simulation" funded by the European Commission.

References

1. A. Averbuch, G. Beylkin, R. Coifman, and M. Israeli, *Multiscale inversion of elliptic operators*, in: Signal and Image Representation in Combined Spaces, J. Zeevi and R. Coifman (eds.), Academic Press, 1995.

Table 2. Comparison of Richardson iteration with different choices of α and steepest descent method (SD). The table shows the relative errors $r_{\mathbf{p}}$ with $\mathbf{p}^0 = 0$.

	$r_{\mathbf{p}}$ for stationary Richardson iteration with different α							$r_{\mathbf{p}}$
It	2.0	1.7	1.3	1.0	0.7	0.5	0.1	SD
1	0.8302	0.6194	0.4401	0.4308	0.5423	0.6565	0.9271	0.4706
2	0.7819	0.5422	0.3799	0.3260	0.3403	0.4387	0.8558	0.3353
3	0.8041	0.4150	0.2471	0.2404	0.2641	0.3306	0.7905	0.2409
4	0.8156	0.3088	0.1857	0.1872	0.2176	0.2684	0.7311	0.1888
5	0.7941	0.2265	0.1483	0.1559	0.1833	0.2254	0.6773	0.1561
6	0.7931	0.1645	0.1136	0.1240	0.1534	0.1938	0.6286	0.1238
7	0.7841	0.1326	0.0983	0.1020	0.1315	0.1690	0.5844	0.0910
8	0.7837	0.1012	0.0813	0.0872	0.1130	0.1491	0.5446	0.0822
9	0.7785	0.0910	0.0751	0.0780	0.0996	0.1326	0.5086	0.0726
10	0.7770	0.0824	0.0723	0.0724	0.0889	0.1190	0.4762	0.0713

2. I. Babuška and W.C. Rheinboldt, *Error estimates for adaptive finite element computations*, SIAM J. Numer. Anal. **15** (1978), 736–754.
3. E. Bänsch, P. Morin, and R.H. Nochetto, *An adaptive Uzawa FEM for Stokes: Convergence without the inf-sup*, WIAS Berlin, preprint, 2001.
4. R.E. Bank and A. Weiser, *Some a posteriori error estimates for elliptic partial differential equations*, Math. Comp., **44** (1985), 283–301.
5. A. Barinka, T. Barsch, K. Urban, and J. Vorloeper, *The Multilevel Library: Software Tools for Multiscale Methods and Wavelets, Version 2.1, Documentation*, RWTH Aachen, IGPM Preprint 199, 2001.
6. A. Barinka, Fast evaluation tools for adaptive wavelet schemes, PhD Thesis, RWTH Aachen, in preparation.
7. A. Barinka, T. Barsch, P. Charton, A. Cohen, S. Dahlke, W. Dahmen, and K. Urban, *Adaptive wavelet schemes for elliptic problems – Implementation and numerical experiments*, SIAM J. Scient. Comput. **23**, No. 3 (2001), 910–939.
8. A. Barinka, S. Dahlke, W. Dahmen, *Adaptive Application of Operators in Standard Representation*, in preparation.
9. S. Bertoluzza, *A-posteriori error estimates for wavelet Galerkin methods*, Appl. Math. Lett., **8** (1995), 1–6.
10. S. Bertoluzza, C. Canuto, and K. Urban, *On the adaptive computation of integrals of wavelets*, Appl. Numer. Math. **34** (2000), 13–38.
11. G. Beylkin, R. Coifman, and V. Rokhlin, Fast wavelet transforms and numerical algorithms. I. Commun. Pure Appl. Math. **44**, No. 2 (1991), 141–183.
12. G. Beylkin and 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, and P. Oswald (eds.), Academic Press, 1997, 137–197.
13. F. Bornemann, B. Erdmann, and R. Kornhuber, *A posteriori error estimates for elliptic problems in two and three space dimensions*, SIAM J. Numer. Anal., **33** (1996), 1188–1204.
14. F. Brezzi and M. Fortin, Mixed and Hybrid Finite Element Methods, Springer, 1991.
15. C. Canuto, A. Tabacco, and K. Urban, *The wavelet element method, part I:*

Construction and analysis, Appl. Comp. Harm. Anal., **6** (1999), 1–52.
16. C. Canuto, A. Tabacco, and K. Urban, *The wavelet element method, part II: Realization and additional features in 2D and 3D*, Appl. Comp. Harm. Anal., **8** (2000), 123–165.
17. A. Cohen, *Wavelet methods in numerical analysis*, in: Handbook of Numerical Analysis, vol. VII, P.-G. Ciarlet and J.-L. Lions (eds.), Elsevier, Amsterdam, 2000, 417–711.
18. A. Cohen, W. Dahmen, and R. DeVore, *Adaptive wavelet methods for elliptic operator equations – Convergence rates*, Math. Comp., **70** (2001), 27–75.
19. A. Cohen, W. Dahmen, and R. DeVore, *Adaptive wavelet methods II - Beyond the elliptic case*, IGPM Report # 199, RWTH Aachen, 2000, to appear in Found. Comput. Math.
20. A. Cohen, I. Daubechies, and J.-C. Feauveau, *Biorthogonal bases of compactly supported wavelets*, Comm. Pure and Appl. Math., **45** (1992), 485–560.
21. A. Cohen and R. Masson, *Wavelet adaptive methods for second order elliptic problems: Boundary conditions and domain decomposition*, Numer. Math., **86** (2000), 193–238.
22. S. Dahlke, *Besov regularity for elliptic boundary value problems on polygonal domains*, Appl. Math. Lett., **12** (1999), 31–36.
23. S. Dahlke, W. Dahmen, R. Hochmuth, and R. Schneider, *Stable multiscale bases and local error estimation for elliptic problems*, Appl. Numer. Math. **8** (1997), 21–47.
24. S. Dahlke, W. Dahmen, and R. DeVore, *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, London, 1997, 237–283.
25. S. Dahlke, W. Dahmen, and K. Urban, *Adaptive wavelet methods for saddle point problems – Convergence rates*, IGPM Report # 204, RWTH Aachen, 2001, to appear in SIAM J. Numer. Anal.
26. S. Dahlke and R. DeVore, *Besov regularity for elliptic boundary value problems*, Comm. Part. Diff. Eq., **22** (1997), 1–16.
27. S. Dahlke, R. Hochmuth, and K. Urban, *Adaptive wavelet methods for saddle point problems*, Math. Model. Numer. Anal. (M2AN), **34(5)** (2000), 1003–1022.
28. W. Dahmen, *Wavelet and Multiscale Methods for Operator Equations*, Acta Numerica, **6** (1997), 55–228.
29. W. Dahmen, *Wavelet methods for PDEs – Some recent developments*, J. Comp. Appl. Math., **128** (2001), 133–185.
30. W. Dahmen, H. Harbrecht, and R. Schneider *Compression Techniques for Boundary Integral Equations — Optimal Complexity Estimates*, IGPM Report # 213, RWTH Aachen, 2002.
31. 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.
32. W. Dahmen, A. Kunoth, and K. Urban, *Wavelets in Numerical Analysis and their Quantitative Properties*, in: Surface Fitting and Multiresolution Methods, A. Le Mehaute, C. Rabut, and L.L. Schumaker (eds.), Vanderbilt University Press, Nashville, TN, 1997, 93–130.

33. W. Dahmen, S. Prößdorf, and R. Schneider, *Multiscale methods for pseudo-differential equations on smooth closed manifolds*, in: Wavelets: theory, algorithms, and applications, C.K. Chui, L. Montefusco, and L. Puccio (eds.), Academic Press 1994, 385–424.
34. W. Dahmen and R. Schneider, *Composite Wavelet Bases for Operator Equations*, Math. Comp. **68** (1999), 1533–1567.
35. W. Dahmen and R. Schneider, Wavelets with Complementary Boundary Conditions - Function Spaces on the Cube, Results in Mathematics, 34 (1998), 255–293.
36. W. Dörfler, *A convergent adaptive algorithm for Poisson's equation*, SIAM J. Numer. Anal., **33** (1996), 1106–1124.
37. K. Eriksson, D. Estep, P. Hansbo, and C. Johnson, *Introduction to adaptive methods for differential equations*, Acta Numerica **4** (1995), 105–158.
38. V. Girault and P.-A. Raviart, Finite Element Methods for Navier-Stokes Equations, Springer, 1986.
39. P. Grisvard, Singularities in Boundary Value Problems, Research Notes in Applied Mathematics **22**, Springer-Verlag, 1992.
40. H. Harbrecht, *Wavelet Galerkin Schemes for the Boundary Element Method in Three Dimensions*, PhD thesis, Technical University of Chemnitz, 2001.
41. P.G. Lemarié-Rieusset, *Analyses multi-résolutions non orthogonales, Commutation entre Projecteurs et Derivation et Ondelettes Vecteurs à divergence nulle* (in french), Rev. Mat. Iberoam. **8** (1992), 221-236.
42. H. Meurant. *Computer Solution of Large Linear Systems.* North–Holland, Amsterdam, 1999.
43. E. Novak, *On the power of adaptation*, J. Complexity **12** (1996), 199–237.
44. J. Osborn, *Regularity of solutions to the Stokes problem in a polygonal domain*, in: Symposium on Numerical Solutions of Partial Differential Equations III, B. Hubbart (ed.), Academic Press, 1975, 393-411.
45. T. v. Petersdorff and C. Schwab, *Wavelet approximations for first kind boundary integral equations on polygons*, Numer. Math. **74**, No. 4 (1996), 479–516.
46. K. Urban, *Wavelet bases in* $\mathbf{H}(\mathrm{div})$ *and* $\mathbf{H}(\mathbf{curl})$, Math. Comp. **70** (2001), 739-766.
47. K. Urban, Wavelets in Numerical Simulation, Springer 2002.