

A ROBUST MULTIGRID METHOD FOR THE TIME-DEPENDENT STOKES PROBLEM

STEFAN TAKACS

ABSTRACT. In the present paper we propose an all-at-once multigrid method for generalized Stokes flow problems. Such problems occur as subproblems in implicit time-stepping approaches for time-dependent Stokes problems. The discretized optimality system is a large scale linear system whose condition number depends on the grid size of the spacial discretization and of the length of the time step. Recently, for this problem an all-at-once multigrid method has been proposed, where in each smoothing step the Poisson problem has to be solved (approximatively) for the pressure field. In the present paper, we propose an all-at-once multigrid method where the solution of such subproblems is not needed. We prove that the proposed method shows robust convergence behavior in the grid size of the spacial discretization and of the length of the time-step.

1. INTRODUCTION

In the present paper, we consider the following model problem (*generalized Stokes flow problem*). Let $\Omega \subseteq \mathbb{R}^d$ be a bounded domain with $d \in \{2, 3\}$ and assume $f \in [L^2(\Omega)]^d$ and $g \in L^2(\Omega)$ to be given. Find a velocity field u and a pressure distribution p such that

$$(1.1) \quad \begin{aligned} -\Delta u + \beta u + \nabla p &= f \text{ in } \Omega, \\ \nabla \cdot u &= g \text{ in } \Omega, \\ u &= 0 \text{ on } \partial\Omega \end{aligned}$$

is satisfied. $\beta > 0$ is assumed to be a given parameter. To obtain existence and uniqueness of the solution, we further require $\int_{\Omega} p \, dx = \int_{\Omega} g \, dx = 0$.

The problem (1.1) appears as auxiliary problem for some implicit time-stepping approach to solve an incompressible, time-dependent Stokes flow problem. In this case, the parameter β is proportional to the inverse of the length of the time-step, scaled by a viscosity parameter.

The discretization of the problem leads to an indefinite linear system with saddle-point structure. The main goal of this work is to construct and to analyze numerical methods that produce an approximate solution to the problem, where the computational complexity can be bounded by the number of unknowns times a constant which is independent of the grid level (of the spacial discretization) and the choice of β , in particular for large values of β .

For the solution of such a saddle point problem, there are several possibilities. In [5, 11, 13, 14, 17, 25], various kinds of preconditioners have been proposed for this

Last modification: May 24, 2013.

2010 *Mathematics Subject Classification*. Primary 65N55 65N22 76D07 65N30.

The research was funded by the Austrian Science Fund (FWF): J3362-N25.

problem which can be combined with an Krylov-space method as outer iteration scheme to an iterative solver for the problem.

An alternative are all-at-once multigrid methods. Such methods are on the one hand typically quite fast, on the other hand using those methods one does not need an outer iteration scheme. Originally, multigrid methods have been designed and analyzed for elliptic problems. They also work well for saddle point problems (like the generalized Stokes problem). For $\beta = 0$, problem (1.1) is the standard Stokes problem. For this case several multigrid solvers are available, see, e.g., [18, 21, 6, 4, 24] and the papers cited in [23, 15]. The construction of a multigrid method for $\beta > 0$, particularly if the method should show robust convergence behavior in β , is more involved, see [12] for an overview and numerical results. Recently, a robust all-at-once multigrid method has been proposed, see [15]. In the named paper, for each step of the smoothing iteration, the Poisson problem has to be solved (approximatively) for the pressure distribution.

The goal of the present paper is to drop this requirement, i.e., we present an all-at-once multigrid method where the solution of such sub-problems is not needed. We prove that the proposed method is robust in the grid size of the spacial discretization and in the choice of β .

In this paper we present a convergence proof for the proposed multigrid method based on the classical splitting of the analysis into smoothing property and approximation property, see [9].

This paper is organized as follows. In Section 2 we will introduce the variational formulation and discuss its discretization. In Section 3 we will introduce an all-at-once multigrid approach. A convergence proof is given in Section 4. Numerical results which illustrate the convergence result will be presented in Section 5. In Section 6 we will close with conclusions.

2. VARIATIONAL FORMULATION AND DISCRETIZATION

Here and in what follows, $L^2(\Omega)$ and $H^1(\Omega)$ denote the standard Lebesgue and Sobolev spaces with associated standard norms $\|\cdot\|_{L^2(\Omega)}$ and $\|\cdot\|_{H^1(\Omega)}$, respectively. The space $L_0^2(\Omega)$ is the space of all L^2 -functions with mean value 0, i.e.,

$$L_0^2(\Omega) := \left\{ v \in L^2(\Omega) : \int_{\Omega} v \, d\xi = 0 \right\}.$$

The space $H_0^1(\Omega)$ is the space of all functions in $H^1(\Omega)$, vanishing on the boundary. Both spaces are equipped with standard norms, i.e., $\|\cdot\|_{L_0^2(\Omega)} := \|\cdot\|_{L^2(\Omega)}$ and $\|\cdot\|_{H_0^1(\Omega)} := \|\cdot\|_{H^1(\Omega)}$.

Using these spaces, we can set up the variational formulation of (1.1), which reads as follows. Find $u \in U := [H_0^1(\Omega)]^d$ and $p \in P := L_0^2(\Omega)$ such that

$$\begin{aligned} (\nabla u, \nabla \tilde{u})_{L^2(\Omega)} + \beta(u, \tilde{u})_{L^2(\Omega)} + (p, \nabla \cdot \tilde{u})_{L^2(\Omega)} &= (f, \tilde{u})_{L^2(\Omega)} \\ (\nabla \cdot u, \tilde{p})_{L^2(\Omega)} &= (g, \tilde{p})_{L^2(\Omega)} \end{aligned}$$

holds for all $\tilde{u} \in U$ and $\tilde{p} \in P$. Certainly, the variational problem can be rewritten as one variational equation as follows. Find $x \in X$ such that

$$(2.1) \quad \mathcal{B}(x, \tilde{x}) = \mathcal{F}(\tilde{x}) \quad \text{for all } \tilde{x} \in X,$$

where $X := U \times P$ and

$$\begin{aligned}\mathcal{B}((u, p), (\tilde{u}, \tilde{p})) &:= (\nabla u, \nabla \tilde{u})_{L^2(\Omega)} + \beta(u, \tilde{u})_{L^2(\Omega)} + (p, \nabla \cdot \tilde{u})_{L^2(\Omega)} + (\nabla \cdot u, \tilde{p})_{L^2(\Omega)}, \\ \mathcal{F}(\tilde{u}, \tilde{p}) &:= (f, \tilde{u})_{L^2(\Omega)} + (g, \tilde{p})_{L^2(\Omega)}.\end{aligned}$$

We are interested in finding an approximative solution for equation (2.1). Both, the proposed solution strategy and the convergence analysis, follow the abstract framework introduced in [20]. The conditions, **(A1)**, **(A1a)**, **(A3)** and **(A4)**, mentioned in the present paper are the same conditions as in [20].

For simplicity, we introduce the following notation.

Notation 2.1. Throughout this paper, $C > 0$ is a generic constant, independent of the grid level k and the choice of the parameter β . For any scalars a and b , we write $a \lesssim b$ (or $b \gtrsim a$) if there is a constant $C > 0$ such that $a < Cb$. We write $a \approx b$ if $a \lesssim b \lesssim a$.

Let the Hilbert spaces X , U and P (introduced above) be equipped with the following norms:

$$\begin{aligned}\|x\|_X^2 &:= \|(u, p)\|_X^2 := \|u\|_U^2 + \|p\|_P^2, \\ \|u\|_U^2 &:= \|u\|_{H^1(\Omega)}^2 + \beta\|u\|_{L^2(\Omega)}^2 \text{ and} \\ \|p\|_P^2 &:= \sup_{0 \neq w \in [H_0^1(\Omega)]^d} \frac{(p, \nabla \cdot w)_{L^2(\Omega)}^2}{\|w\|_{H^1(\Omega)}^2 + \beta\|w\|_{L^2(\Omega)}^2}.\end{aligned}$$

Lemma 2.1 in [15] states the following result.

(A1): The relation

$$\|x\|_X \lesssim \sup_{0 \neq \tilde{x} \in X} \frac{\mathcal{B}(x, \tilde{x})}{\|\tilde{x}\|_X} \lesssim \|x\|_X$$

holds for all $x \in X$.

Using the following notation, we can express the norms in a nicer way.

Notation 2.2. For any Hilbert space A , the symbol A^* denotes its dual space equipped with the dual norm

$$\|u\|_{A^*} := \sup_{0 \neq w \in A} \frac{\langle u, w \rangle}{\|w\|_A},$$

where $\langle u, \cdot \rangle := u(w)$ denotes the duality pairing.

For any Hilbert space A and any scalar $a > 0$, the symbol aA denotes the space on the underlying set of the Hilbert space A equipped with the norm

$$\|u\|_{aA}^2 := a\|u\|_A^2.$$

For any two Hilbert spaces A and B , the symbol $A \cap B$ denotes the space on the intersection of the underlying sets, $\{u \in A \cap B\}$, equipped with the norm

$$\|u\|_{A \cap B}^2 := \|u\|_A^2 + \|u\|_B^2,$$

and the symbol $A + B$ denotes the space on the algebraic sum of the underlying sets, $\{u_1 + u_2 : u_1 \in A, u_2 \in B\}$, equipped with the norm

$$\|u\|_{A+B}^2 := \inf_{u_1 \in A, u_2 \in B, u = u_1 + u_2} \|u_1\|_A^2 + \|u_2\|_B^2.$$

The spaces A^* , aA , $A \cap B$ and $A + B$ are Hilbert spaces. The fact that A^* is a Hilbert space follows directly from the Riesz representation theorem, see, e.g., Theorem 1.2 in [1]. The fact that aA is a Hilbert space is obvious and for the latter two see, e.g., Lemma 2.3.1 in [3].

We immediately see that the norm on U can be rewritten as follows

$$\|u\|_U = \|u\|_{H^1(\Omega) \cap \beta L^2(\Omega)}.$$

To reformulate the norm on P , we need a regularity assumption.

(R): *Regularity of the Stokes problem.* Let $f \in [L^2(\Omega)]^d$ and $g \in H_0^1(\Omega) \cap L_0^2(\Omega)$ be arbitrarily but fixed and $(u, p) \in [H_0^1(\Omega)]^d \times L_0^2(\Omega)$ be the solution of the Stokes problem, i.e., such that

$$\begin{aligned} (\nabla u, \nabla \tilde{u})_{L^2(\Omega)} + (p, \nabla \tilde{u})_{L^2(\Omega)} &= (f, \tilde{u})_{L^2(\Omega)} \\ (\nabla u, \tilde{p})_{L^2(\Omega)} &= (g, \tilde{p})_{L^2(\Omega)} \end{aligned}$$

holds for all $(\tilde{u}, \tilde{p}) \in [H_0^1(\Omega)]^d \times L_0^2(\Omega)$.

Then $(u, p) \in [H^2(\Omega)]^d \times H^1(\Omega)$ and

$$\|u\|_{H^2(\Omega)}^2 + \|p\|_{H^1(\Omega)}^2 \lesssim \|f\|_{L^2(\Omega)}^2 + \|g\|_{H^1(\Omega)}^2.$$

Lemma 2.3. *Condition (R) is satisfied for convex polygonal domains.*

Proof. Theorem 2 in [10] states (in the notation of the present paper) that provided $f \in [L^2(\Omega)]^2$, $g \in H^1(\Omega) \cap L_0^2(\Omega)$ and $\delta^{-1}g \in L^2(\Omega)$ that

$$\|u\|_{H^2(\Omega)}^2 + \|\nabla p\|_{L^2(\Omega)}^2 \lesssim \|f\|_{L^2(\Omega)}^2 + \|\nabla g\|_{L^2(\Omega)}^2 + \|\delta^{-1}g\|_{L^2(\Omega)}^2,$$

is satisfied, where $\delta : \Omega \rightarrow \mathbb{R}$ is the distance to the closest vertex.

Lemma 2 in [10] states that $\|\delta^{-1}g\|_{L^2(\Omega)} \lesssim \|g\|_{H^1(\Omega)}$ is satisfied for all $g \in H_0^1(\Omega)$.

Combining these results, we obtain that

$$\|u\|_{H^2(\Omega)}^2 + \|\nabla p\|_{L^2(\Omega)}^2 \lesssim \|f\|_{L^2(\Omega)}^2 + \|g\|_{H^1(\Omega)}^2$$

is satisfied for all $f \in [L^2(\Omega)]^2$ and $g \in H_0^1(\Omega) \cap L_0^2(\Omega)$. As $p \in L_0^2(\Omega)$ was assumed, Poincaré's inequality states that $\|p\|_{H^1(\Omega)} \lesssim \|\nabla p\|_{L^2(\Omega)}$, which finishes the proof. \square

Note the fact that we assume g to satisfy homogeneous Dirichlet boundary conditions. This condition can be weakened but it is not possible to drop such a condition completely, cf. [10].

Lemma 2.4. *If (R) is satisfied, then*

$$\|p\|_P \approx \|p\|_{L^2(\Omega) + \beta^{-1}H^1(\Omega)}$$

holds for all $p \in L_0^2(\Omega)$.

For a proof of this lemma, see Theorem 3.2 in [16], which requires a regularity assumption which is weaker than assumption (R).

The discretization of problem (2.1) is done using standard finite element techniques. We assume to have for $k = 0, 1, 2, \dots$ a sequence of grids obtained by uniform refinement. On each grid level k , we discretize the problem using the Galerkin approach, i.e., we have finite dimensional spaces $X_k \subseteq X$ and consider the following problem. Find $x_k \in X_k$ such that

$$(2.2) \quad \mathcal{B}(x_k, \tilde{x}_k) = \mathcal{F}(\tilde{x}_k) \quad \text{for all } \tilde{x}_k \in X_k.$$

Using a nodal basis, we can represent this problem in matrix-vector notation as follows:

$$(2.3) \quad \mathcal{A}_k \underline{x}_k = \underline{f}_k,$$

where

$$\mathcal{A}_k = \begin{pmatrix} A_k & B_k^T \\ B_k & 0 \end{pmatrix}, \quad \underline{x}_k = \begin{pmatrix} u_k \\ p_k \end{pmatrix}, \quad \underline{f}_k = \begin{pmatrix} f_k \\ g_k \end{pmatrix}$$

and the matrices A_k and B_k represent the scalar products $a(u, \tilde{u}) := (\nabla u, \nabla \tilde{u})_{L^2(\Omega)} + \beta(u, \tilde{u})_{L^2(\Omega)}$ and $b(u, \tilde{p}) := (\nabla \cdot u, \tilde{p})_{L^2(\Omega)}$, respectively.

Here and in what follows, any underlined quantity, like \underline{x}_k , is the representation of the corresponding non-underlined quantity, here x_k , with respect to a nodal basis of the corresponding Hilbert space, here X_k .

The next step is to show existence and uniqueness of the discretized problem, which is guaranteed by the following condition.

(A1a): The relation

$$\|x_k\|_X \lesssim \sup_{0 \neq \tilde{x}_k \in X_k} \frac{\mathcal{B}(x_k, \tilde{x}_k)}{\|\tilde{x}_k\|_X} \lesssim \|x_k\|_X$$

holds for all $x_k \in X_k$.

To achieve this condition, we have to choose a discretization which is stable for the Stokes flow problem.

Lemma 2.5. *Let $X_k := U_k \times P_k$, where $U_k \subseteq H^1(\Omega)$, $P_k \subseteq H_0^1(\Omega)$ such that the weak inf-sup condition*

$$(2.4) \quad \sup_{0 \neq u_k \in U_k} \frac{(\nabla \cdot u_k, p_k)_{L^2(\Omega)}}{\|u_k\|_{L^2(\Omega)}} \gtrsim \|\nabla p_k\|_{L^2(\Omega)}$$

holds for all $p_k \in P_k$. Then condition (A1a) is satisfied.

For a proof, see Lemma 2.2 in [15].

Note that condition (2.4) is a standard condition which guarantees that the chosen discretization is stable for the Stokes problem. In [2, 22] it was shown that condition (S) is satisfied for the Taylor-Hood element ($P1 - P2$ -element) for polygonal domains where at least one vertex of each element is located in the interior of the domain. Here and in what follows we assume that the problem is discretized with the Taylor-Hood element and that the mesh satisfies the named condition.

Note that due to the fact that the grids are obtained by uniform refinement, the discrete subsets are nested, i.e., $U_k \subseteq U_{k+1}$ and $P_k \subseteq P_{k+1}$. Therefore, also $X_k \subseteq X_{k+1}$ holds.

3. AN ALL-AT-ONCE MULTIGRID METHOD

The problem shall be solved using an all-at-once multigrid method. The abstract algorithm for solving the discretized equation (2.3) on grid level k reads as follows. Starting from an initial approximation $\underline{x}_k^{(0)}$, one iterate of the multigrid method is given by the following two steps:

- *Smoothing procedure:* Compute

$$(3.1) \quad \underline{x}_k^{(0,m)} := \underline{x}_k^{(0,m-1)} + \hat{\mathcal{A}}_k^{-1} \left(\underline{f}_k - \mathcal{A}_k \underline{x}_k^{(0,m-1)} \right) \quad \text{for } m = 1, \dots, \nu$$

with $\underline{x}_k^{(0,0)} = \underline{x}_k^{(0)}$. The choice of the smoother (or, in other words, of the preconditioning matrix $\hat{\mathcal{A}}_k^{-1}$) will be discussed below.

- *Coarse-grid correction:*

- Compute the defect $\underline{r}_k^{(1)} := \underline{f}_k - \mathcal{A}_k \underline{x}_k^{(0,\nu)}$ and restrict it to grid level $k-1$ using an restriction matrix I_k^{k-1} :

$$\underline{r}_{k-1}^{(1)} := I_k^{k-1} \left(\underline{f}_k - \mathcal{A}_k \underline{x}_k^{(0,\nu)} \right).$$

- Solve the coarse-grid problem

$$(3.2) \quad \mathcal{A}_{k-1} \underline{p}_{k-1}^{(1)} = \underline{r}_{k-1}^{(1)}$$

approximatively.

- Prolongate \underline{p}_{k-1} to the grid level k using an prolongation matrix I_{k-1}^k and add the result to the previous iterate:

$$\underline{x}_k^{(1)} := \underline{x}_k^{(0,\nu)} + I_{k-1}^k \underline{p}_{k-1}^{(1)}.$$

As we have assumed to have nested spaces, the intergrid-transfer matrices I_{k-1}^k and I_k^{k-1} can be chosen in a canonical way: I_{k-1}^k is the canonical embedding and the restriction I_k^{k-1} is its properly scaled transpose.

If the problem on the coarser grid is solved exactly (two-grid method), the coarse-grid correction is given by

$$(3.3) \quad \underline{x}_k^{(1)} := \underline{x}_k^{(0,\nu)} + I_{k-1}^k \mathcal{A}_{k-1}^{-1} I_k^{k-1} \left(\underline{f}_k - \mathcal{A}_k \underline{x}_k^{(0,\nu)} \right).$$

In practice, the problem (3.2) is approximatively solved by applying one step (V-cycle) or two steps (W-cycle) of the multigrid method, recursively. On grid level $k=0$ the problem (3.2) is solved exactly.

To construct a multigrid convergence result based on Hackbusch's splitting of the analysis into smoothing property and approximation property, we have to introduce an appropriate framework.

Convergence is shown on the spaces X_k , which are equipped with an L^2 -like norms $\|\cdot\|_k$, which are defined as follows.

$$\|\underline{x}_k\|_k^2 := \|\underline{x}_k\|_{\mathcal{L}_k}^2 := (\mathcal{L}_k \underline{x}_k, \underline{x}_k)_{\ell^2},$$

where

$$(3.4) \quad \mathcal{L}_k := \begin{pmatrix} (\beta + h_k^{-2}) M_{U,k} & \\ & h_k^{-2} (\beta + h_k^{-2})^{-1} M_{P,k} \end{pmatrix},$$

where the matrices $M_{U,k}$ and $M_{P,k}$ are mass-matrices, representing the L^2 -inner product in U_k and P_k , respectively.

The smoothing property and the approximation property read as follows.

- *Smoothing property:*

$$(3.5) \quad \sup_{\tilde{x}_k \in X_k} \frac{\mathcal{B} \left(\underline{x}_k^{(0,\nu)} - \underline{x}_k^*, \tilde{x}_k \right)}{\|\tilde{x}_k\|_k} \leq \eta(\nu) \|\underline{x}_k^{(0)} - \underline{x}_k^*\|_k$$

should hold for some function $\eta(\nu)$ with $\lim_{\nu \rightarrow \infty} \eta(\nu) = 0$. Here and in what follows, $\underline{x}_k^* \in X_k$ is the exact solution of the discretized problem (2.2).

- *Approximation property:*

$$(3.6) \quad |||x_k^{(1)} - x_k^*|||_k \leq C_A \sup_{\tilde{x}_k \in X_k} \frac{\mathcal{B}(x_k^{(0,\nu)} - x_k^*, \tilde{x}_k)}{|||\tilde{x}_k|||_k}$$

should hold for some constant $C_A > 0$.

It is easy to see that, if we combine both conditions, we obtain

$$|||x_k^{(1)} - x_k^*|||_k \leq q(\nu) |||x_k^{(0)} - x_k^*|||_k,$$

where $q(\nu) = C_A \eta(\nu)$, i.e., that the two-grid method converges for ν large enough. The convergence of the W-cycle multigrid method can be shown under mild assumptions, see e.g. [9].

The choice of an appropriate smoother is a key issue in constructing an all-at-once multigrid method for an indefinite problem. In this paper, we introduce two kinds of smoothers. The first smoother is appropriate for a large class of problems including the model problem: the *normal equation smoother*, cf. [7], which reads as follows.

$$\underline{x}_k^{(0,m)} := \underline{x}_k^{(0,m-1)} + \tau \underbrace{\mathcal{L}_k^{-1} \mathcal{A}_k \mathcal{L}_k^{-1}}_{\hat{\mathcal{A}}_k^{-1}} \left(\underline{f}_k - \mathcal{A}_k \underline{x}_k^{(0,m-1)} \right) \quad \text{for } m = 1, \dots, \nu.$$

Here, a fixed $\tau > 0$ has to be chosen such that the spectral radius $\rho(\tau \hat{\mathcal{A}}_k^{-1} \mathcal{A}_k)$ is bounded away from 2 on all grid levels k and for all choices of the parameter β .

Note that also the distributive smoother which was proposed in [15] for the generalized Stokes flow problem was a normal equation smoother. There, the normal equation was taken in another Hilbert space. Therefore, the application of the inverse of the matrix, which represents the scalar product on the Hilbert space, to some given vector \underline{w}_k (in our notation: solving the linear system $\mathcal{L}_k \underline{v}_k = \underline{w}_k$), a Poisson problem has to be solved for the pressure distribution.

Using a standard inverse inequality, one can show that

$$|||x_k|||_X \lesssim |||x_k|||_k$$

is satisfied for all $x_k \in X_k$. Based on this result, using an eigenvalue analysis one can show the following lemma, cf. [7].

Lemma 3.1. *The damping parameter $\tau > 0$ can be chosen independent of grid level k and the choice of the parameter β such that*

$$\tau \rho(\hat{\mathcal{A}}_k^{-1} \mathcal{A}_k) \leq 2 - \epsilon < 2,$$

holds for some constant $\epsilon > 0$. For this choice of τ , there is a constant $C_S > 0$, independent of the grid level k and the choice of the parameter β , such that the smoothing property (3.5) is satisfied with rate $\eta(\nu) := C_S \nu^{-1/2}$.

Certainly, the iteration procedure (3.5) should be efficient-to-apply. Using the fact, that the mass matrices $M_{U,k}$ and $M_{P,k}$ in (3.4) and their diagonals are spectrally equivalent under weak assumptions, for the practical realization of the normal equation smoother these mass matrices can be replaced by their diagonals.

The second smoother, which we propose, is a *Uzawa type smoother*, cf. [19]. Here, one step of the smoother to compute $\underline{x}_k^{(0,m)} = (\underline{u}_k^{(0,m)}, \underline{p}_k^{(0,m)})$ based on $\underline{x}_k^{(0,m-1)} =$

$(\underline{u}_k^{(0,m-1)}, \underline{p}_k^{(0,m-1)})$ reads as follows:

$$\begin{aligned}\underline{u}_k^{(0,m-1/2)} &:= \underline{u}_k^{(0,m-1)} + \tau \hat{A}_k^{-1} \left(\underline{f}_k - A_k \underline{u}_k^{(0,m-1)} - B_k^T \underline{p}_k^{(0,m-1)} \right) \\ \underline{p}_k^{(0,m)} &:= \underline{p}_k^{(0,m-1)} - \sigma \hat{S}_k^{-1} \left(\underline{g}_k - B_k \underline{u}_k^{(0,m-1/2)} \right) \\ \underline{u}_k^{(0,m)} &:= \underline{u}_k^{(0,m-1)} + \tau \hat{A}_k^{-1} \left(\underline{f}_k - A_k \underline{u}_k^{(0,m-1)} - B_k^T \underline{p}_k^{(0,m)} \right),\end{aligned}$$

where \hat{A}_k and \hat{S}_k are the (1,1)-block and the (2,2)-block of \mathcal{L}_k , respectively.

The smoother can be rewritten in the compact notation (3.1), where

$$\hat{\mathcal{A}}_k := \begin{pmatrix} \tau^{-1} \hat{A}_k & B_k^T \\ B_k & \tau B_k \hat{A}_k^{-1} B_k^T - \sigma^{-1} \hat{S}_k \end{pmatrix}.$$

Here, the smoothing property can be shown using the theory introduced in [19].

Lemma 3.2. *Let \hat{A} and \hat{S} be as above. Then $\tau > 0$ and $\sigma > 0$ can be chosen independent of the grid level and the choice of the parameter β such that*

$$\tau^{-1} \hat{A} \geq A \quad \text{and} \quad \sigma^{-1} \hat{S} \geq \tau B_k \hat{A}_k^{-1} B_k^T.$$

For this choice of τ and σ , there is a constant $C_S > 0$, independent of the grid level k and the choice of the parameter β , such that the smoothing property (3.5) is satisfied with rate $\eta(\nu) := C_S \nu^{-1/2}$.

Proof. The fact that $\tau > 0$ and $\sigma > 0$ can be chosen independent of the grid level and the choice of the parameter β follow from standard inverse inequalities.

For the smoothing property, we apply Theorem 4 in [19] with the choice $\mathcal{K}_k := \mathcal{L}_k^{-1/2} \mathcal{A}_k \mathcal{L}_k^{-1/2}$ and $\hat{\mathcal{K}}_k := \mathcal{L}_k^{-1/2} \hat{\mathcal{A}}_k \mathcal{L}_k^{-1/2}$. This immediately implies

$$\|\mathcal{L}_k^{-1/2} \mathcal{A}_k (I - \hat{\mathcal{A}}_k \mathcal{A}_k)^\nu \mathcal{L}_k^{-1/2}\|_{\ell^2} \leq \eta(\nu) \|\mathcal{L}_k^{-1/2} \mathcal{A}_k \mathcal{L}_k^{-1/2}\|_{\ell^2}.$$

As $\underline{x}_k^T \mathcal{A}_k \underline{\tilde{x}}_k = \mathcal{B}(x_k, \tilde{x}_k) \lesssim \|x_k\|_X \|\tilde{x}_k\|_X \lesssim \|x_k\|_k \|\tilde{x}_k\|_k \approx \|\mathcal{L}_k^{1/2} \underline{x}_k\|_{\ell^2} \|\mathcal{L}_k^{1/2} \underline{\tilde{x}}_k\|_{\ell^2}$ holds for all $x_k, \tilde{x}_k \in X_k$, we obtain $\|\mathcal{L}_k^{-1/2} \mathcal{A}_k \mathcal{L}_k^{-1/2}\|_{\ell^2} \lesssim 1$, which finishes the proof. \square

4. A PROOF OF THE APPROXIMATION PROPERTY

The proof of the approximation property is done using the approximation theorem introduced in [20] which requires besides the conditions **(A1)** and **(A1a)** two more conditions (conditions **(A3)** and **(A4)**) involving, besides the Hilbert space X , two more Hilbert spaces $X_{-,k} := (X_-, \|\cdot\|_{X_{-,k}})$ and $X_{+,k} := (X_+, \|\cdot\|_{X_{+,k}})$, which are chosen as follows.

As weaker space we choose $X_- := U_- \times P_-$, where $U_- := [L^2(\Omega)]^d$ and $P_- := [H_0^1(\Omega) \cap L_0^2(\Omega)]^*$. These Hilbert spaces are equipped with norms

$$\begin{aligned}\|x\|_{X_{-,k}}^2 &:= \|(u, p)\|_{X_{-,k}}^2 := \|u\|_{U_{-,k}}^2 + \|p\|_{P_{-,k}}^2, \\ \|u\|_{U_{-,k}}^2 &:= h_k^{-2} \|u\|_{[L^2(\Omega) + \beta^{-1} H_0^1(\Omega)]^*}^2 \quad \text{and} \\ \|p\|_{P_{-,k}}^2 &:= h_k^{-2} \|p\|_{[H_0^1(\Omega) \cap \beta L_0^2(\Omega)]^*}^2.\end{aligned}$$

Note that dual spaces are $(X_-)^* := (U_-)^* \times (P_-)^*$, where $(U_-)^* = [L^2(\Omega)]^d$ and $(P_-)^* = H_0^1(\Omega) \cap L_0^2(\Omega)$, equipped with norms

$$\begin{aligned}\|\mathcal{F}\|_{(X_-,k)^*}^2 &= \|(f, g)\|_{(X_-,k)^*}^2 := \|f\|_{(U_-,k)^*}^2 + \|g\|_{(P_-,k)^*}^2, \\ \|f\|_{(U_-,k)^*}^2 &= h_k^2 \|f\|_{L^2(\Omega) + \beta^{-1} H_0^1(\Omega)}^2 \quad \text{and} \\ \|g\|_{(P_-,k)^*}^2 &= h_k^2 \|g\|_{H_0^1(\Omega) \cap \beta L_0^2(\Omega)}^2.\end{aligned}$$

As stronger space we choose $X_+ := U_+ \times P_+$, where $U_+ := [H^2(\Omega) \cap H_0^1(\Omega)]^d$ and $P_+ := H^1(\Omega) \cap L_0^2(\Omega)$, equipped with norms

$$\begin{aligned}\|x\|_{X_+,k}^2 &:= \|(u, p)\|_{X_+,k}^2 := \|u\|_{U_+,k}^2 + \|p\|_{P_+,k}^2, \\ \|u\|_{U_+,k}^2 &:= h_k^2 \|u\|_{H^2(\Omega) \cap \beta H^1(\Omega)}^2 \quad \text{and} \\ \|p\|_{P_+,k}^2 &:= h_k^2 \|p\|_{H^1(\Omega) + \beta^{-1} H^2(\Omega)}^2.\end{aligned}$$

The additional conditions read as follows.

(A3): On all grid levels k , the approximation error result

$$\inf_{x_k \in X_k} \|x - x_k\|_X \lesssim \|x\|_{X_+,k} \quad \text{for all } x \in X_+$$

is satisfied.

(A4): For all grid levels k and all $\mathcal{F} \in (X_-)^*$, the solution $x_{\mathcal{F}} \in X$ of the problem,

$$(4.1) \quad \text{find } x \in X \text{ such that} \quad \mathcal{B}(x, \tilde{x}) = \mathcal{F}(\tilde{x}) \quad \text{for all } \tilde{x} \in X,$$

satisfies $x_{\mathcal{F}} \in X_+$ and the inequality

$$\|x_{\mathcal{F}}\|_{X_+,k} \lesssim \|\mathcal{F}\|_{(X_-,k)^*}.$$

Based on these assumptions, the following theorem shows the approximation property.

Theorem 4.1. *Let for $k = 0, 1, 2, \dots$ the symmetric matrices \mathcal{A}_k be obtained by discretizing problem (2.2) using a sequence of finite-dimensional nested subspaces $X_{k-1} \subseteq X_k \subset X$. Assume that there are Hilbert spaces $X_+ \subseteq X \subseteq X_-$ with mesh-dependent norms $\|\cdot\|_{X_+,k}$, $\|\cdot\|_X$ and $\|\cdot\|_{X_-,k}$ such that the conditions **(A1)**, **(A1a)**, **(A3)** and **(A4)** are satisfied. Then the coarse-grid correction (3.3) satisfies the approximation property*

$$(4.2) \quad \|x_k^{(1)} - x_k^*\|_{X_-,k} \leq C_A \sup_{\tilde{x}_k \in X_k} \frac{\mathcal{B}(x_k^{(0,\nu)} - x_k^*, \tilde{x}_k)}{\|\tilde{x}_k\|_{X_-,k}},$$

where C_A only depends on the constants that appear (implicitly) in the named conditions.

For a proof, see [20], Theorem 4.1.

Theorem 4.2. *Condition **(A3)** is satisfied.*

Proof. Note that it suffices to show approximation error results for the individual variables separately. Using a standard interpolation operator $\Pi_k : [H^2(\Omega)]^d \rightarrow U_k$, we obtain for the velocity field

$$\|u - \Pi_k u\|_{L^2(\Omega)}^2 \lesssim h_k^2 \|u\|_{H^1(\Omega)}^2 \quad \text{and} \quad \|u - \Pi_k u\|_{H^1(\Omega)}^2 \lesssim h_k^2 \|u\|_{H^2(\Omega)}^2,$$

for all $u \in [H^2(\Omega)]^d$ and therefore

$$\begin{aligned} \inf_{u_k \in U_k} \|u - u_k\|_U^2 &\leq \|u - \Pi_k u\|_U^2 = \|u - \Pi_k u\|_{H^1(\Omega)}^2 + \beta \|u - \Pi_k u\|_{L^2(\Omega)}^2 \\ &\lesssim h_k^2 \left(\|u\|_{H^2(\Omega)}^2 + \beta \|u\|_{H^1(\Omega)}^2 \right) = \|u\|_{U_{+,k}}^2. \end{aligned}$$

Also for the pressure distribution we can do a similar estimate. The estimates

$$\inf_{p_k \in P_k} \|p - p_k\|_{L^2(\Omega)}^2 \lesssim h_k^2 \|p\|_{H^1(\Omega)}^2 \quad \text{and} \quad \inf_{p_k \in P_k} \|p - p_k\|_{H^1(\Omega)}^2 \lesssim h_k^2 \|p\|_{H^2(\Omega)}^2$$

are standard approximation error results which imply

$$\begin{aligned} \inf_{p_k \in P_k} \|p - p_k\|_P^2 &= \inf_{p_k \in P_k} \|p - p_k\|_{L^2(\Omega) + \beta^{-1} H^1(\Omega)}^2 \\ &= \inf_{\substack{p_k \in P_k \\ q_1 \in L^2(\Omega) \\ q_2 \in H^1(\Omega) \\ q_1 + q_2 = p - p_k}} \|q_1\|_{L^2(\Omega)}^2 + \|q_2\|_{\beta^{-1} H^1(\Omega)}^2 \\ &= \inf_{\substack{p_1 \in L^2(\Omega) \\ p_2 \in H^1(\Omega) \\ p_1 + p_2 = p}} \inf_{p_{1,k} \in P_k} \|p_1 - p_{1,k}\|_{L^2(\Omega)}^2 + \inf_{p_{2,k} \in P_k} \|p_2 - p_{2,k}\|_{\beta^{-1} H^1(\Omega)}^2 \\ &\leq \inf_{\substack{p_1 \in H^1(\Omega) \\ p_2 \in H^2(\Omega) \\ p_1 + p_2 = p}} \inf_{p_{1,k} \in P_k} \|p_1 - p_{1,k}\|_{L^2(\Omega)}^2 + \inf_{p_{2,k} \in P_k} \|p_2 - p_{2,k}\|_{\beta^{-1} H^1(\Omega)}^2 \\ &\lesssim h_k^2 \inf_{\substack{p_1 \in H^1(\Omega) \\ p_2 \in H^2(\Omega) \\ p_1 + p_2 = p}} \|p_1\|_{H^1(\Omega)}^2 + \|p_2\|_{\beta^{-1} H^2(\Omega)}^2 = h_k^2 \|p\|_{H^1(\Omega) + \beta^{-1} H^2(\Omega)}^2, \end{aligned}$$

for all $p \in H^1(\Omega) \cap L_0^2(\Omega)$, which finishes the proof. \square

Using the regularity assumption **(R)**, we can show the following lemma.

Lemma 4.3. *Suppose that the assumption **(R)** is satisfied. Let $\mathcal{F} \in (X_-)^*$. Then, $x_{\mathcal{F}}$, the solution of (4.1), satisfies $x_{\mathcal{F}} \in X_+$.*

Proof. Let $\mathcal{F}(\tilde{y}, \tilde{p}) := (f, \tilde{u})_{L^2(\Omega)} + (g, \tilde{p})_{L^2(\Omega)}$ for $f \in [L^2(\Omega)]^d$ and $g \in H_0^1(\Omega) \cap L_0^2(\Omega)$. Rewrite the problem as follows:

$$\begin{aligned} (\nabla u_{\mathcal{F}}, \nabla \tilde{u})_{L^2(\Omega)} + (p_{\mathcal{F}}, \nabla \cdot \tilde{u})_{L^2(\Omega)} &= (f, \tilde{u})_{L^2(\Omega)} - \beta (u_{\mathcal{F}}, \tilde{u})_{L^2(\Omega)} \\ (\nabla \cdot u_{\mathcal{F}}, \tilde{p})_{L^2(\Omega)} &= (g, \tilde{p})_{L^2(\Omega)} \end{aligned}$$

for all $\tilde{u} \in U$ and $\tilde{p} \in P$. As $f - \beta u_{\mathcal{F}} \in [L^2(\Omega)]^d$ and $g \in H_0^1(\Omega) \cap L_0^2(\Omega)$, we obtain using regularity assumption **(R)** immediately that $x_{\mathcal{F}} \in X_+$. \square

Note that the combination of the argument used in the proof of this lemma and condition **(A1)** immediately implies $\|x_{\mathcal{F}}\|_{X_{+,k}} \leq C(\beta) \|\mathcal{F}\|_{(X_{-,k})^*}$, where $C(\beta)$ is a constant depending on β . For showing a robust estimate, we need to do some more work. First, we need some preliminary results. For the next lemma we need, besides regularity assumption **(R)**, the following standard regularity assumption for the Poisson problem with homogeneous Neumann boundary conditions.

(R1): Regularity of the Poisson problem. Let $g \in L^2(\Omega)$ and $p \in H^1(\Omega) \cap L_0^2(\Omega)$ be such that

$$(\nabla p, \nabla \tilde{p})_{H^1(\Omega)} = (g, \tilde{p})_{L^2(\Omega)} \quad \text{for all } \tilde{p} \in H^1(\Omega) \cap L_0^2(\Omega).$$

Then $p \in H^2(\Omega)$ and $\|p\|_{H^2(\Omega)} \lesssim \|g\|_{L^2(\Omega)}$.

Such a regularity assumption can be guaranteed for convex polygonal domains (see, e.g., [8]).

Lemma 4.4. *Suppose that the assumptions **(R)** and **(R1)** are satisfied. Let $\mathcal{F} \in (X_-)^*$ and let $x_{\mathcal{F}} = (u_{\mathcal{F}}, p_{\mathcal{F}}) \in X_+$ be the solution of (4.1). Then $p_{\mathcal{F}}$ satisfies the estimate*

$$\|p_{\mathcal{F}}\|_{P_+,k}^2 \lesssim \|u_{\mathcal{F}}\|_{U_+,k}^2 + \|\mathcal{F}\|_{(X_-,k)^*}^2.$$

Proof. Let $\mathcal{F}(\tilde{y}, \tilde{p}) := (f, \tilde{u})_{L^2(\Omega)} + (g, \tilde{p})_{L^2(\Omega)}$ for $f \in [L^2(\Omega)]^d$ and $g \in H_0^1(\Omega) \cap L_0^2(\Omega)$. Let $f_2 \in [H_0^1(\Omega)]^d$ be arbitrarily but fix and set $f_1 := f - f_2 \in [L^2(\Omega)]^d$. We have that

$$(p_{\mathcal{F}}, \nabla \cdot \tilde{u})_{L^2(\Omega)} = -(\nabla u_{\mathcal{F}}, \nabla \tilde{u})_{L^2(\Omega)} - \beta(u_{\mathcal{F}}, \tilde{u})_{L^2(\Omega)} + (f_1 + f_2, \tilde{u})_{L^2(\Omega)}$$

is satisfied for all $\tilde{u} \in [H_0^1(\Omega)]^d$. This is equivalent to

$$-(\nabla p_{\mathcal{F}}, \tilde{u})_{L^2(\Omega)} = (\Delta u_{\mathcal{F}}, \tilde{u})_{L^2(\Omega)} - \beta(u_{\mathcal{F}}, \tilde{u})_{L^2(\Omega)} + (f_1 + f_2, \tilde{u})_{L^2(\Omega)}$$

for all $\tilde{u} \in [H_0^1(\Omega)]^d$. Note that $\nabla p_{\mathcal{F}}, \Delta u_{\mathcal{F}}, u_{\mathcal{F}}, f_1$ and f_2 are in $[L^2(\Omega)]^d$. Therefore, the above statement holds for all $\tilde{u} \in [L^2(\Omega)]^d$ (because $[H_0^1(\Omega)]^d$ is dense in $[L^2(\Omega)]^d$), particularly for $\tilde{u} := \nabla \tilde{p}$, where $\tilde{p} \in H^1(\Omega) \cap L_0^2(\Omega)$. So, we obtain

$$(4.3) \quad -(\nabla p_{\mathcal{F}}, \nabla \tilde{p})_{L^2(\Omega)} = (\Delta u_{\mathcal{F}}, \nabla \tilde{p})_{L^2(\Omega)} - \beta(u_{\mathcal{F}}, \nabla \tilde{p})_{L^2(\Omega)} + (f_1 + f_2, \nabla \tilde{p})_{L^2(\Omega)}$$

for all $\tilde{p} \in H^1(\Omega) \cap L_0^2(\Omega)$.

Let $p_1 \in H^1(\Omega) \cap L_0^2(\Omega)$ be such that

$$(4.4) \quad -(\nabla p_1, \nabla \tilde{p})_{L^2(\Omega)} = (\Delta u_{\mathcal{F}}, \nabla \tilde{p})_{L^2(\Omega)} + (f_1, \nabla \tilde{p})_{L^2(\Omega)}$$

holds for all $\tilde{p} \in H^1(\Omega) \cap L_0^2(\Omega)$.

Note that $\Delta u_{\mathcal{F}} + f_1 \in [L^2(\Omega)]^d$ and therefore the right-hand-side is a functional in $[H_0^1(\Omega)]^*$. So, existence and uniqueness of $p_1 \in H_0^1(\Omega) \cap L_0^2(\Omega)$ is guaranteed. Using the choice $\tilde{p} := p_1$, we obtain

$$\|\nabla p_1\|_{L^2(\Omega)} \leq \|\Delta u_{\mathcal{F}}\|_{L^2(\Omega)} + \|f_1\|_{L^2(\Omega)} \leq \|u_{\mathcal{F}}\|_{H^2(\Omega)} + \|f_1\|_{L^2(\Omega)}.$$

Using Poincaré's inequality, we obtain further

$$(4.5) \quad \|p_1\|_{H^1(\Omega)} \lesssim \|u_{\mathcal{F}}\|_{H^2(\Omega)} + \|f_1\|_{L^2(\Omega)}.$$

Let $p_2 \in H^1(\Omega) \cap L_0^2(\Omega)$ be such that

$$-(\nabla p_2, \nabla \tilde{p})_{L^2(\Omega)} = -\beta(u_{\mathcal{F}}, \nabla \tilde{p})_{L^2(\Omega)} + (f_2, \nabla \tilde{p})_{L^2(\Omega)}$$

for all $\tilde{p} \in H^1(\Omega) \cap L_0^2(\Omega)$. This implies, as $u_{\mathcal{F}} \in [H_0^1(\Omega)]^d$ and $f_2 \in [H_0^1(\Omega)]^d$, that

$$(4.6) \quad -(\nabla p_2, \nabla \tilde{p})_{L^2(\Omega)} = \beta(\nabla \cdot u_{\mathcal{F}}, \tilde{p})_{L^2(\Omega)} - (\nabla \cdot f_2, \tilde{p})_{L^2(\Omega)}$$

holds. As $\beta \nabla \cdot u_{\mathcal{F}} - \nabla \cdot f_2 \in L^2(\Omega)$, existence and uniqueness of p_2 is guaranteed. Condition **(R1)** implies moreover $p_2 \in H^2(\Omega)$ and

$$(4.7) \quad \|p_2\|_{H^2(\Omega)}^2 \lesssim \beta^2 \|\nabla \cdot u_{\mathcal{F}}\|_{L^2(\Omega)}^2 + \|\nabla f_2\|_{L^2(\Omega)}^2 \lesssim \beta^2 \|u_{\mathcal{F}}\|_{H^1(\Omega)}^2 + \|f_2\|_{H^1(\Omega)}^2.$$

Note that from (4.3), (4.4) and (4.6), we obtain

$$(\nabla(p_1 + p_2), \nabla \tilde{p})_{L^2(\Omega)} = (\nabla p_{\mathcal{F}}, \nabla \tilde{p})_{L^2(\Omega)}$$

is satisfied for all $\tilde{p} \in H^1(\Omega) \cap L_0^2(\Omega)$, which implies (because $p_{\mathcal{F}} \in L_0^2(\Omega)$ and $p_1 + p_2 \in L_0^2(\Omega)$) that $p_{\mathcal{F}} = p_1 + p_2$ is satisfied.

So, we have using (4.5) and (4.7)

$$\begin{aligned}
\|p_{\mathcal{F}}\|_{P_{+,k}}^2 &= h_k^2 \|p_{\mathcal{F}}\|_{H^1(\Omega) + \beta^{-1}H^2(\Omega)}^2 \\
&\leq \inf_{\substack{p=q_1+q_2, \\ q_1 \in H^1(\Omega) \cap L_0^2(\Omega) \\ q_2 \in H^2(\Omega) \cap L_0^2(\Omega)}} h_k^2 \|q_1\|_{H^1(\Omega)}^2 + h_k^2 \|q_2\|_{\beta^{-1}H^2(\Omega)}^2 \\
&\leq h_k^2 \|p_1\|_{H^1(\Omega)}^2 + h_k^2 \|p_2\|_{\beta^{-1}H^2(\Omega)}^2 \\
&\lesssim h_k^2 \|u_{\mathcal{F}}\|_{H^2(\Omega)}^2 + h_k^2 \|u_{\mathcal{F}}\|_{\beta H^1(\Omega)}^2 + h_k^2 \|f_1\|_{L^2(\Omega)}^2 + h_k^2 \|f_2\|_{\beta^{-1}H^1(\Omega)}^2 \\
&= h_k^2 \|u_{\mathcal{F}}\|_{H^2(\Omega) \cap \beta H^1(\Omega)}^2 + h_k^2 \|f_1\|_{L^2(\Omega)}^2 + h_k^2 \|f_2\|_{\beta^{-1}H^1(\Omega)}^2.
\end{aligned}$$

As $f_2 \in [H_0^1(\Omega)]^d$ was chosen arbitrarily, we can take the infimum over all f_2 , which finishes the proof. \square

For the next lemma, we need some notation. As $H_0^1(\Omega)$ is dense in $L^2(\Omega)$, for $u \in [H^2(\Omega)]^d$ the function $-\Delta u \in [L^2(\Omega)]^d$ can be approximated by some function $w^\epsilon \in [H_0^1(\Omega)]^d$ such that

$$\|(-\Delta)u - w^\epsilon\|_{L^2(\Omega)}^2 \leq \epsilon.$$

So, we can introduce an operator $-\Delta^\epsilon : [H^2(\Omega)]^d \rightarrow [H_0^1(\Omega)]^d$ such that

$$\|(-\Delta)u - (-\Delta^\epsilon)u\|_{L^2(\Omega)}^2 \leq \epsilon$$

for all $u \in H^2(\Omega)$.

Analogously, we introduce the operator $\nabla^\epsilon : H^1(\Omega) \rightarrow [H_0^1(\Omega)]^d$ such that

$$\|\nabla p - \nabla^\epsilon p\|_{L^2(\Omega)}^2 \leq \epsilon$$

for all $p \in H^1(\Omega)$.

Lemma 4.5. *Let $\mathcal{F} := (f, g) \in (X_-)^*$ and let $x_{\mathcal{F}} = (u_{\mathcal{F}}, p_{\mathcal{F}})$ be the solution of (4.1). Then $u_{\mathcal{F}}$ satisfies the estimate*

$$\|u_{\mathcal{F}}\|_{U_{+,k}}^2 \lesssim \|\mathcal{F}\|_{(X_{-,k})^*} \|x_{\mathcal{F}}\|_{X_{+,k}}.$$

Proof. Let $\mathcal{F}(\tilde{y}, \tilde{p}) := (f, \tilde{u})_{L^2(\Omega)} + (g, \tilde{p})_{L^2(\Omega)}$ for $f \in [L^2(\Omega)]^d$ and $g \in H_0^1(\Omega)$.

The idea of this proof is to show that for all $\epsilon > 0$ there is some $\tilde{x}^\epsilon \in X$ such that

$$\begin{aligned}
(4.8) \quad &\mathcal{F}(\tilde{x}^\epsilon) - \mathcal{B}(x_{\mathcal{F}}, \tilde{x}^\epsilon) \\
&\lesssim h_k^{-2} \|\mathcal{F}\|_{[X_{-,k}]^*} \|x_{\mathcal{F}}\|_{X_{+,k}} - h_k^{-2} \|u_{\mathcal{F}}\|_{U_{+,k}}^2 \\
&\quad + \epsilon(1 + \beta) h_k^{-1} (\|x_{\mathcal{F}}\|_{X_{+,k}} + \|\mathcal{F}\|_{(X_{-,k})^*}) + \epsilon^2.
\end{aligned}$$

Note that the left-hand-side of the inequality is 0. Therefore, this would be sufficient to show the statement of the lemma, as $\epsilon > 0$ can be chosen arbitrarily small.

In the following, we show that (4.8) is satisfied for the choice $\tilde{x}^\epsilon := (-\Delta^\epsilon u_{\mathcal{F}}, \nabla^\epsilon p_{\mathcal{F}})$. We estimate the individual summands of $\mathcal{F}(\tilde{x}^\epsilon) - \mathcal{B}(x_{\mathcal{F}}, \tilde{x}^\epsilon)$ separately.

$$\begin{aligned}
-\beta(u_{\mathcal{F}}, -\Delta^\epsilon u_{\mathcal{F}})_{L^2(\Omega)} &\leq -\beta(u_{\mathcal{F}}, -\Delta u_{\mathcal{F}})_{L^2(\Omega)} + \epsilon\beta \|u_{\mathcal{F}}\|_{L^2(\Omega)} \\
&= -\beta(\nabla u_{\mathcal{F}}, \nabla u_{\mathcal{F}})_{L^2(\Omega)} + \epsilon\beta \|u_{\mathcal{F}}\|_{L^2(\Omega)} \lesssim -\beta \|u_{\mathcal{F}}\|_{H^1(\Omega)}^2 + \epsilon\beta h_k^{-1} \|x_{\mathcal{F}}\|_{X_{+,k}}
\end{aligned}$$

is satisfied due to the fact that $u_{\mathcal{F}} \in H^2(\Omega) \cap H_0^1(\Omega)$ and due to Friedrichs' inequality.

For the next summand, we obtain

$$\begin{aligned} & -(\nabla u_{\mathcal{F}}, \nabla(-\Delta^\epsilon)u_{\mathcal{F}})_{L^2(\Omega)} = -(\Delta u_{\mathcal{F}}, \Delta^\epsilon u_{\mathcal{F}})_{L^2(\Omega)} \\ & \leq -(\Delta u_{\mathcal{F}}, \Delta u_{\mathcal{F}})_{L^2(\Omega)} + \epsilon \|\Delta u_{\mathcal{F}}\|_{L^2(\Omega)}^2 \leq -\|u_{\mathcal{F}}\|_{H^2(\Omega)}^2 + \epsilon \beta^{1/2} h_k^{-1} \|x_{\mathcal{F}}\|_{X_{+,k}} \end{aligned}$$

due to the fact that Δ^ϵ maps into $H_0^1(\Omega)$. Moreover we use $u_{\mathcal{F}} \in H^2(\Omega) \cap H_0^1(\Omega)$ and Friedrichs' inequality.

For the next two summands, we obtain

$$\begin{aligned} & -(\nabla \cdot u_{\mathcal{F}}, \nabla \cdot \nabla^\epsilon p_{\mathcal{F}})_{L^2(\Omega)} - (\nabla \cdot (-\Delta^\epsilon)u_{\mathcal{F}}, p_{\mathcal{F}})_{L^2(\Omega)} \\ & = (\nabla \nabla \cdot u_{\mathcal{F}}, \nabla^\epsilon p_{\mathcal{F}})_{L^2(\Omega)} - (\Delta^\epsilon u_{\mathcal{F}}, \nabla p_{\mathcal{F}})_{L^2(\Omega)} \\ & \leq (\nabla \nabla \cdot u_{\mathcal{F}}, \nabla^\epsilon p_{\mathcal{F}})_{L^2(\Omega)} - (\Delta^\epsilon u_{\mathcal{F}}, \nabla^\epsilon p_{\mathcal{F}})_{L^2(\Omega)} + \epsilon \|\Delta^\epsilon u_{\mathcal{F}}\|_{L^2(\Omega)} \\ & \leq -(\nabla u_{\mathcal{F}}, \nabla \nabla^\epsilon p_{\mathcal{F}})_{L^2} - (\Delta u_{\mathcal{F}}, \nabla^\epsilon p_{\mathcal{F}})_{L^2} + \epsilon(\|\Delta u_{\mathcal{F}}\|_{L^2} + \|\nabla^\epsilon p_{\mathcal{F}}\|_{L^2} + \epsilon) \\ & = -(\nabla u_{\mathcal{F}}, \nabla \nabla^\epsilon p_{\mathcal{F}})_{L^2} - (\nabla \cdot \nabla u_{\mathcal{F}}, \nabla^\epsilon p_{\mathcal{F}})_{L^2} + \epsilon(\|\Delta u_{\mathcal{F}}\|_{L^2} + \|\nabla^\epsilon p_{\mathcal{F}}\|_{L^2} + \epsilon) \\ & = -(\nabla u_{\mathcal{F}}, \nabla \nabla^\epsilon p_{\mathcal{F}})_{L^2} + (\nabla u_{\mathcal{F}}, \nabla \nabla^\epsilon p_{\mathcal{F}})_{L^2} + \epsilon(\|\Delta u_{\mathcal{F}}\|_{L^2} + \|\nabla^\epsilon p_{\mathcal{F}}\|_{L^2} + \epsilon) \\ & \leq \epsilon(\|\Delta u_{\mathcal{F}}\|_{L^2(\Omega)} + \|\nabla p_{\mathcal{F}}\|_{L^2(\Omega)} + \epsilon) \leq \epsilon(\|u_{\mathcal{F}}\|_{H^2(\Omega)} + \|p_{\mathcal{F}}\|_{H^1(\Omega)} + \epsilon) \\ & \leq \epsilon(1 + \beta^{1/2})h_k^{-1} \|x_{\mathcal{F}}\|_{X_{+,k}} + \epsilon^2. \end{aligned}$$

Let $f_2 \in [H_0^1(\Omega)]^d$ and $f_1 := f - f_2$. Then

$$(f_1, -\Delta^\epsilon u_{\mathcal{F}})_{L^2(\Omega)} \lesssim \|f_1\|_{L^2(\Omega)} \|u_{\mathcal{F}}\|_{H^2(\Omega)} + \epsilon \|f_1\|_{L^2(\Omega)}$$

holds as well as

$$\begin{aligned} (f_2, -\Delta^\epsilon u_{\mathcal{F}})_{L^2(\Omega)} & \lesssim (\nabla f_2, \nabla u_{\mathcal{F}})_{L^2(\Omega)} + \epsilon \|f_2\|_{L^2(\Omega)} \\ & \lesssim \|f_2\|_{\beta^{-1}H^1(\Omega)} \|u_{\mathcal{F}}\|_{\beta H^1(\Omega)} + \epsilon \beta^{1/2} \|f_2\|_{\beta^{-1}H^1(\Omega)}. \end{aligned}$$

This implies

$$\begin{aligned} (f, -\Delta^\epsilon u_{\mathcal{F}})_{L^2(\Omega)} & \lesssim \|f\|_{L^2(\Omega) + \beta^{-1}H_0^1(\Omega)} \|u_{\mathcal{F}}\|_{H^2(\Omega) \cap \beta H^1(\Omega)} + \epsilon(1 + \beta^{1/2}) \|f\|_{L^2(\Omega) + \beta^{-1}H_0^1(\Omega)} \\ & \lesssim \|f\|_{L^2(\Omega) + \beta^{-1}H_0^1(\Omega)} \|u_{\mathcal{F}}\|_{H^2(\Omega) \cap \beta H^1(\Omega)} + \epsilon h_k^{-1} (1 + \beta^{1/2}) \|\mathcal{F}\|_{[X_{-,k}]^*}. \end{aligned}$$

Let $p_2 \in H^2(\Omega)$ and $p_1 := p - p_2$. We have

$$(g, \nabla \cdot \nabla^\epsilon p_1)_{L^2(\Omega)} = -(\nabla g, \nabla^\epsilon p_1)_{L^2(\Omega)} \lesssim \|g\|_{H^1(\Omega)} \|p_1\|_{H^1(\Omega)} + \epsilon \|g\|_{H^1(\Omega)}.$$

Moreover, using $g \in H_0^1(\Omega)$, we have also

$$\begin{aligned} (g, \nabla \cdot \nabla^\epsilon p_2)_{L^2(\Omega)} & = -(\nabla g, \nabla^\epsilon p_2)_{L^2(\Omega)} \lesssim -(\nabla g, \nabla p_2)_{L^2(\Omega)} + \epsilon \|g\|_{H^1(\Omega)} \\ & = (g, \nabla \cdot \nabla p_2)_{L^2(\Omega)} + \epsilon \|g\|_{H^1(\Omega)} \leq \|g\|_{\beta L^2(\Omega)} \|p_1\|_{\beta^{-1}H^2(\Omega)} + \epsilon \|g\|_{H^1(\Omega)} \end{aligned}$$

and therefore

$$\begin{aligned} (g, \nabla \cdot \nabla^\epsilon p_{\mathcal{F}})_{L^2(\Omega)} & \lesssim \|g\|_{H_0^1(\Omega) + \beta L_0^2(\Omega)} \|p_{\mathcal{F}}\|_{H^1(\Omega) + \beta^{-1}H^2(\Omega)} + \epsilon \|g\|_{H^1(\Omega)} \\ & \lesssim \|g\|_{H_0^1(\Omega) + \beta L_0^2(\Omega)} \|p_{\mathcal{F}}\|_{H^1(\Omega) + \beta^{-1}H^2(\Omega)} + \epsilon h_k^{-1} \|\mathcal{F}\|_{(X_{-,k})^*} \end{aligned}$$

is satisfied. Combining these results, we immediately obtain (4.8), which finishes the proof. \square

So, we can show the following theorem.

Theorem 4.6. *Suppose that the assumptions **(R)** and **(R1)** are satisfied. Then condition **(A4)** is satisfied.*

Proof. Note that Lemma 4.3 already states that $x_{\mathcal{F}} \in X_{+,k}$. It remains to show

$$(4.9) \quad \|x_{\mathcal{F}}\|_{X_+} \lesssim \|\mathcal{F}\|_{(X_{-,k})^*}.$$

Note that Lemma 4.4 implies immediately

$$\|x_{\mathcal{F}}\|_{X_{+,k}}^2 = \|u_{\mathcal{F}}\|_{U_{+,k}}^2 + \|p_{\mathcal{F}}\|_{P_{+,k}}^2 \lesssim \|u_{\mathcal{F}}\|_{U_{+,k}}^2 + \|\mathcal{F}\|_{(X_{-,k})^*}^2$$

If we combine this result with the statement of Lemma 4.5, we obtain

$$\|x_{\mathcal{F}}\|_{X_{+,k}}^2 \leq C \left(\|x_{\mathcal{F}}\|_{X_{+,k}} \|\mathcal{F}\|_{(X_{-,k})^*} + \|\mathcal{F}\|_{(X_{-,k})^*}^2 \right)$$

for some constant $C > 0$ (independent of k and β) which implies

$$\|x_{\mathcal{F}}\|_{X_{+,k}} \leq \frac{1}{2} \left(C + \sqrt{4C + C^2} \right) \|\mathcal{F}\|_{(X_{-,k})^*},$$

and further (4.9). This finishes the proof. \square

So, we have shown condition **(A4)**. Therefore, Theorem 4.1 implies the approximation property. Note, that we have now shown the approximation property in the norm $\|\cdot\|_{X_{-,k}}$, i.e., (4.2). The next step is to show the approximation property in the norm $\|\cdot\|_k$, i.e., (3.6).

To show (3.6), the following lemma is sufficient.

Lemma 4.7. *The inequality*

$$\| \|x_k\| \|_k \lesssim \|x_k\|_{X_{-,k}}$$

is satisfied for all $x_k \in X_k$.

For showing this lemma, we need some preliminary results.

We define on each grid level the functions $\psi_k : \Omega \rightarrow \mathbb{R}$ as follows. The function ψ_k is a continuous function that is linear on each element (cf. the Courant element). The function ψ_k is defined to be 0 on all vertices which are located on $\partial\Omega$ and to be 1 on all other vertices (interior vertices).

In what follows, \hat{P}_k is the space of functions obtained by discretizing the space $H^1(\Omega)$ using the standard Courant element. Note that – as we have used the Taylor Hood element – the identity $\hat{P}_k = \{p_k + a : p_k \in P_k \text{ and } a \in \mathbb{R}\}$ holds.

Lemma 4.8. *The inequality*

$$\|\psi_k p_k\|_{L^2(\Omega)}^2 \leq \|p_k\|_{L^2(\Omega)}^2$$

holds for all $p_k \in \hat{P}_k$.

Proof. It suffices to note that $\psi_k \leq 1$ holds on the whole domain Ω . \square

Lemma 4.9. *The inequality*

$$\|\psi_k p_k\|_{H^1(\Omega)}^2 \lesssim h_k^{-2} \|p_k\|_{L^2(\Omega)}^2$$

holds for all $p_k \in \hat{P}_k$.

Proof. First note that

$$\|\psi_k p_k\|_{H^1(\Omega)}^2 = \|\nabla(\psi_k p_k)\|_{L^2(\Omega)}^2 + \|\psi_k p_k\|_{L^2(\Omega)}^2 \lesssim \|\nabla(\psi_k p_k)\|_{L^2(\Omega)}^2 + h_k^{-2} \|p_k\|_{L^2(\Omega)}^2$$

is satisfied due to the last lemma and $h_k \lesssim 1$. So, it suffices to bound $\|\nabla(\psi_k p_k)\|_{L^2(\Omega)}^2$ from above.

Assume that the mesh consists of the elements \mathcal{T}_i for $i = 1, \dots, M_k$. First note that

$$\|p_k\|_{L^2(\Omega)}^2 = \sum_{i=1}^{M_k} \|p_k\|_{L^2(\mathcal{T}_i)}^2 \quad \text{and} \quad \|\nabla(\psi_k p_k)\|_{L^2(\Omega)}^2 = \sum_{i=1}^{M_k} \|\nabla(\psi_k p_k)\|_{L^2(\mathcal{T}_i)}^2$$

is satisfied.

So it suffices to show

$$(4.10) \quad \|\nabla(\psi_k p_k)\|_{L^2(\mathcal{T}_i)}^2 \lesssim h_k^{-2} \|p_k\|_{L^2(\mathcal{T}_i)}^2$$

for all $i = 1, \dots, M_k$.

Note that on all interior elements (no vertex is located on $\partial\Omega$) we have $\psi_k = 1$. So, the inequality (4.10) is a standard inverse inequality. So, it remains to show (4.10) on all elements where one or two vertices are located on $\partial\Omega$. (Note that we have assumed already in Section 2 that each element has at least one vertex in the interior of Ω .)

We have

$$(4.11) \quad \begin{aligned} \|\nabla(\psi_k(\xi)p_k(\xi))\|_{L^2(\mathcal{T}_i)}^2 &= \int_{\mathcal{T}_i} \|\nabla(\psi_k(\xi)p_k(\xi))\|_{\ell^2}^2 d\xi \\ &= \int_{\mathcal{T}_i} \|(\nabla\psi_k(\xi))p_k(\xi) + \psi_k(\xi)(\nabla p_k(\xi))\|_{\ell^2}^2 d\xi \\ &\lesssim \int_{\mathcal{T}_i} \|\nabla\psi_k(\xi)\|_{\ell^2}^2 p_k(\xi)^2 d\xi + \int_{\mathcal{T}_i} \psi_k(\xi)^2 \|\nabla p_k(\xi)\|_{\ell^2}^2 d\xi. \end{aligned}$$

For estimating the first summand, $\int_{\mathcal{T}_i} \|\nabla\psi_k(\xi)\|_{\ell^2}^2 p_k(\xi)^2 d\xi$, we use the fact that ψ_k is linear and therefore $\nabla\psi_k$ is constant. So, we obtain

$$\int_{\mathcal{T}_i} \|\nabla\psi_k(\xi)\|_{\ell^2}^2 p_k(\xi)^2 d\xi \leq \|\nabla\psi_k\|_{\ell^2}^2 \|p_k\|_{L^2(\Omega)}^2.$$

Note that that ψ_k takes the value 0 on two vertices of the element, say P_1 and P_2 , and the value 1 on the third vertex P_3 (the case that ψ_k takes the value 1 on two vertices, say P_1 and P_2 and the value 0 on the third vertex is completely analogous). It is geometrically evident that $\|\nabla\psi_k\|_{\ell^2}$ is equal to the reciprocal of the length of the altitude on the edge P_1P_2 . The reciprocal of the length of the altitude is bounded from above by h_k^{-1} . This shows

$$\int_{\mathcal{T}_i} \|\nabla\psi_k(\xi)\|_{\ell^2}^2 p_k(\xi)^2 d\xi \leq h_k^{-2} \|p_k\|_{L^2(\Omega)}^2.$$

The second summand in (4.11), $\int_{\mathcal{T}_i} \psi_k(\xi)^2 \|\nabla p_k(\xi)\|_{\ell^2}^2 d\xi$, can be bounded from above using $\psi_k(\xi)^2 \leq 1$ by $\|\nabla p_k\|_{L^2(\Omega)}^2$, which can be bounded from above by $h_k^{-2} \|p_k\|_{L^2(\Omega)}^2$ using a standard inverse inequality. This finishes the proof. \square

Lemma 4.10. *The inequality*

$$(4.12) \quad (p_k, \psi_k p_k)_{L^2(\Omega)} \gtrsim (p_k, p_k)_{L^2(\Omega)}$$

holds for all $p_k \in \hat{P}_k$.

Proof. Assume that the mesh consists of the elements \mathcal{T}_i for $i = 1, \dots, M_k$. First note that

$$(p_k, \psi_k p_k)_{L^2(\Omega)} = \sum_{i=1}^{M_k} (p_k, \psi_k p_k)_{L^2(\mathcal{T}_i)} \quad \text{and} \quad (p_k, p_k)_{L^2(\Omega)} = \sum_{i=1}^{M_k} (p_k, p_k)_{L^2(\mathcal{T}_i)}.$$

So, it suffices to show (4.12) for the individual elements \mathcal{T}_i , i.e.,

$$(4.13) \quad (p_k, p_k)_{L^2(\mathcal{T}_i)} \lesssim (p_k, \psi_k p_k)_{L^2(\mathcal{T}_i)}$$

for all $p_k \in \hat{P}_k$.

Note that on all interior elements (no vertex is located on $\partial\Omega$) we have $\psi_k = 1$. So, the inequality (4.13) is obviously satisfied. So, it remains to show (4.10) on all elements where one (case 1) or two vertices (case 2) are located on $\partial\Omega$. (Note that we have assumed already in Section 2 that each element has at least one vertex in the interior of Ω).

Case 1. The statement which we have to show reads as follows:

$$(4.14) \quad \int_{\mathcal{T}_i} p_k^2(\xi) \, d\xi \lesssim \int_{\mathcal{T}_i} \psi_k(\xi) p_k^2(\xi) \, d\xi.$$

Note that, both, $p_k(\xi)$ and $\psi_k(\xi)$, are linear functions. These integrals can be computed using the reference element $\Delta := \{(\xi_1, \xi_2) \in (0, 1)^2 : \xi_1 + \xi_2 < 1\}$. It is well known that there is a linear transformation $\Phi_i : \Delta \rightarrow \mathcal{T}_i$. Using the standard substitution rule, we obtain that (4.14) is equivalent to

$$\int_{\Delta} \hat{p}_k^2(\hat{\xi}) |\det \nabla^2 \Phi_i(\hat{\xi})| \, d\hat{\xi} \lesssim \int_{\Delta} \hat{\psi}_k(\hat{\xi}) \hat{p}_k^2(\hat{\xi}) |\det \nabla^2 \Phi_i(\hat{\xi})| \, d\hat{\xi},$$

where $\hat{p}(\xi) = u(\Phi_i^{-1}(\xi))$ and $\hat{\psi}(\xi) = \psi(\Phi_i^{-1}(\xi))$ and $|\det \nabla^2 \Phi_i(\hat{\xi})|$ is the absolute value of the Jacobi determinant of the transformation. Because the transformation is linear, both, \hat{p} and $\hat{\psi}$, are linear functions and, moreover, the Jacobi determinant is a constant. So it suffices to show

$$(4.15) \quad \int_{\Delta} \hat{p}_k^2(\hat{\xi}) \, d\hat{\xi} \lesssim \int_{\Delta} \hat{\psi}_k(\hat{\xi}) \hat{p}_k^2(\hat{\xi}) \, d\hat{\xi},$$

for all linear functions \hat{p}_k . As mentioned above, the function $\hat{\psi}_k$ takes the value 1 on two vertices and the value 0 on one vertex. We assume without loss of generality, that it takes the value 0 on $(0, 1)$. This directly implies that $\hat{\psi}_k(\xi) = 1 - \xi_2$.

Using a general approach for \hat{p}_k , like $\hat{p}_k(\xi_1, \xi_2) := a_0 + a_1 \xi_1 + a_2 \xi_2$, we can compute both integrals in (4.15) and obtain for $\mathbf{a} := (a_0, a_1, a_2)^T$ that

$$\int_{\Delta} \hat{p}_k^2(\hat{\xi}) \, d\hat{\xi} = \mathbf{a}^T F \mathbf{a} \quad \text{with} \quad F := \frac{1}{24} \begin{pmatrix} 12 & 4 & 4 \\ 4 & 2 & 1 \\ 4 & 1 & 2 \end{pmatrix}$$

and

$$\int_{\Delta} \hat{\psi}_k(\hat{\xi}) \hat{p}_k^2(\hat{\xi}) \, d\hat{\xi} = \mathbf{a}^T G \mathbf{a} \quad \text{with} \quad G := \frac{1}{120} \begin{pmatrix} 40 & 15 & 10 \\ 15 & 8 & 3 \\ 10 & 3 & 4 \end{pmatrix}.$$

Obviously, we obtain

$$\frac{\int_{\Delta} \hat{p}_k^2(\hat{\xi}) \, d\hat{\xi}}{\int_{\Delta} \hat{\psi}_k(\hat{\xi}) \hat{p}_k^2(\hat{\xi}) \, d\hat{\xi}} \leq \lambda_{\max}(G^{-1}F) = \frac{1}{3} (6 + \sqrt{6}),$$

where λ_{max} is the largest eigenvalue. This finishes the proof for Case 1.

Case 2. The analysis of Case 2 is analogous to the analysis of Case 1. Here, we assume that $\hat{\psi}_k(\xi)$ takes the value 0 on the vertices $(0, 0)$ and $(1, 0)$ and the value 1 on the vertex $(0, 1)$. So, we obtain $\hat{\psi}_k(\xi) = \xi_2$. Here we obtain, using the same arguments as above, that the inequality

$$\int_{\mathcal{T}_i} p_k^2(\xi) d\xi \leq (4 + \sqrt{6}) \int_{\mathcal{T}_i} \psi_k(\xi) p_k^2(\xi) d\xi,$$

holds, which finishes the proof. \square

Now we can show Lemma 4.7.

Proof. (of Lemma 4.7). The analysis can be done component-wise. So it suffices to show that

$$\|u_k\|_{U,k} \lesssim \|u_k\|_{U_{-,k}} \quad \text{and} \quad \|p_k\|_{P,k} \lesssim \|p_k\|_{P_{-,k}}$$

is satisfied for all $u_k \in U_k$ and all $p_k \in P_k$.

We obtain using the definition of $\|\cdot\|_{U_{-,k}}$, the choice $\tilde{u} := u_k \in [H_0^1(\Omega)]^d$ and a standard inverse inequality

$$\begin{aligned} \|u_k\|_{U_{-,k}}^2 &= \sup_{0 \neq \tilde{u} \in [L^2(\Omega)]^d} \frac{(u_k, \tilde{u})_{L^2(\Omega)}^2}{h_k^2 \|\tilde{u}\|_{L^2(\Omega) + \beta^{-1} H_0^1(\Omega)}^2} \gtrsim \frac{(u_k, u_k)_{L^2(\Omega)}^2}{h_k^2 \|u_k\|_{L^2(\Omega) + \beta^{-1} H_0^1(\Omega)}^2} \\ &= \frac{(u_k, u_k)_{L^2(\Omega)}^2}{\inf_{w \in [H_0^1(\Omega)]^d} h_k^2 \|w - u_k\|_{L^2(\Omega)}^2 + h_k^2 \|w\|_{\beta^{-1} H^1(\Omega)}^2} \\ &\gtrsim \frac{(u_k, u_k)_{L^2(\Omega)}^2}{\inf_{w_k \in U_k} h_k^2 \|w_k - u_k\|_{L^2(\Omega)}^2 + h_k^2 \|w_k\|_{\beta^{-1} H^1(\Omega)}^2} \\ &\gtrsim \frac{(u_k, u_k)_{L^2(\Omega)}^2}{\inf_{w_k \in U_k} h_k^2 \|w_k - u_k\|_{L^2(\Omega)}^2 + \|w_k\|_{\beta^{-1} L^2(\Omega)}^2} \\ &= (\beta + h_k^{-2}) \|u_k\|_{L^2(\Omega)}^2 = \|u_k\|_{U,k}^2, \end{aligned}$$

i.e., the first inequality.

Note that we obtain using the definition of $\|\cdot\|_{P_{-,k}}$ and the fact that $p_k \in P_k \subseteq L_0^2(\Omega)$ that

$$\|p_k\|_{P_{-,k}}^2 = \sup_{0 \neq q \in H_0^1(\Omega) \cap L_0^2(\Omega)} \frac{(p_k, q)_{L^2(\Omega)}^2}{h_k^2 \|q\|_{H_0^1(\Omega) \cap \beta L^2(\Omega)}^2} = \sup_{0 \neq q \in H_0^1(\Omega) \cap L_0^2(\Omega)} \frac{(p_k - a, q)_{L^2(\Omega)}^2}{h_k^2 \|q\|_{H_0^1(\Omega) \cap \beta L^2(\Omega)}^2}$$

is satisfied, where $a \in \mathbb{R}$ is such that $\psi_k * (p_k - a) \in L_0^2(\Omega)$.

By plugging in $q := \psi_k * (p_k - a)$, we obtain using Lemmas 4.8, 4.9 and 4.10 (note that $p_k \in P_k$ implies $p_k - a \in \hat{P}_k$) that

$$\begin{aligned} \|p_k\|_{P_{-,k}}^2 &\gtrsim \frac{(p_k - a, p_k - a)_{L^2(\Omega)}^2}{\|p_k - a\|_{L^2(\Omega)}^2 + \beta h_k^2 \|p_k - a\|_{L^2(\Omega)}^2} \\ &= (1 + \beta h_k^2)^{-1} \|p_k - a\|_{L^2(\Omega)}^2 \geq (1 + \beta h_k^2)^{-1} \|p_k\|_{L^2(\Omega)}^2 = \|p_k\|_{P,k}^2, \end{aligned}$$

where the last inequality is satisfied due to the fact that $p_k \in P_k \subseteq L_0^2(\Omega)$. So, this shows the second inequality and finishes the proof. \square

So, we have shown the approximation property. So, we obtain as follows.

Theorem 4.11. *Assume that*

- *the regularity assumptions **(R)** and **(R1)** are satisfied on the domain Ω ,*
- *the problem is discretized using the Taylor-Hood element and*
- *one of the smoothers proposed in this paper is used.*

Then the two-grid method converges if sufficiently many smoothing steps are applied, i.e., we have

$$\|x_k^{(1)} - x_k^*\|_k \leq q(\nu) \|x_k^{(0)} - x_k^*\|_k,$$

with $q(\nu) := C_S C_A \nu^{-1/2}$, where the constants C_A and C_S are independent of the grid level k and the choice of the parameter β .

The convergence of the W-cycle multigrid method follows under weak assumptions, cf. [9].

5. NUMERICAL RESULTS

In this section, we illustrate the convergence theory presented within this paper with numerical results. The numerical experiments were done as follows.

For the numerical experiments, the domain Ω was chosen to be the unit square $\Omega := (0, 1)^2$. As mentioned in Section 2, the weak inf-sup-condition (2.4) can be shown for the Taylor-Hood element only if at least one vertex of each element is in the interior of the domain Ω . As this is not satisfied for the the standard decomposition of the unit square into two triangular elements, we choose the coarsest grid level $k = 0$ to be a decomposition of the domain Ω into 8 triangles, as seen in Figure 1. The grid levels $k = 1, 2, \dots$ were constructed by uniform refinement, i.e., every triangle was decomposed into four subtriangles.

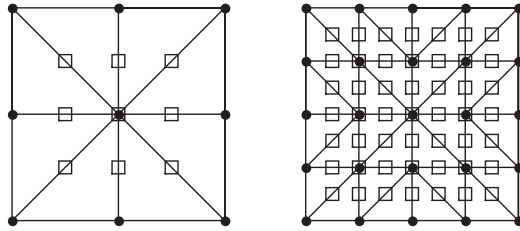


FIGURE 1. Discretization on grid levels $k = 1$ and $k = 2$, where the squares denote the degrees of freedom of (the components of) the velocity field u are the the dots denote the degrees of freedom of pressure distribution p

The right-hand-side functions f and g have been chosen such that the solution of the problem on grid level k is the L^2 -projection of the exact solution

$$u(\xi_1, \xi_2) := \phi(\xi_1, \xi_2) \begin{pmatrix} \xi_2 - \frac{1}{2} \\ \frac{1}{2} - \xi_1 \end{pmatrix} \quad \text{and} \quad p(\xi_1, \xi_2) := \phi(\xi_1, \xi_2),$$

where

$$\phi(\xi_1, \xi_2) := \max \left\{ 0, \min \left\{ 1, 2 - 4\sqrt{(\xi_1 - \frac{1}{2})^2 + (\xi_2 - \frac{1}{2})^2} \right\} \right\},$$

into X_k .

For solving the discretized problem, we have used the proposed W-cycle multigrid method. To obtain a proper scaling, the following choice of matrix \mathcal{L}_k was taken for the numerical experiments

$$(5.1) \quad \mathcal{L}_k := \begin{pmatrix} \hat{A}_k & \\ & \hat{S}_k \end{pmatrix}, \quad \text{where} \quad \hat{A}_k := \text{diag } A_k \quad \text{and} \quad \hat{S}_k := \text{diag } B_k \hat{A}_k^{-1} B_k^T.$$

Note that the matrix \mathcal{L}_k , introduced above, is spectrally equivalent to the matrix \mathcal{L}_k , introduced in Section 3. So, the choice introduced above is still covered by the convergence theory. The damping parameter is chosen to be $\tau = 0.35$, for all grid levels k and all choices of β .

For the Uzawa type smoother, the matrices \hat{A}_k and \hat{S}_k have been chosen as introduced in (5.1) and the damping parameters have been chosen to be $\tau := \sigma := 0.8$. Note that the matrices \hat{A}_k and \hat{S}_k , introduced in this section, are spectrally equivalent to the choices introduced in Section 3. So, the choice introduced above is still covered by the convergence theory.

The number of iterations and the convergence rate were measured as follows: we start with $\underline{x}_k^{(0)}$ and measure the reduction of the error in each step using the norm $\|\cdot\|_k$. The iteration was stopped when the initial error was reduced by a factor of $\epsilon = 10^{-9}$. The convergence rate q is the mean convergence rate in this iteration, i.e.,

$$q = \left(\frac{\|\underline{x}_k^{(n)} - \underline{x}_k^*\|_k}{\|\underline{x}_k^{(0)} - \underline{x}_k^*\|_k} \right)^{1/n},$$

where n is the number of iterations needed to reach the stopping criterion. Here, \underline{x}_k^* is the exact solution and $\underline{x}_k^{(i)}$ is the i -th iterate.

$\nu = 1 + 1$		$\nu = 2 + 2$		$\nu = 3 + 3$		$\nu = 4 + 4$		$\nu = 8 + 8$		$\nu = 16 + 16$	
n	q	n	q	n	q	n	q	n	q	n	q
Normal equation smoother											
88	0.789	46	0.634	30	0.496	24	0.412	17	0.293	11	0.148
Uzawa type smoother											
divergent		divergent		14	0.212	20	0.347	6	0.031	5	0.008

TABLE 1. Number of iterations n and convergence rate q for the normal equation smoother and the Uzawa type smoother depending with $\nu = \nu_{pre} + \nu_{post}$ smoothing steps on grid level $k = 4$ for $\beta = 1$

In Table 1 we compare for a fixed grid level (level $k = 4$) and a fixed choice $\beta = 1$ the convergence rates for several choices of ν , the number of pre- and post-smoothing steps. We see that the convergence rate behaves approximately like $\nu^{-1/2}$, if the number of smoothing steps is increased. This is consistent with the theory which guarantees the convergence rate being bounded by $C \nu^{-1/2}$. We observe that the preconditioned normal equation smoother already converges for $\nu = 1 + 1$ smoothing steps, while for the Uzawa type smoother $\nu = 3 + 3$ smoothing steps are necessary.

	$\beta = 1$		$\beta = 10^2$		$\beta = 10^4$		$\beta = 10^6$		$\beta = 10^8$		$\beta = 10^{10}$	
	n	q	n	q	n	q	n	q	n	q	n	q
$k = 4$	30	0.496	30	0.493	30	0.496	68	0.736	71	0.745	71	0.745
$k = 5$	29	0.489	29	0.488	22	0.388	64	0.722	70	0.744	71	0.745
$k = 6$	29	0.484	29	0.486	27	0.457	52	0.670	70	0.743	71	0.745
$k = 7$	28	0.475	28	0.475	28	0.470	35	0.553	70	0.742	71	0.745
$k = 8$	28	0.469	28	0.469	28	0.468	20	0.347	67	0.732	71	0.746

TABLE 2. Number of iterations n and convergence rate q for the normal equation smoother with $\nu = 3 + 3$ smoothing steps

	$\beta = 1$		$\beta = 10^2$		$\beta = 10^4$		$\beta = 10^6$		$\beta = 10^8$		$\beta = 10^{10}$	
	n	q	n	q	n	q	n	q	n	q	n	q
$k = 4$	14	0.212	13	0.201	6	0.030	7	0.051	8	0.059	8	0.059
$k = 5$	13	0.194	13	0.193	9	0.095	7	0.043	7	0.050	7	0.050
$k = 6$	12	0.176	12	0.176	11	0.145	6	0.024	7	0.047	7	0.048
$k = 7$	12	0.166	12	0.166	11	0.150	5	0.013	7	0.044	7	0.045
$k = 8$	11	0.147	11	0.147	11	0.145	8	0.058	7	0.038	7	0.043

TABLE 3. Number of iterations n and convergence rate q for the Uzawa type smoother with $\nu = 3 + 3$ smoothing steps

	$\beta = 1$		$\beta = 10^2$		$\beta = 10^4$		$\beta = 10^6$		$\beta = 10^8$		$\beta = 10^{10}$	
	n	q	n	q	n	q	n	q	n	q	n	q
$k = 4$	88	0.789	87	0.787	101	0.814	267	0.925	268	0.925	268	0.925
$k = 5$	87	0.787	87	0.787	65	0.727	263	0.924	272	0.927	272	0.927
$k = 6$	85	0.783	85	0.783	79	0.768	223	0.911	274	0.927	274	0.927
$k = 7$	84	0.780	84	0.780	82	0.776	130	0.854	274	0.927	275	0.927
$k = 8$	82	0.776	82	0.776	81	0.774	58	0.697	270	0.926	276	0.928

TABLE 4. Number of iterations n and convergence rate q for the normal equation smoother with $\nu = 1 + 1$ smoothing steps

In Tables 2, 3 and 4 we compare various grid levels k and choices of the parameter β . For Tables 2 and 3, we have used a fixed choice of $\nu = 3 + 3$ smoothing steps. First we observe that, for both smoothers, the number of iterations seems to be well-bounded for all grid levels k which yields an optimal convergence behavior. Moreover, we see that the number of iterations is also well-bounded for a wide range of choices of the parameter β , i.e., we observe also robust convergence as predicted by the convergence theory.

Comparing both kinds of smoothers, we see that the Uzawa type smoother leads to much faster convergence rates than the preconditioned normal equation smoother. Note that moreover the computational complexity of the Uzawa type smoother (per iteration) is slightly smaller than the complexity of the normal equation smoother.

In Table 4, we observe that also for the case that only $\nu = 1 + 1$ smoothing steps are applied, the preconditioned normal equation smoother converges for all choices of β and all grid levels. Also in this case the convergence rates are – as proposed by the theory – uniformly bounded.

It has to be mentioned that for the model problem, also the (more efficient) V-cycle multigrid method converges with rates comparable to the convergence rates of the W-cycle multigrid method. However, the V-cycle is not covered by the convergence theory.

The numerical experiments done by the author have shown that the convergence rates can be improved slightly by adjusting the choice of the parameters to the grid levels and the choice of β . However, the main goal of this paper is to show that the proposed method also works well for fixed choices of the parameter.

6. CONCLUSIONS AND FURTHER WORK

In the present paper we have proposed an all-at-once multigrid solver for the generalized Stokes problem where the smoothing property is needed in the scaled L^2 -norm $||| \cdot |||_k$. This allows to construct a multigrid method where the smoother is a simple linear iteration (which consists only of divisions and the multiplication of vectors with the system matrix \mathcal{A}_k). In the present paper, a preconditioned normal equation smoother and an Uzawa type smoother have been chosen but it seems possible to find also other simple smoothers which satisfy the smoothing property in the norm $||| \cdot |||_k$.

The convergence rates observed for the multigrid method proposed in the present paper are comparable with the rates observed for the methods proposed in [15]. Note that for applying the methods proposed in the named paper, it is necessary to solve a Poisson problem in each smoothing step. This is not needed for the method proposed in the present paper.

So, the extension of the analysis presented in this paper to other smoothers would be of interest.

Acknowledgements. The author thanks Markus Kollmann for providing code used to compute the numerical results presented in this paper. Moreover, the support of the numerical analysis group of the Mathematical Institute, University of Oxford, is gratefully acknowledged.

REFERENCES

1. R. Adams and J. Fournier, *Sobolev Spaces*, Academic Press, 2008, 2nd ed.
2. M. Bercovier and O. Pironneau, *Error estimates for finite element method solution of the Stokes problem in primitive variables*, Numerische Mathematik **33** (1979), 211 – 224.
3. J. Bergh and J. Löfström, *Interpolation Spaces, an Introduction*, Springer, Berlin, 1976.
4. D. Braess and R. Sarazin, *An efficient smoother for the Stokes problem*, Applied Numerical Mathematics **23** (1997), no. 1, 3 – 19.
5. J.H. Bramble and J.E. Pasciak, *Iterative techniques for time dependent stokes problems*, Comp. Math. Appl. **33** (1997), 13–30.
6. S. Brenner, *A nonconforming multigrid method for the stationary stokes problem*, Math. Comp. **55** (1990), 411–473.
7. S.C. Brenner, *Multigrid methods for parameter dependent problems*, RAIRO, Modélisation Math. Anal. Numér **30** (1996), 265 – 297.
8. M. Dauge, *Elliptic boundary value problems on corner domains. Smoothness and asymptotics of solutions*, Lecture Notes in Mathematics, 1341. Berlin etc.: Springer-Verlag (1988).
9. W. Hackbusch, *Multi-Grid Methods and Applications*, Springer, Berlin, 1985.

10. R.B. Kellogg and J.E Osborn, *A regularity result for the Stokes problem in a convex polygon*, Journal of Functional Analysis **21** (1976), no. 4, 397–431.
11. G. M. Kobelkov and M. A. Olshanskii, *Effective preconditioning of uzawa type schemes for generalized stokes problem*, Numer. Math. **86** (2000), 443–470.
12. M. Larin and A. Reusken, *A comparative study of efficient iterative solvers for generalized stokes problem*, Numer. Linear Algebra Appl. **15** (2008), 13–34.
13. K. Mardal and R. Winther, *Uniform preconditioners for the time dependent stokes problem*, Numerische Mathematik **98** (2004), 305–327 (English).
14. ———, *Uniform preconditioners for the time dependent stokes problem (erratum)*, Numerische Mathematik **103** (2006), 171–172 (English).
15. M. Olshanskii, *Multigrid Analysis for the Time Dependent Stokes Problem*, Mathematics of Computation **81** (2012), no. 277, 57 – 79.
16. M. Olshanskii, J. Peters, and A. Reusken, *Uniform preconditioners for a parameter dependent saddle point problem with application to generalized stokes interface equations*, Numerische Mathematik (2005), 159–191.
17. ———, *Uniform preconditioners for a parameter dependent saddle point problem with application to generalized Stokes interface equations*, Numerische Mathematik **105** (2006), 159 – 191.
18. Verfürth R., *A multilevel algorithm for mixed problems*, SIAM J. on Numerical Analysis **21** (1984), 264–271.
19. J. Schöberl and W. Zulehner, *On Schwarz-type Smoothers for Saddle Point Problems*, Numerische Mathematik **95** (2002), 3777 – 399.
20. S. Takacs and W. Zulehner, *Convergence analysis of all-at-once multigrid methods for elliptic control problems under partial elliptic regularity*, SIAM J. on Numerical Analysis (2012), Accepted for publication.
21. S. P. Vanka, *Block-implicit multigrid solution of Navier-Stokes equations in primitive variables*, Math. Comp. **65** (1986), 138 – 158.
22. R. Verfürth, *Error estimates for a mixed finite element approximation of the Stokes equations*, RAIRO **18** (1984), 175 – 182.
23. P. Wesseling and C. W. Oosterlee, *Geometric multigrid with applications to computational fluid dynamics*, J. Comput. Appl. Math. **128** (2001), 311–334.
24. W. Zulehner, *A class of smoothers for saddle point problems*, Computing. **65** (2000), 227–246.
25. ———, *Non-standard Norms and Robust Estimates for Saddle Point Problems*, SIAM J. on Matrix Anal. & Appl **32** (2011), 536 – 560.

VISITING POSTDOC, MATHEMATICAL INSTITUTE, UNIVERSITY OF OXFORD, UNITED KINGDOM
E-mail address: takacs@maths.ox.ac.uk, stefan.takacs@numa.uni-linz.ac.at