

# ADAPTIVE FINITE ELEMENT METHODS FOR THE STOKES PROBLEM WITH DISCONTINUOUS VISCOSITY

ANDREA BONITO <sup>\*</sup> AND DENIS DEVAUD <sup>†</sup>

**Abstract.** Discontinuity in viscosities is of interest in many applications. Classical adaptive numerical methods perform under the restricting assumption that the discontinuities of the viscosity are captured by the initial partition. This excludes applications where the jump of the viscosity takes place across curves, manifolds or at a priori unknown positions. We present a novel estimate measuring the distortion of the viscosity in  $L^q$  for a  $q < \infty$ , thereby allowing for any type of discontinuities. This estimate requires the velocity  $\mathbf{u}$  of the Stokes system to satisfy the extra regularity assumption  $\nabla \mathbf{u} \in L^r(\Omega)^{d \times d}$  for some  $r > 2$ . We show that the latter holds on any bounded Lipschitz domain provided the data belongs to a smaller class than those required to obtain well-posedness. Based on this theory, we introduce adaptive finite element methods which approximate the solution of Stokes equations with possible discontinuous viscosities. We prove that these algorithms are quasi-optimal in terms of error compared to the number of cells. Finally, the performance of the adaptive algorithm is numerically illustrated on insightful examples.

**1. Introduction.** Let  $\Omega \subset \mathbb{R}^d$  be a polyhedral domain with  $d \geq 2$  and  $\mu : \Omega \rightarrow \mathbb{R}$ . We consider the following form of the Stokes problem

$$-\operatorname{div}(\mu \nabla \mathbf{u}) + \nabla p = \mathbf{f}, \quad \text{in } \Omega, \quad (1.1a)$$

$$\operatorname{div}(\mathbf{u}) = g, \quad \text{in } \Omega, \quad (1.1b)$$

$$\mathbf{u} = \mathbf{0}, \quad \text{on } \partial\Omega. \quad (1.1c)$$

For  $1 \leq r \leq \infty$ , we denote by  $L^r(\Omega)$  the usual Lebesgue spaces. If  $\mathbf{w} : \Omega \rightarrow \mathbb{R}^d$  is a vector valued function, we define its  $L^r(\Omega)^d$  norm as

$$\|\mathbf{w}\|_{L^r(\Omega)} := \|\ |\mathbf{w}|\ \|_{L^r(\Omega)}, \quad (1.2)$$

where  $|\cdot|$  denotes the Euclidean norm in  $\mathbb{R}^d$ . Throughout this paper, we will use boldface characters to denote vector valued functions. Similarly, the  $L^r(\Omega)^{d \times d}$  norm of a matrix valued function  $A : \Omega \rightarrow \mathbb{R}^{d \times d}$  is denoted by

$$\|A\|_{L^r(\Omega)} := \|\ |A|\ \|_{L^r(\Omega)},$$

where  $|B|$  is the Frobenius norm of  $B \in \mathbb{R}^{d \times d}$ . The functional space  $L_0^r(\Omega)$  is the set of functions in  $L^r(\Omega)$  with zero mean value. We define  $W_0^{1,r}(\Omega)$ ,  $r \geq 1$ , as the Sobolev space of real valued functions in  $W^{1,r}(\Omega)$  whose traces vanish on the boundary of  $\Omega$ . These are Banach spaces with the norm

$$\|v\|_{W_0^{1,r}} := \|\nabla v\|_{L^r(\Omega)}.$$

The dual space of  $W_0^{1,r}(\Omega)$  is denoted by  $W^{-1,r'}(\Omega)$  with  $\frac{1}{r} + \frac{1}{r'} = 1$  and is equipped with the norm

$$\|f\|_{W^{-1,r'}} := \sup_{v \in W_0^{1,r}(\Omega), v \neq 0} \frac{\langle f, v \rangle}{\|v\|_{W_0^{1,r}}},$$

where  $\langle \cdot, \cdot \rangle$  is the  $W^{-1,r'} - W_0^{1,r}$  duality product. For  $r = 2$ , we set  $H^{-1}(\Omega) := W^{-1,2}(\Omega)$ . The same definitions hold for any subdomain  $\omega \subset \Omega$ . To simplify the notations, we will write  $\|v\|_{L^r(\Omega)} = \|v\|_{L^r}$  if no confusion is possible.

We consider the weak formulation of (1.1): given  $\mathbf{f} \in H^{-1}(\Omega)^d$  and  $g \in L_0^2(\Omega)$ , find  $(\mathbf{u}, p) \in H_0^1(\Omega)^d \times L_0^2(\Omega)$  such that

$$\int_{\Omega} \mu \nabla \mathbf{u} \cdot \nabla \mathbf{v} - \int_{\Omega} p \operatorname{div}(\mathbf{v}) + \int_{\Omega} q \operatorname{div}(\mathbf{u}) = \langle \mathbf{f}, \mathbf{v} \rangle + \int_{\Omega} qg, \quad \forall (\mathbf{v}, q) \in H_0^1(\Omega)^d \times L_0^2(\Omega). \quad (1.3)$$

---

<sup>\*</sup>Texas A&M University, Department of Mathematics, TAMU 3368, College Station, TX 77843, USA (bonito@math.tamu.edu)

<sup>†</sup>EPFL SMA, CH-1015, Lausanne, Switzerland (denis.devaud@epfl.ch).

We assume that there exists  $\mu_{\min}$  and  $\mu_{\max}$  such that

$$0 < \mu_{\min} \leq \mu(\mathbf{x}) \leq \mu_{\max} < \infty, \quad \text{a.e. in } \Omega. \quad (1.4)$$

This guarantees [16] that (1.3) admits a unique weak solution  $(\mathbf{u}, p) \in H_0^1(\Omega)^d \times L_0^2(\Omega)$ . Moreover, there exists a constant  $C$  only depending on  $\Omega$ ,  $\mu_{\min}$  and  $\mu_{\max}$  such that

$$\|\mathbf{u}\|_{H_0^1} + \|p\|_{L^2} \leq C (\|\mathbf{f}\|_{H^{-1}} + \|g\|_{L^2}). \quad (1.5)$$

In practice, the first step of any algorithm to find the solution of (1.3) is to approximate  $\mathbf{f}$  by  $\hat{\mathbf{f}}$ ,  $g$  by  $\hat{g}$  and  $\mu$  by  $\hat{\mu}$ . The simplest example of this is the application of quadrature rules. We then need to estimate the influence of such a replacement on our solution. In this aim, we derive in Section 2.1 (Lemma 2.1) the following bound<sup>1</sup>

$$\|\mathbf{u} - \hat{\mathbf{u}}\|_{H_0^1} + \|p - \hat{p}\|_{L^2} \lesssim \|\mathbf{f} - \hat{\mathbf{f}}\|_{H^{-1}} + \|g - \hat{g}\|_{L^2} + (\|g\|_{L^2} + \|\mathbf{f}\|_{H^{-1}}) \|\hat{\mu} - \mu\|_{L^\infty}, \quad (1.6)$$

where  $(\hat{\mathbf{u}}, \hat{p})$  is the solution of (1.3) with coefficients  $\hat{\mathbf{f}}$ ,  $\hat{g}$  and  $\hat{\mu}$ . In the case of a discontinuous viscosity function  $\mu$ , (1.6) reveals that its approximation  $\hat{\mu}$  is required to match exactly these discontinuities. If these discontinuities occur on curves or manifolds or if the discontinuity interface is a priori unknown, this condition is generally not fulfilled.

To circumvent this issue, we extend the results for elliptic partial differential equations (PDEs) by Bonito, DeVore and Nochetto [8] and present a second perturbation theory (Theorem 2.2), which relaxes the constraints on the viscosity approximation suggested by (1.6):

$$\|\mathbf{u} - \hat{\mathbf{u}}\|_{H_0^1} + \|p - \hat{p}\|_{L^2} \lesssim \|\mathbf{f} - \hat{\mathbf{f}}\|_{H^{-1}} + \|g - \hat{g}\|_{L^2} + \|\hat{\mu} - \mu\|_{L^q} \|\nabla \mathbf{u}\|_{L^r}, \quad q := \frac{2r}{r-2}, \quad (1.7)$$

provided  $\nabla \mathbf{u} \in L^r(\Omega)^{d \times d}$  for some  $r \geq 2$ . The distortion of the viscosity function is now measured in  $L^q$  for  $q < \infty$ , thereby not requiring the approximation  $\hat{\mu}$  to exactly match the discontinuities of  $\mu$ .

For the estimate (1.7) to be relevant, we need that  $\nabla \mathbf{u} \in L^r(\Omega)^{d \times d}$  for some  $r > 2$ . In the case  $\mu \equiv 1$ , Maz'ya and Rossmann [20] proved the validity of such estimate for any  $r \in [2, R)$  where  $R > 2$  only depends on  $\Omega$  and provided  $\mathbf{f} \in W^{-1,r}(\Omega)^d$ ,  $g \in L_0^r(\Omega)$ . Using a perturbation argument originally due to Jochmann [18], see also [9], we extend in Section 2.2 this result to any viscosity satisfying (1.4) upon restricting further the range of parameter, i.e.  $r \in [2, \tilde{R})$ ,  $2 < \tilde{R} \leq R$ . We refer to Meyers [21], see also [8], for similar estimates on elliptic PDEs.

Only few results are available in the literature regarding adaptive finite element methods (AFEMs) suited to approximate the solution of the Stokes system. Optimal rate of convergence of an adaptive method was first achieved by Dahlike, Dahmen and Urban [13] using a wavelet discretization. The adaptive method relies on an Uzawa algorithm to handle the saddle point structure of the Stokes system and to take advantage of the ellipticity of the pressure Schur complement. In [2], Bänsch, Morin and Nochetto proposed a similar algorithm using piecewise polynomial finite elements pairs for the velocity-pressure discretization (inf-sup stable or not). Each iteration consists of an adaptive finite element module constructing an approximation of the velocity within a prescribed tolerance followed by a pressure update. The prescribed tolerance is reduced by a constant factor  $0 < \rho < 1$  between each iterations. The authors proved a contraction property for the energy error between two Uzawa iterates and observed numerically optimal convergence rates provided  $\rho$  is sufficiently close to 1. Then, Kondratyuk and Stevenson designed and studied in [19] a new algorithm which also relies on a Uzawa algorithm and is based on unstable piecewise polynomial finite element pairs. One of the key feature of this algorithm is its ability to automatically balance the errors on the velocity and the pressure, thereby leading to optimal rate of convergence without any parameter restriction. Without relying on a Uzawa algorithm, Hu and Xu [17] derived optimal rate of convergence using the Scott-Vogelius finite elements and non-conforming piecewise linear finite elements. Both finite element methods allow to separate

<sup>1</sup>Throughout this paper we use the notation  $A \lesssim B$  to denote  $A \leq CB$  with a constant  $C$  independent of  $A$  and  $B$ . We shall indicate if appropriate on which quantities the constant  $C$  depends.

the convergence of the velocity and the pressure. We mention again that all these methods assume that the viscosity, if not constant, is piecewise smooth on the initial subdivision.

In view of estimate (1.7), we then describe in Section 3 an AFEM to approximate the solution of (1.3) when  $\mu$  has discontinuities not necessarily matching the initial mesh. This algorithm is based on four subroutines, `RHS_MOM`, `RHS_MASS`, `VISC` and `STOKES`. The first two produce piecewise polynomial approximations  $\hat{\mathbf{f}}$  and  $\hat{g}$  of  $\mathbf{f}$  and  $g$  respectively. The routine `VISC` returns a piecewise polynomial approximation of the viscosity  $\hat{\mu}$  (close in  $L^q$ ) which satisfies

$$0 < C^{-1}\mu_{\min} \leq \hat{\mu} \leq C\mu_{\max}, \quad (1.8)$$

where  $C$  is an absolute constant. Note that requirement (1.8) can be ensured by post-processing as shown in [8]. The last routine `STOKES` is an AFEM producing an approximation of the solution to (1.3) where  $\mathbf{f}$ ,  $g$  and  $\mu$  are replaced by the approximations  $\hat{\mathbf{f}}$ ,  $\hat{g}$  and  $\hat{\mu}$  whose possible discontinuities are now aligned with the subdivision.

In Section 3, we introduce the main algorithm and present its structure. Its performances are discussed in Section 4, in particular we prove (see Theorem 4.3) its optimal rate of convergence on a certain class of functions. To achieve this performance, we assume that each subroutine `RHS_MOM`, `RHS_MASS`, `VISC` and `STOKES` also deliver optimal rates of convergence. Section 5 discusses realizations of such algorithm. We conclude in Section 6 by presenting numerical illustrations of the performance of our algorithm.

**2. Perturbation Theory.** We present in this section a perturbation theorem (Theorem 2.2) which enables to approximate the viscosity in a  $L^q$  norm for  $q \leq \infty$ . To allow for finite  $q$ , it requires that  $\nabla \mathbf{u} \in L^r(\Omega)^{d \times d}$  for some  $r > 2$ . We then show in Theorem 2.3 the existence of such  $r$  provided that the data  $\mathbf{f}$  and  $g$  fulfill additional regularity conditions.

**2.1. Perturbation Theorems.** We start by recalling that the divergence operator is surjective as an operator from  $H_0^1(\Omega)$  to  $L_0^2(\Omega)$  [16] and therefore the following inf-sup condition holds: There exists  $\beta > 0$  such that

$$\inf_{q \in L_0^2(\Omega)} \sup_{\mathbf{v} \in H_0^1(\Omega), \mathbf{v} \neq 0} \frac{\int_{\Omega} q \operatorname{div}(\mathbf{v})}{\|q\|_{L^2} \|\mathbf{v}\|_{H_0^1}} \geq \beta. \quad (2.1)$$

Let  $\mathbf{f}, \hat{\mathbf{f}} \in H^{-1}(\Omega)$ ,  $g, \hat{g} \in L_0^2(\Omega)$  and  $\mu, \hat{\mu}$  satisfying

$$0 < \mu_{\min} \leq \mu(\mathbf{x}) \leq \mu_{\max} < \infty, \quad 0 < \hat{\mu}_{\min} \leq \hat{\mu}(\mathbf{x}) \leq \hat{\mu}_{\max} < \infty, \quad \text{a.e. in } \Omega. \quad (2.2)$$

In what follows,  $(\mathbf{u}, p) \in H_0^1(\Omega)^d \times L_0^2(\Omega)$  denotes the weak solution of the Stokes system satisfying

$$\int_{\Omega} \mu \nabla \mathbf{u} \cdot \nabla \mathbf{v} - \int_{\Omega} p \operatorname{div}(\mathbf{v}) + \int_{\Omega} q \operatorname{div}(\mathbf{u}) = \langle \mathbf{f}, \mathbf{v} \rangle + \int_{\Omega} g q, \quad \forall (\mathbf{v}, q) \in H_0^1(\Omega)^d \times L_0^2(\Omega), \quad (2.3)$$

while  $(\hat{\mathbf{u}}, \hat{p}) \in H_0^1(\Omega)^d \times L_0^2(\Omega)$  denotes the solution with data  $\hat{\mathbf{f}}, \hat{g}$  and  $\hat{\mu}$ :

$$\int_{\Omega} \hat{\mu} \nabla \hat{\mathbf{u}} \cdot \nabla \mathbf{v} - \int_{\Omega} \hat{p} \operatorname{div}(\mathbf{v}) + \int_{\Omega} q \operatorname{div}(\hat{\mathbf{u}}) = \langle \hat{\mathbf{f}}, \mathbf{v} \rangle + \int_{\Omega} \hat{g} q, \quad \forall (\mathbf{v}, q) \in H_0^1(\Omega)^d \times L_0^2(\Omega). \quad (2.4)$$

We start with a perturbation result based on a distortion of the viscosity measured in  $L^\infty$ . This result is instrumental to derive the  $L^q$  perturbation estimate (2.10) provided in Theorem 2.2.

**LEMMA 2.1.** *The following estimate holds*

$$\|\mathbf{u} - \hat{\mathbf{u}}\|_{H_0^1} + \|p - \hat{p}\|_{L^2} \lesssim \|\mathbf{f} - \hat{\mathbf{f}}\|_{H^{-1}} + \|g - \hat{g}\|_{L^2} + (\|g\|_{L^2} + \|\mathbf{f}\|_{H^{-1}}) \|\hat{\mu} - \mu\|_{L^\infty}, \quad (2.5)$$

where the constant hidden in “ $\lesssim$ ” depends on  $\mu_{\min}$ ,  $\mu_{\max}$ ,  $\hat{\mu}_{\min}$ ,  $\hat{\mu}_{\max}$  and  $\beta$ .

*Proof.* We start by defining a lifting of  $g - \hat{g}$ . Let  $(\boldsymbol{\omega}, \pi) \in H_0^1(\Omega)^d \times L_0^2(\Omega)$  be the unique weak solution of the following Stokes problem

$$\int_{\Omega} \nabla \boldsymbol{\omega} \cdot \nabla \mathbf{v} - \int_{\Omega} \pi \operatorname{div}(\mathbf{v}) + \int_{\Omega} q \operatorname{div}(\boldsymbol{\omega}) = \int_{\Omega} (g - \hat{g})q, \quad \forall (\mathbf{v}, q) \in H_0^1(\Omega)^d \times L_0^2(\Omega).$$

Setting  $\tilde{\mathbf{u}} := \mathbf{u} - \boldsymbol{\omega} \in H_0^1(\Omega)^d$ , we realize that relations (2.3) and (2.4) yield for any  $\mathbf{v} \in H_0^1(\Omega)^d$

$$\int_{\Omega} \hat{\mu} \nabla(\tilde{\mathbf{u}} - \hat{\mathbf{u}}) \cdot \nabla \mathbf{v} - \int_{\Omega} (p - \hat{p}) \operatorname{div}(\mathbf{v}) = \langle \mathbf{f} - \hat{\mathbf{f}}, \mathbf{v} \rangle - \int_{\Omega} \hat{\mu} \nabla \boldsymbol{\omega} \cdot \nabla \mathbf{v} + \int_{\Omega} (\hat{\mu} - \mu) \nabla \mathbf{u} \cdot \nabla \mathbf{v} \quad (2.6)$$

and for any  $q \in L_0^2(\Omega)$

$$\int_{\Omega} q \operatorname{div}(\tilde{\mathbf{u}} - \hat{\mathbf{u}}) = 0. \quad (2.7)$$

Hence, choosing  $\mathbf{v} := \tilde{\mathbf{u}} - \hat{\mathbf{u}}$  in (2.6) and  $q := p - \hat{p}$  in (2.7) together with a Cauchy-Schwarz inequality, we deduce that

$$\begin{aligned} \int_{\Omega} \hat{\mu} \nabla(\tilde{\mathbf{u}} - \hat{\mathbf{u}}) \cdot \nabla(\tilde{\mathbf{u}} - \hat{\mathbf{u}}) &\leq \left( \|\mathbf{f} - \hat{\mathbf{f}}\|_{H^{-1}} + \|\hat{\mu} - \mu\|_{L^\infty} \|\mathbf{u}\|_{H_0^1} + \hat{\mu}_{\max} \|\boldsymbol{\omega}\|_{H_0^1} \right) \|\tilde{\mathbf{u}} - \hat{\mathbf{u}}\|_{H_0^1} \\ &\lesssim \left( \|\mathbf{f} - \hat{\mathbf{f}}\|_{H^{-1}} + \|\hat{\mu} - \mu\|_{L^\infty} \|\mathbf{u}\|_{H_0^1} + \|g - \hat{g}\|_{L^2} \right) \|\tilde{\mathbf{u}} - \hat{\mathbf{u}}\|_{H_0^1}, \end{aligned}$$

where we used in the last step the estimate  $\|\boldsymbol{\omega}\|_{H_0^1} + \|\pi\|_{L^2} \lesssim \|g - \hat{g}\|_{L^2}$ ; see (1.5). Therefore, the definition (2.2) of  $\hat{\mu}_{\min}$  directly implies

$$\|\mathbf{u} - \hat{\mathbf{u}}\|_{H_0^1} \leq \|\tilde{\mathbf{u}} - \hat{\mathbf{u}}\|_{H_0^1} + \|\boldsymbol{\omega}\|_{L^2} \lesssim \|\mathbf{f} - \hat{\mathbf{f}}\|_{H^{-1}} + \|\hat{\mu} - \mu\|_{L^\infty} \|\mathbf{u}\|_{H_0^1} + \|g - \hat{g}\|_{L^2}. \quad (2.8)$$

We now turn to the pressure contribution. Owing to the inf-sup condition (2.1) and since  $p - \hat{p} \in L_0^2(\Omega)$ , there exists  $\mathbf{v} \in H_0^1(\Omega)^d$  such that  $\|\mathbf{v}\|_{H_0^1} = 1$  and

$$\beta \|p - \hat{p}\|_{L^2} \leq \int_{\Omega} (p - \hat{p}) \operatorname{div}(\mathbf{v}).$$

As a consequence, relations (2.3) and (2.4) satisfied by  $(\mathbf{u}, p)$  and  $(\hat{\mathbf{u}}, \hat{p})$  together with the definition (2.2) of  $\hat{\mu}_{\max}$  imply

$$\begin{aligned} \beta \|p - \hat{p}\|_{L^2} &\leq \int_{\Omega} (\mu \nabla \mathbf{u} - \hat{\mu} \nabla \hat{\mathbf{u}}) \cdot \nabla \mathbf{v} - \langle \mathbf{f} - \hat{\mathbf{f}}, \mathbf{v} \rangle \\ &= \int_{\Omega} (\mu - \hat{\mu}) \nabla \mathbf{u} \cdot \nabla \mathbf{v} + \int_{\Omega} \hat{\mu} \nabla(\mathbf{u} - \hat{\mathbf{u}}) \cdot \nabla \mathbf{v} - \langle \mathbf{f} - \hat{\mathbf{f}}, \mathbf{v} \rangle \\ &\leq \|\mu - \hat{\mu}\|_{L^\infty} \|\mathbf{u}\|_{H_0^1} + \hat{\mu}_{\max} \|\mathbf{u} - \hat{\mathbf{u}}\|_{H_0^1} + \|\mathbf{f} - \hat{\mathbf{f}}\|_{H^{-1}}. \end{aligned} \quad (2.9)$$

The desired bound (2.5) follows upon combining (2.8), (2.9) and the bound (1.5).  $\square$

We now provide the perturbation theorem extending the above lemma and used to derive the AFEM for the Stokes system with general discontinuous viscosities discussed in Section 3.2. This result extends to the Stokes system the argumentation provided in [8] for elliptic PDEs.

**THEOREM 2.2.** (Perturbation Theorem) *Assume that  $\nabla \mathbf{u} \in L^r(\Omega)^{d \times d}$  for some  $r \geq 2$ . There exists a constant  $D$  only depending on  $\mu_{\min}$ ,  $\mu_{\max}$ ,  $\hat{\mu}_{\min}$ ,  $\hat{\mu}_{\max}$  and  $\beta$  such that*

$$\|\mathbf{u} - \hat{\mathbf{u}}\|_{H_0^1} + \|p - \hat{p}\|_{L^2} \leq D \left( \|\mathbf{f} - \hat{\mathbf{f}}\|_{H^{-1}} + \|g - \hat{g}\|_{L^2} + \|\hat{\mu} - \mu\|_{L^q} \|\nabla \mathbf{u}\|_{L^r} \right), \quad q := \frac{2r}{r-2}. \quad (2.10)$$

*Proof.* Let us assume that  $(\bar{\mathbf{u}}, \bar{p}) \in H_0^1(\Omega)^d \times L_0^2(\Omega)$  is given by

$$\int_{\Omega} \hat{\mu} \nabla \bar{\mathbf{u}} \cdot \nabla \mathbf{v} - \int_{\Omega} \bar{p} \operatorname{div}(\mathbf{v}) + \int_{\Omega} q \operatorname{div}(\bar{\mathbf{u}}) = \langle \mathbf{f}, \mathbf{v} \rangle + \int_{\Omega} g q, \quad \forall (\mathbf{v}, q) \in H_0^1(\Omega)^d \times L_0^2(\Omega). \quad (2.11)$$

Thanks to (1.5), see also Lemma 2.1, we have

$$\|\bar{\mathbf{u}} - \hat{\mathbf{u}}\|_{H_0^1} + \|\bar{p} - \hat{p}\|_{L^2} \lesssim \|\mathbf{f} - \hat{\mathbf{f}}\|_{H^{-1}} + \|g - \hat{g}\|_{L^2}. \quad (2.12)$$

We proceed as in Lemma 2.1 except that Hölder inequalities are invoked to take advantage of the regularity assumed on  $\nabla \mathbf{u}$ . The equations (2.3) and (2.11) satisfied by  $(\mathbf{u}, p)$  and  $(\bar{\mathbf{u}}, \bar{p})$  directly imply

$$\int_{\Omega} (\mu \nabla \mathbf{u} - \hat{\mu} \nabla \bar{\mathbf{u}}) \cdot \nabla (\mathbf{u} - \bar{\mathbf{u}}) = 0,$$

whence the definition (2.2) of  $\hat{\mu}_{\min}$  yields

$$\hat{\mu}_{\min} \|\mathbf{u} - \bar{\mathbf{u}}\|_{H_0^1}^2 \leq \int_{\Omega} \hat{\mu} \nabla (\mathbf{u} - \bar{\mathbf{u}}) \cdot \nabla (\mathbf{u} - \bar{\mathbf{u}}) = \int_{\Omega} (\hat{\mu} - \mu) \nabla \mathbf{u} \cdot \nabla (\mathbf{u} - \bar{\mathbf{u}}) \leq \int_{\Omega} |\hat{\mu} - \mu| |\nabla \mathbf{u}| |\nabla (\mathbf{u} - \bar{\mathbf{u}})|.$$

We apply an Hölder inequality to the last term of the above estimate with  $q$  and  $q' := \frac{q}{q-1} \in [1, 2]$  to get

$$\int_{\Omega} |\hat{\mu} - \mu| |\nabla \mathbf{u}| |\nabla (\mathbf{u} - \bar{\mathbf{u}})| \leq \|\hat{\mu} - \mu\|_{L^q} \|\nabla \mathbf{u}\|_{L^q} \|\nabla (\mathbf{u} - \bar{\mathbf{u}})\|_{L^{q'}}. \quad (2.13)$$

We set  $s := \frac{2}{q'} \in [1, 2]$  and apply a second Hölder inequality to write

$$\|\nabla \mathbf{u}\|_{L^q} \|\nabla (\mathbf{u} - \bar{\mathbf{u}})\|_{L^{q'}} \leq \|\nabla \mathbf{u}\|_{L^{s'q'}} \|\mathbf{u} - \bar{\mathbf{u}}\|_{H_0^1}. \quad (2.14)$$

Since the argument holds for any  $q \in [2, \infty]$ , we define  $q := \frac{2r}{r-2}$  so that  $r = s'q' \in [2, \infty]$  and it follows

$$\|\mathbf{u} - \bar{\mathbf{u}}\|_{H_0^1} \leq \frac{1}{\hat{\mu}_{\min}} \|\hat{\mu} - \mu\|_{L^q} \|\nabla \mathbf{u}\|_{L^r}. \quad (2.15)$$

We infer the desired bound on  $\|\mathbf{u} - \hat{\mathbf{u}}\|_{H_0^1}$  combining (2.12) and (2.15).

The estimate of  $\|p - \hat{p}\|_{L^2}$  follows from the inf-sup condition (2.1). Indeed, since  $p - \bar{p} \in L_0^2(\Omega)$ , there exists  $\mathbf{v} \in H_0^1(\Omega)^d$  with  $\|\mathbf{v}\|_{H_0^1} = 1$  such that

$$\beta \|p - \bar{p}\|_{L^2} \leq \int_{\Omega} (p - \bar{p}) \operatorname{div}(\mathbf{v}).$$

Hence, recalling that  $(\mathbf{u}, p)$  and  $(\bar{\mathbf{u}}, \bar{p})$  are the solutions of (2.3) and (2.11), we obtain the following bound for the pressure term

$$\begin{aligned} \beta \|p - \bar{p}\|_{L^2} &\leq \int_{\Omega} (\mu \nabla \mathbf{u} - \hat{\mu} \nabla \bar{\mathbf{u}}) \cdot \nabla \mathbf{v} = \int_{\Omega} (\mu - \hat{\mu}) \nabla \mathbf{u} \cdot \nabla \mathbf{v} + \int_{\Omega} \hat{\mu} \nabla (\mathbf{u} - \bar{\mathbf{u}}) \cdot \nabla \mathbf{v} \\ &\leq \int_{\Omega} (\mu - \hat{\mu}) \nabla \mathbf{u} \cdot \nabla \mathbf{v} + \hat{\mu}_{\max} \|\mathbf{u} - \bar{\mathbf{u}}\|_{H_0^1}, \end{aligned}$$

where we have used the definition (2.2) of  $\hat{\mu}_{\max}$  to derive the last estimate. Using the same Hölder inequalities leading to (2.13) and (2.14) on the first term of the above estimate, we get

$$\|p - \bar{p}\|_{L^2} \leq \frac{1}{\beta} \left( \hat{\mu}_{\max} \|\mathbf{u} - \bar{\mathbf{u}}\|_{H_0^1} + \|\hat{\mu} - \mu\|_{L^q} \|\nabla \mathbf{u}\|_{L^r} \right). \quad (2.16)$$

The desired bound for  $\|p - \hat{p}\|_{L^2}$  is obtained combining estimates (2.12) and (2.16).  $\square$

REMARK 1 (Explicit Constant). *Tracking the constants in all the estimations in the proof of Theorem 2.2, we obtain an explicit formula for the constant  $D$ :*

$$D = 2 \max \left\{ \frac{1}{\beta} + \frac{1}{\hat{\mu}_{\min}} \left( 1 + \frac{\hat{\mu}_{\max}}{\beta} \right), \frac{1}{\beta} \left( 1 + \frac{\hat{\mu}_{\max}}{\beta} \right) \left( 1 + \frac{\hat{\mu}_{\max}}{\hat{\mu}_{\min}} \right) \right\}.$$

REMARK 2. (Multiple subdomains) *The choice of the  $L^r$  norm in the bound (2.10) could be different from one subdomain of  $\Omega$  to another. Assume that  $\bar{\Omega} = \bar{\Omega}_1 \cup \bar{\Omega}_2$  for some disjoint open subsets  $\Omega_i \subset \Omega$ . A similar argument leads to the following bound instead of (2.10)*

$$\|\mathbf{u} - \hat{\mathbf{u}}\|_{H_0^1} + \|p - \hat{p}\|_{L^2} \lesssim \|\mathbf{f} - \hat{\mathbf{f}}\|_{H^{-1}} + \|g - \hat{g}\|_{L^2} + \|\hat{\mu} - \mu\|_{L^{q_1}(\Omega_1)} \|\nabla \mathbf{u}\|_{L^{r_1}(\Omega_1)} + \|\hat{\mu} - \mu\|_{L^{q_2}(\Omega_2)} \|\nabla \mathbf{u}\|_{L^{r_2}(\Omega_2)},$$

with  $q_i := \frac{2r_i}{r_i - 2}$  and  $r_i \geq 2$  is such that  $\nabla \mathbf{u} \in L^{r_i}(\Omega_i)^{d \times d}$ .

**2.2. Regularity Result for the Stokes Problem.** In order not to constraint the approximation  $\hat{\mu}$  to capture exactly the discontinuities of  $\mu$ , one needs  $q < \infty$ , i.e.  $r := \frac{2q}{q-2} > 2$ , in the perturbation estimate (2.10). More precisely, we need the additional regularity  $\nabla \mathbf{u} \in L^r(\Omega)^{d \times d}$ ,  $r > 2$ , on the velocity gradient. This is referenced as **Condition r**. When  $\mu \equiv 1$ , Maz'ya and Rossmann [20] showed that there exists  $R > 2$  depending on  $\Omega$  such that this condition is fulfilled for any  $2 \leq r \leq R$ , provided  $\mathbf{f} \in W^{-1,R}(\Omega)$  and  $g \in L_0^R(\Omega)$ . Note that when  $\Omega$  is not convex,  $R < 3$ . We denote by  $C_R$  the continuity constant such that

$$\|\nabla \mathbf{u}\|_{L^R} + \|p\|_{L^R} \leq C_R (\|\mathbf{f}\|_{W^{-1,R}} + \|g\|_{L^R}). \quad (2.17)$$

In the following theorem, we extend this result to general functions  $\mu$  satisfying (2.2) using a perturbation argument as originally proposed in [18]; see also [8, 9, 21].

THEOREM 2.3 ( $L^r$  estimate on the Stokes solution). *Let  $R > 2$  be such that (2.17) holds whenever  $\mathbf{f} \in W^{-1,R}(\Omega)^d$  and  $g \in L_0^R(\Omega)$  and assume that  $\mu$  satisfies (2.2). Then there exists  $2 < \tilde{R} \leq R$  such that the unique weak solution  $(\mathbf{u}, p) \in H_0^1(\Omega)^d \times L_0^2(\Omega)$  of (1.3) satisfies*

$$\|\nabla \mathbf{u}\|_{L^r} + \|p\|_{L^r} \leq C_r (\|\mathbf{f}\|_{W^{-1,r}} + \|g\|_{L^r}), \quad (2.18)$$

provided  $\mathbf{f} \in W^{-1,r}(\Omega)^d$  and  $g \in L_0^r(\Omega)$ ,  $2 \leq r < \tilde{R}$ , and where  $C_r$  is a constant only depending on  $\Omega$ ,  $r$ ,  $R$ ,  $\mu_{\min}$  and  $\mu_{\max}$ .

*Proof.* Let  $S : H_0^1(\Omega)^d \times L_0^2(\Omega) \rightarrow H^{-1}(\Omega)^d \times L_0^2(\Omega)$  be defined as

$$S(\mathbf{v}, q) := \begin{pmatrix} -\Delta \mathbf{v} + \nabla q \\ \operatorname{div}(\mathbf{v}) \end{pmatrix}, \quad \mathbf{v} \in H_0^1(\Omega)^d, \quad q \in L_0^2(\Omega).$$

The mapping  $S$  is an isomorphism from  $H_0^1(\Omega)^d \times L_0^2(\Omega)$  to  $H^{-1}(\Omega)^d \times L_0^2(\Omega)$ ; see for instance [16]. Moreover, we have that

$$\|S(\mathbf{v}, q)\|_{H^{-1} \times L^2} := \|-\Delta \mathbf{v} + \nabla q\|_{H^{-1}} + \|\operatorname{div}(\mathbf{v})\|_{L^2} \leq 2 \left( \|\mathbf{v}\|_{H_0^1} + \|q\|_{L^2} \right),$$

which implies that  $S$  is a bounded operator and so is  $S^{-1}$  as an operator from  $H^{-1}(\Omega)^d \times L_0^2(\Omega)$  to  $H_0^1(\Omega)^d \times L_0^2(\Omega)$  with norm denoted by  $C_2$  for later reference. In view of estimate (2.17),  $S^{-1}$  is also bounded as an operator from  $W^{-1,R}(\Omega)^d \times L_0^R(\Omega)$  to  $W_0^{1,R}(\Omega)^d \times L_0^R(\Omega)$  with norm  $C_R$ . As a consequence, the real interpolation theory [14, 24] implies that  $S^{-1}$  is a bounded operator from

$$\left( H_0^1(\Omega)^d \times L_0^2(\Omega), W_0^{1,R}(\Omega)^d \times L_0^R(\Omega) \right)_{\eta(r),r} = W_0^{1,r}(\Omega)^d \times L_0^r(\Omega)$$

to

$$\left( H^{-1}(\Omega)^d \times L_0^2(\Omega), W^{-1,R}(\Omega)^d \times L_0^R(\Omega) \right)_{\eta(r),r} = W^{-1,r}(\Omega)^d \times L_0^r(\Omega),$$

where  $\eta(r) := \frac{1/2-1/r}{1/2-1/R}$ . Its norm is given by

$$\|S^{-1}\| = C_2 \left( \frac{C_R}{C_2} \right)^{\eta(r)}. \quad (2.19)$$

We now define the operator  $S_\mu : W_0^{1,r}(\Omega)^d \times L_0^r(\Omega) \rightarrow W^{-1,r}(\Omega)^d \times L_0^r(\Omega)$  as

$$S_\mu(\mathbf{v}, q) := \begin{pmatrix} -\operatorname{div}(k\mu\nabla\mathbf{v}) + \nabla q \\ \operatorname{div}(\mathbf{v}) \end{pmatrix},$$

for a parameter  $k > 0$  to be chosen later. Note that  $S_\mu$  is a bounded operator and

$$\|S - S_\mu\| \leq \|1 - k\mu\|_{L^\infty}.$$

Therefore,  $S^{-1}(S - S_\mu)$  is bounded as an operator from  $W_0^{1,r}(\Omega)^d \times L_0^r(\Omega)$  to  $W_0^{1,r}(\Omega)^d \times L_0^r(\Omega)$  with

$$\|S^{-1}(S - S_\mu)\| \leq \|S^{-1}\| \|S - S_\mu\| \leq C_2 \left( \frac{C_R}{C_2} \right)^{\eta(r)} \|1 - k\mu\|_{L^\infty}.$$

Assume for the moment that  $k$  is chosen such that  $C_2\|1 - k\mu\|_{L^\infty} < 1$ . Then, a sufficient condition for  $S_\mu = S(I - S^{-1}(S - S_\mu))$  to be invertible is  $C_2 \left( \frac{C_R}{C_2} \right)^{\eta(r)} \|1 - k\mu\|_{L^\infty} < 1$ , that is  $2 \leq r < \tilde{R}(R)$ , where  $\tilde{R}(R)$  is given by

$$\tilde{R}(R) := \operatorname{argmax} \left\{ 2 < r < R \mid C_2 \left( \frac{C_R}{C_2} \right)^{\eta(r)} \|1 - k\mu\|_{L^\infty} < 1 \right\}. \quad (2.20)$$

Moreover, we deduce the following bound for  $\|S_\mu^{-1}\|$  as an operator from  $W^{-1,r}(\Omega)^d \times L_0^r(\Omega)$  to  $W_0^{1,r}(\Omega)^d \times L_0^r(\Omega)$ :

$$\|S_\mu^{-1}\| \leq \frac{\|S^{-1}\|}{1 - C_2 \left( \frac{C_R}{C_2} \right)^{\eta(r)} \|1 - k\mu\|_{L^\infty}} \leq \frac{C_2 \left( \frac{C_R}{C_2} \right)^{\eta(r)}}{1 - C_2 \left( \frac{C_R}{C_2} \right)^{\eta(r)} \|1 - k\mu\|_{L^\infty}}. \quad (2.21)$$

Noting that  $(\mathbf{u}, kp) = S_\mu^{-1}(k\mathbf{f}, g)$ , we have

$$\begin{aligned} \|\nabla\mathbf{u}\|_{L^r} + \|p\|_{L^r} &\leq \max\{1, 1/k\} (\|\nabla\mathbf{u}\|_{L^r} + k\|p\|_{L^r}) \\ &\leq \max\{1, 1/k\} \left( \frac{C_2 \left( \frac{C_R}{C_2} \right)^{\eta(r)}}{1 - C_2 \left( \frac{C_R}{C_2} \right)^{\eta(r)} \|1 - k\mu\|_{L^\infty}} \right) (k\|\mathbf{f}\|_{W^{-1,r}} + \|g\|_{L^r}) \\ &\leq \max\{k, 1/k\} \left( \frac{C_2 \left( \frac{C_R}{C_2} \right)^{\eta(r)}}{1 - C_2 \left( \frac{C_R}{C_2} \right)^{\eta(r)} \|1 - k\mu\|_{L^\infty}} \right) (\|\mathbf{f}\|_{W^{-1,r}} + \|g\|_{L^r}). \end{aligned}$$

The bound (2.18) follows upon setting

$$C_r := \max\{k, 1/k\} \left( \frac{C_2 \left( \frac{C_R}{C_2} \right)^{\eta(r)}}{1 - C_2 \left( \frac{C_R}{C_2} \right)^{\eta(r)} \|1 - k\mu\|_{L^\infty}} \right).$$

It remains to prove that there exists  $k$  satisfying the condition  $C_2\|1 - k\mu\|_{L^\infty} < 1$ . Notice that

$$\begin{aligned} C_2\|1 - k\mu\|_{L^\infty} < 1 &\Leftrightarrow -\frac{1}{C_2} < k\mu(\mathbf{x}) - 1 < \frac{1}{C_2} \quad \text{for a.e. } \mathbf{x} \in \Omega \\ &\Leftrightarrow \frac{C_2 - 1}{C_2\mu(\mathbf{x})} < k < \frac{C_2 + 1}{C_2\mu(\mathbf{x})} \quad \text{for a.e. } \mathbf{x} \in \Omega \end{aligned}$$

Therefore any parameter  $k$  satisfying

$$\max \left( 0, \frac{C_2 - 1}{C_2 \mu_{\min}} \right) < k < \frac{C_2 + 1}{C_2 \mu_{\max}}$$

ensures that  $\|1 - k\mu\|_{L^\infty} < 1$ .  $\square$

**3. Adaptive Finite Element Methods.** The algorithm developed in this paper differs from others in the assumptions made on the viscosity function. More precisely, existing methods assume that the viscosity is at least piecewise smooth on the initial subdivision. Our new procedure does not require such assumption but relies on adaptive methods to approximate (i) the data and (ii) the Stokes system with approximate data and in particular piecewise polynomial viscosity subordinate to the current subdivision. Each of these modules are already available in the literature. We refer for instance to [4, 12] for the discussion on algorithms approximating  $g$  and  $\mathbf{f}$  and to [8] for the approximation of the viscosity which must satisfy the constraint (2.2). AFEM for the Stokes system with viscosity jump aligned with the initial subdivision can be found in [2, 17, 19]. In this section, we present the structure of our main algorithm and discuss each of its modules.

**3.1. Mesh Refinement and Finite Element Approximation.** One of the basic and necessary tools underlying any AFEM is a refinement strategy; refer to [10] for the study of different procedures including bisection, quad-refinement and red-refinement rules with hanging nodes. In this work, we briefly introduce the main ingredients restraining our discussion to the *newest vertex bisection* method for  $d = 2$ . However, what is presented here can be extended to  $d \geq 2$  and other refinement rules satisfying Conditions 3, 4 and 6 in [10].

Let  $\mathcal{T}_0$  be an initial partition of a polyhedral (or polygonal in two dimensions) domain  $\Omega$  into triangles. For a triangle  $T \in \mathcal{T}_0$  and a vertex  $v(T)$  of  $T$ , the bisection method will partition  $T$  into its two children  $T_1$  and  $T_2$  by adding an edge from  $v(T)$  to the middle of the opposite edge. The new created node is then the vertex of  $T_1$  and  $T_2$  used as starting point to bisect those simplices into their own children. This process can then be applied recursively. The key point when using bisection method to partition the whole mesh is an initial labeling. This is not addressed here but one can see the work from Binev *et al.* [3] and Stevenson [23].

It results from this procedure a unique master forest  $\mathbb{F}$  whose root is given by  $\mathcal{T}_0$ . Any partition  $\mathcal{T}$  that can be obtained from several applications of the bisection method is associated to a unique tree  $\mathcal{F} \subset \mathbb{F}$  and is called *admissible*. Given two admissible partitions  $\mathcal{T}$  and  $\mathcal{T}^*$ , we say that  $\mathcal{T}^*$  is a refinement of  $\mathcal{T}$ , and we denote it  $\mathcal{T} \preceq \mathcal{T}^*$ , if their associated forest are such that  $\mathcal{F} \subset \mathcal{F}^*$ . The complexity of a mesh  $\mathcal{T}$  is measured by the number of bisections  $n(\mathcal{T})$  necessary to obtain it from  $\mathcal{T}_0$  since  $\#\mathcal{T} = \#\mathcal{T}_0 + n(\mathcal{T})$ . We denote by  $\mathbb{T}_n$  the set of all admissible partitions obtained from at most  $n$  bisections of  $\mathcal{T}_0$ .

The grids emanating from the bisection method are not necessarily conforming. Additional refinements may be performed by bisecting recursively all the necessary triangles required to ensure conformity. This completion procedure terminates and there exists  $\Lambda_0$  depending only on  $\mathcal{T}_0$  and the initial labeling such that if  $\{\mathcal{T}_k\}_{k=1}^\infty$  is a sequence of conforming meshes obtained successively by refining the *marked* triangles  $\{\mathcal{M}_k\}_{k=0}^\infty$  and possible additional refinements to ensure conformity, the following bound holds

$$\#\mathcal{T}_k - \#\mathcal{T}_0 \leq \Lambda_0 \sum_{j=0}^{k-1} \#\mathcal{M}_j. \quad (3.1)$$

We refer to [3, 23] for a detailed discussion. Estimate (3.1) turns out to be critical to derive quasi-optimal bound on the performance of the AFEM designed for the Stokes system [19].

Let  $\mathcal{T} \in \mathbb{T}_n$  be a conforming partition of  $\Omega$  and consider finite element spaces  $\mathcal{V}(\mathcal{T}) \subset H_0^1(\Omega)^d$  and  $\mathcal{Q}(\mathcal{T}) \subset L_0^2(\Omega)$ . The finite element formulation of the problem (1.3) reads: seek  $(\mathbf{U}, P) \in \mathcal{V}(\mathcal{T}) \times \mathcal{Q}(\mathcal{T})$  such that

$$\int_{\Omega} \mu \nabla \mathbf{U} \cdot \nabla \mathbf{V} - \int_{\Omega} P \operatorname{div}(\mathbf{V}) + \int_{\Omega} Q \operatorname{div}(\mathbf{U}) = \langle \mathbf{f}, \mathbf{V} \rangle + \int_{\Omega} g Q, \quad \forall (\mathbf{V}, Q) \in \mathcal{V}(\mathcal{T}) \times \mathcal{Q}(\mathcal{T}). \quad (3.2)$$

Throughout this paper, we use capital letters to denote members of  $\mathcal{V}(\mathcal{T})$  or  $\mathcal{Q}(\mathcal{T})$ .



**3.2. Structure of the AFEM.** The available AFEMs designed to approximate the solution  $(\mathbf{u}, p)$  of (1.3) assume that if the viscosity is discontinuous, its discontinuities are aligned with the initial subdivision [2, 17, 19]. In all cases, sequences of conforming nested partitions  $\{\mathcal{T}_k\}_{k=0}^\infty$  are generated adaptively together with a subordinate sequence of finite element functions  $\{(\mathbf{U}_k, P_k)\}_{k=0}^\infty$  converging to the solution of (1.3). Notice that  $(\mathbf{U}_k, P_k) \in \mathbf{V}(\mathcal{T}_k) \times \mathcal{Q}(\mathcal{T}_k)$  is not necessarily the solution of (3.2) with  $\mathcal{T} = \mathcal{T}_k$  but could be, for instance, an Uzawa approximation as in [2, 19]. The iterative process is stopped when a prescribed tolerance  $\varepsilon > 0$  on the estimated error is achieved.

In fact, we do not restrict our considerations to a particular algorithm. Instead, given a conforming subdivision  $\mathcal{T}$ , a viscosity  $\hat{\mu}$  satisfying (2.2) with possible discontinuities aligned with  $\mathcal{T}$ , inputs data functions  $\hat{\mathbf{f}}, \hat{g}$  and a target tolerance  $\varepsilon > 0$ , we denote generically by

$$[\mathcal{T}_\varepsilon, (\mathbf{U}_\varepsilon, P_\varepsilon)] = \text{STOKES}(\mathcal{T}, \hat{\mu}, \hat{\mathbf{f}}, \hat{g}, \varepsilon)$$

any algorithm which returns a conforming refinement  $\mathcal{T} \preceq \mathcal{T}_\varepsilon$  and a subordinate finite element solution  $(\mathbf{U}_\varepsilon, P_\varepsilon)$  satisfying

$$\|\hat{\mathbf{u}} - \mathbf{U}_\varepsilon\|_{H_0^1} + \|\hat{p} - P_\varepsilon\|_{L^2} \leq \varepsilon.$$

The main difference between the algorithm introduced in this work and the existing STOKES procedures is that we allow for viscosities with possible discontinuities not necessarily aligned with the subdivision produced. Instead, we generate iteratively adequate approximations of  $\mu$ ,  $\mathbf{f}$  and  $g$  used as input to the STOKES routine. The metrics used to define these approximations is based on the perturbation estimate (2.10). We write

$$[\mathcal{T}_\varepsilon, \mathbf{f}_\varepsilon] = \text{RHS\_MOM}(\mathcal{T}, \mathbf{f}, \varepsilon)$$

the subroutine taking as input a conforming subdivision  $\mathcal{T}_0 \preceq \mathcal{T}$ , a function  $\mathbf{f}$  and a target tolerance  $\varepsilon$  and returning a conforming refinement  $\mathcal{T} \preceq \mathcal{T}_\varepsilon$  together with a piecewise polynomial function  $\mathbf{f}_\varepsilon$  subordinate to  $\mathcal{T}_\varepsilon$  satisfying

$$\|\mathbf{f} - \mathbf{f}_\varepsilon\|_{H^{-1}} \leq \varepsilon. \quad (3.3)$$

Similarly, we write for the approximation of  $g$

$$[\mathcal{T}_\varepsilon, g_\varepsilon] = \text{RHS\_MASS}(\mathcal{T}, g, \varepsilon) \quad \text{such that} \quad \|g - g_\varepsilon\|_{L^2} \leq \varepsilon. \quad (3.4)$$

The viscosity approximation needs to satisfy the additional positivity constraint (2.2). We denote by

$$[\mathcal{T}_\varepsilon, \mu_\varepsilon] = \text{VISC}(\mathcal{T}, \mu, \varepsilon)$$

the routine providing a conforming refinement  $\mathcal{T} \preceq \mathcal{T}_\varepsilon$  and a piecewise polynomial approximation  $\mu_\varepsilon$  subordinate to  $\mathcal{T}_\varepsilon$  satisfying

$$\|\mu - \mu_\varepsilon\|_{L^q} \leq \varepsilon \quad \text{and} \quad 0 < C_\mu^{-1} \mu_{\min} \leq \mu_\varepsilon \leq C_\mu \mu_{\max}.$$

Here  $q := \frac{2r}{r-2}$ , where  $r > 2$  is such that **Condition r** is fulfilled (see Section 2),  $C_\mu$  is an absolute constant and  $\mu_{\min}, \mu_{\max}$  satisfy (2.2).

**REMARK 3.** (*Choice of  $q$* ) Notice that the subroutine VISC requires in general the a-priori knowledge of  $q$  and therefore in what  $L_r$  space  $\nabla \mathbf{u}$  lies. However, when  $\mu$  is piecewise constant with a finite number of pieces, the generic choice  $q = 2$  is always possible independently of the integrability of  $\nabla \mathbf{u}$ . Indeed, for  $q > 2$ , we can take advantage of the interpolation estimate

$$\|\mu - \hat{\mu}\|_{L_q(\Omega)} \leq (2C)^{2/\bar{q}} \|\mu - \hat{\mu}\|_{L_2(\Omega)}^{2/q},$$

where  $\bar{q} := \frac{2p}{p-2}$  and  $C$  is a constant such that  $\|\mu\|_{L_\infty(\Omega)}, \|\hat{\mu}\|_{L_\infty(\Omega)} \leq C$ . We refer to [8, Section 6] for more details.

Several algorithms are able to realize `RHS_MOM`, `RHS_MASS` and `VISC`; see Section 5 for examples.

Given an initial tolerance  $\varepsilon_0 > 0$ , the algorithm we propose to handle general discontinuous viscosity consists in successive applications of the above routines

```

STOKES_DISC( $\mathcal{T}_0, \mathbf{f}, g, \mu, \varepsilon$ )
   $k = 0$ 
  LOOP
    [ $\mathcal{T}_k(\mathbf{f}), \mathbf{f}_k$ ] = RHS_MOM( $\mathcal{T}_k, \mathbf{f}, \omega\varepsilon_k$ );
    [ $\mathcal{T}_k(g), g_k$ ] = RHS_MASS( $\mathcal{T}_k(\mathbf{f}), g, \omega\varepsilon_k$ );
    [ $\mathcal{T}_k(\mu), \mu_k$ ] = VISC( $\mathcal{T}_k(g), \mu, \omega\varepsilon_k$ );
    [ $\mathcal{T}_{k+1}, (\mathbf{U}_{k+1}, P_{k+1})$ ] = STOKES( $\mathcal{T}_k(\mu), \mu_k, \mathbf{f}_k, g_k, \frac{\varepsilon_k}{2}$ );
     $\varepsilon_{k+1} = \xi\varepsilon_k$ ;
     $k \leftarrow k + 1$ ;
  END LOOP
END STOKES_DISC

```

In the above procedure,  $0 < \xi < 1$  and  $\omega$  are positive parameter yet to be determined.

The rationale behind this procedure is that at each iteration  $k$  of `STOKES_DISC`, the constructed finite element approximation  $(\mathbf{U}_{k+1}, P_{k+1})$  satisfies

$$\|\mathbf{u} - \mathbf{U}_{k+1}\|_{H_0^1} + \|p - P_{k+1}\|_{L^2} \leq \varepsilon_k. \quad (3.5)$$

Indeed, it suffices to realize that the call `STOKES`( $\mathcal{T}_k(\mu), \mu_k, \mathbf{f}_k, g_k, \frac{\varepsilon_k}{2}$ ) returns approximations  $(\mathbf{U}_{k+1}, P_{k+1})$  satisfying

$$\|\mathbf{u}_k - \mathbf{U}_{k+1}\|_{H_0^1} + \|p_k - P_{k+1}\|_{L^2} \leq \frac{\varepsilon_k}{2},$$

where  $(\mathbf{u}_k, p_k)$  is the solution of (1.3) with  $\mathbf{f}$ ,  $g$  and  $\mu$  replaced by  $\mathbf{f}_k$ ,  $g_k$  and  $\mu_k$ . Hence, (3.5) follows owing to the perturbation theorem 2.2, the estimates provided in (3.3), (3.4) and choosing  $\omega > 0$  sufficiently small. Notice that the call of the `STOKES` procedure in `STOKES_DISC` is justified since by construction  $\mu_k$ ,  $\mathbf{f}_k$  and  $g_k$  are piecewise polynomial functions subordinate to  $\mathcal{T}_k(\mu)$ .

**4. Performances of `STOKES_DISC`.** In this section we introduce the requirements on the routines presented in Section 3.2 and describe their complexity. Based on these considerations, we provide in Theorem 4.3 the optimal decay rate delivered by `STOKES_DISC` in terms of degrees of freedom invested.

**4.1. Performances of `RHS_MOM` and `RHS_MASS`.** For  $m \geq 0$  and  $n \in \mathbb{N}$ , let  $\Sigma_n^m$  be the set of piecewise polynomial functions of degree at most  $m$  subordinate to a partition  $\mathcal{T} \in \mathbb{T}_n$ . Given a function  $g \in L^2(\Omega)$ , we define the measure of the best approximation of  $g$  on  $\Sigma_n^m$  as

$$\sigma_n(g) := \sigma_n^m(g)_{L^2} := \inf_{S \in \Sigma_n^m} \|g - S\|_{L^2}. \quad (4.1)$$

Based on this definition and given  $s > 0$ , we introduce the space of functions in  $L^2(\Omega)$  which can be approximated by elements of  $\Sigma_n^m$  with decay rate at least  $n^{-s}$ , that is

$$\mathcal{B}^s(L^2(\Omega)) := \mathcal{B}^s(\mathcal{T}_0, L^2(\Omega)) := \left\{ g \in L^2(\Omega) \mid |g|_{\mathcal{B}^s(L^2(\Omega))} := \sup_{n \in \mathbb{N}} (n^s \sigma_n(g)) < \infty \right\}. \quad (4.2)$$

We denote by  $\mathcal{B}^s(L_0^2(\Omega))$  the space of functions in  $\mathcal{B}^s(L^2(\Omega))$  with zero mean value. Similarly, for  $\mathbf{f} \in H^{-1}(\Omega)^d$  we define  $\mathcal{B}^s(H^{-1}(\Omega)^d)$  by replacing  $\|\cdot\|_{L^2}$  by  $\|\cdot\|_{H^{-1}}$ .

Let  $\mathcal{T}$  be a refinement of  $\mathcal{T}_0$ . We then say that `RHS_MASS` is of *class optimal performance* in  $\mathcal{B}^s(L_0^2(\Omega))$  if the number  $N(g) := N(g, \mathcal{T})$  of cells of  $\mathcal{T}$  marked for refinement to achieve the prescribed tolerance  $\varepsilon$  on the error satisfies

$$N(g) \lesssim |g|_{\mathcal{B}^s(L_0^2(\Omega))}^{1/s} \varepsilon^{-1/s}.$$

In a same way, we say that **RHS\_MOM** is of class optimal performance in  $\mathcal{B}^s(H^{-1}(\Omega)^d)$  if

$$N(\mathbf{f}) \lesssim |\mathbf{f}|_{\mathcal{B}^s(H^{-1}(\Omega)^d)}^{1/s} \varepsilon^{-1/s}.$$

Note that in the previous definitions, we have not required that the partition underlying the approximation is conforming. However, in view of estimate (3.1), restricting  $\Sigma_n^m$  to conforming partitions does not affect the above approximation classes; see for instance [8].

**4.2. Performance of VISC.** As for the right hand side functions, we denote by

$$\sigma_n(\mu) := \sigma_n^m(\mu)_{L^q} := \inf_{S \in \Sigma_n^m} \|\mu - S\|_{L^q}$$

the best approximation of  $\mu$  in  $L^q$  norm from piecewise polynomials functions of degree at most  $m$ . We define by  $\mathcal{B}^s(L^q(\Omega)) := \mathcal{B}^s(\mathcal{T}_0, L^q(\Omega))$  the set of functions in  $L^q$  satisfying

$$|\mu|_{\mathcal{B}^s(L^q(\Omega))} := \sup_{n \in \mathbb{N}} (n^s \sigma_n(\mu)) < \infty.$$

This class of functions accounts for  $L^q$  approximability. In analogy to **RHS\_MOM** and **RHS\_MASS**, we say that **VISC** is of class optimal performance in  $\mathcal{B}^s(L^q(\Omega))$  if the number of cells marked for refinement to reach the target tolerance  $\varepsilon$  from  $\mathcal{T}$  satisfies

$$N(\mu) \lesssim |\mu|_{\mathcal{B}^s(L^q(\Omega))}^{1/s} \varepsilon^{-1/s}.$$

We refer again to Section 5.1 for example of class optimal performance algorithms. See [5] for other procedures based on wavelet tree approximation.

**4.3. Performance of STOKES.** We define the class optimality in the same way for **STOKES**. Given  $s > 0$  and an initial partition  $\mathcal{T}_0$  of  $\Omega$ , we consider the following approximation class

$$\mathcal{A}_{\mathbf{V}}^s(\mathcal{T}_0, H_0^1(\Omega)^d) := \left\{ \mathbf{v} \in H_0^1(\Omega)^d \mid |\mathbf{v}|_{\mathcal{A}_{\mathbf{V}}^s(H_0^1(\Omega)^d)} := \sup_{n \geq 1} \left( n^s \inf_{\mathcal{T} \in \mathbb{T}_n} \inf_{\mathbf{V} \in \mathbf{V}(\mathcal{T})} \|\mathbf{v} - \mathbf{V}\|_{H_0^1} \right) < \infty \right\}$$

equipped with the norm  $\|\mathbf{v}\|_{\mathcal{A}_{\mathbf{V}}^s(H_0^1(\Omega)^d)} := \|\mathbf{v}\|_{H_0^1} + |\mathbf{v}|_{\mathcal{A}_{\mathbf{V}}^s(H_0^1(\Omega)^d)}$  for the velocity and

$$\mathcal{A}_{\mathcal{Q}}^s(\mathcal{T}_0, L_0^2(\Omega)) := \left\{ q \in L_0^2(\Omega) \mid |q|_{\mathcal{A}_{\mathcal{Q}}^s(L_0^2(\Omega))} := \sup_{n \geq 1} \left( n^s \inf_{\mathcal{T} \in \mathbb{T}_n} \inf_{Q \in \mathcal{Q}(\mathcal{T})} \|q - Q\|_{L^2} \right) < \infty \right\}$$

equipped with the norm  $\|q\|_{\mathcal{A}_{\mathcal{Q}}^s(L_0^2(\Omega))} := \|q\|_{L^2} + |q|_{\mathcal{A}_{\mathcal{Q}}^s(L_0^2(\Omega))}$  for the pressure. To simplify the notations, we shall write  $\mathcal{A}_{\mathbf{V}}^s(H_0^1(\Omega)^d) := \mathcal{A}_{\mathbf{V}}^s(\mathcal{T}_0, H_0^1(\Omega)^d)$  and  $\mathcal{A}_{\mathcal{Q}}^s(L_0^2(\Omega)) := \mathcal{A}_{\mathcal{Q}}^s(\mathcal{T}_0, L_0^2(\Omega)^d)$ . The **STOKES** routine is said to be of class optimal performance in  $\mathcal{A}_{\mathbf{V}}^s(H_0^1(\Omega)^d) \times \mathcal{A}_{\mathcal{Q}}^s(L_0^2(\Omega))$  if the number  $N(\mathbf{u}, p)$  of cells marked for refinement by the algorithm to reach the tolerance  $\varepsilon > 0$  starting from any subdivision  $\mathcal{T}_0 \preceq \mathcal{T}$  satisfies

$$N(\mathbf{u}, p) \lesssim \left( \|\mathbf{u}\|_{\mathcal{A}_{\mathbf{V}}^s(H_0^1(\Omega)^d)}^{1/s} + \|p\|_{\mathcal{A}_{\mathcal{Q}}^s(L_0^2(\Omega))}^{1/s} \right) \varepsilon^{-1/s}.$$

At this point, we emphasize that even if the solution  $(\mathbf{u}, p)$  of the original Stokes system (2.3) belongs to  $\mathcal{A}_{\mathbf{V}}^s(H_0^1(\Omega)^d) \times \mathcal{A}_{\mathcal{Q}}^s(L_0^2(\Omega))$ , the solution  $(\hat{\mathbf{u}}, \hat{p})$  of the perturbed problem (2.4) does not necessarily belong to the same approximation class. Therefore, to study the complexity of the routine **STOKES** within our AFEM **STOKES\_DISC** (called with approximate data), we require that **STOKES** fulfills a slightly different condition to derive the optimality of our method; see Section 5.3 for examples of such procedures. In this aim, we define the concept of  $\varepsilon$ -approximability originally introduced by Bonito, DeVore and Nochetto [8].

**DEFINITION 4.1.** ( $\varepsilon$ -approximation of order  $s$ ) *Let  $p \in \mathcal{A}_{\mathcal{Q}}^s(L_0^2(\Omega))$  and  $\varepsilon > 0$  be given. Then  $q$  is said to be an  $\varepsilon$ -approximation of order  $s$  of  $p$  if  $\|p - q\|_{L^2} \leq \varepsilon$  and there exists a constant  $K$  independent of  $\varepsilon$ ,  $p$  and  $q$  such that for each  $\delta \geq \varepsilon$ , there exists  $n \in \mathbb{N}$  satisfying*

$$\inf_{\mathcal{T} \in \mathbb{T}_n} \inf_{Q \in \mathcal{Q}(\mathcal{T})} \|q - Q\|_{L^2} \leq \delta, \quad n \leq K |p|_{\mathcal{A}_{\mathcal{Q}}^s(L_0^2(\Omega))}^{1/s} \delta^{-1/s}. \quad (4.3)$$

Similarly, for  $\mathbf{u} \in \mathcal{A}_{\mathbf{V}}^s(H_0^1(\Omega)^d)$  and  $\varepsilon > 0$ ,  $\mathbf{v}$  is said to be an  $\varepsilon$ -approximation of order  $s$  of  $\mathbf{u}$  if  $\|\mathbf{u} - \mathbf{v}\|_{H^1} \leq \varepsilon$  and there exists a constant  $K$  independent of  $\varepsilon$ ,  $\mathbf{u}$  and  $\mathbf{v}$  such that for each  $\delta \geq \varepsilon$ , there exists  $n \in \mathbb{N}$  satisfying

$$\inf_{\mathcal{T} \in \mathbb{T}_n} \inf_{\mathbf{v} \in \mathbf{V}(\mathcal{T})} \|\mathbf{u} - \mathbf{v}\|_{H^1} \leq \delta, \quad n \leq K \|\mathbf{u}\|_{\mathcal{A}_{\mathbf{V}}^s(H_0^1(\Omega)^d)}^{1/s} \delta^{-1/s}. \quad (4.4)$$

We say that  $(\hat{\mathbf{u}}, \hat{p})$  is an  $\varepsilon$ -approximation of order  $s$  of  $(\mathbf{u}, p)$  if  $\hat{\mathbf{u}}$  and  $\hat{p}$  are  $\varepsilon$ -approximation of order  $s$  of  $\mathbf{u}$  and  $p$ , respectively. We now recall a result which characterizes  $\varepsilon$ -approximation of element in  $\mathcal{A}_{\mathbf{V}}^s(H_0^1(\Omega)^d)$  and  $\mathcal{A}_{\mathbf{Q}}^s(L_0^2(\Omega))$ ; see [8] for a proof.

LEMMA 4.2. ( $\varepsilon$ -approximation of order  $s$ ) Let  $\varepsilon > 0$ .

- Assume that  $p \in \mathcal{A}_{\mathbf{Q}}^s(L_0^2(\Omega))$  and  $q \in L_0^2(\Omega)$  satisfy  $\|p - q\|_{L^2} \leq \frac{\varepsilon}{2}$ . Then  $q$  is an  $\varepsilon$ -approximation of order  $s$  of  $p$ .
- Assume that  $\mathbf{u} \in \mathcal{A}_{\mathbf{V}}^s(H_0^1(\Omega)^d)$  and  $\mathbf{v} \in H_0^1(\Omega)^d$  satisfy  $\|\mathbf{u} - \mathbf{v}\|_{H^1} \leq \frac{\varepsilon}{2}$ . Then  $\mathbf{v}$  is an  $\varepsilon$ -approximation of order  $s$  of  $\mathbf{u}$ .

We are now in position to describe the condition required on the STOKES procedure to derive the complexity bound provided in Theorem 4.3.

**( $\gamma\varepsilon$ )-approximation optimality condition of order  $s$ .** Let  $(\mathbf{u}, p)$  be the solution of the Stokes system (2.3). We say that for  $0 < \gamma < 1$  the STOKES routine satisfies the  $(\gamma\varepsilon)$ -approximation optimality condition if there exists a constant  $C$  only depending on  $(\mathbf{u}, p)$  such that for any  $\varepsilon > 0$ , any  $(\hat{\mathbf{u}}, \hat{p})$  solution of the perturbed problem (2.4) which is a  $(\gamma\varepsilon)$ -approximation of order  $s$  of  $(\mathbf{u}, p)$  and any subdivision  $\mathcal{T} \succeq \mathcal{T}_0$ , the number  $N(\hat{\mathbf{u}}, \hat{p})$  of cells marked for refinement by  $\text{STOKES}(\mathcal{T}, \hat{\mu}, \hat{\mathbf{f}}, \hat{g}, \varepsilon)$  satisfies

$$N(\hat{\mathbf{u}}, \hat{p}) \leq C \left( \|\mathbf{u}\|_{\mathcal{A}_{\mathbf{V}}^s(H_0^1(\Omega)^d)}^{1/s} + \|p\|_{\mathcal{A}_{\mathbf{Q}}^s(L_0^2(\Omega))}^{1/s} \right) \varepsilon^{-1/s}, \quad (4.5)$$

provided  $(\mathbf{u}, p) \in \mathcal{A}_{\mathbf{V}}^s(H_0^1(\Omega)^d) \times \mathcal{A}_{\mathbf{Q}}^s(L_0^2(\Omega))$ .

Notice that estimate (4.5) does not involve the data  $\hat{\mathbf{f}}$  and  $\hat{g}$ . Generically, this is only possible when  $\hat{\mathbf{f}}$  and  $\hat{g}$  are captured by the finite element method (no oscillation). This is guaranteed within **STOKES\_DISC** where  $\hat{\mathbf{f}}$  and  $\hat{g}$  are polynomial approximations of  $\mathbf{f}$  and  $g$ . We provide in Section 5.3 an argument showing that the adaptive algorithm for the Stokes system introduced in [19] fulfills this condition.

We conclude this section by mentioning an alternative to the notion of  $\varepsilon$ -approximation.

REMARK 4 (Alternative Equivalent Point of View). Consider the following semi-norm

$$|\mathbf{v}|_{\mathcal{A}_{\mathbf{V}}^s(H_0^1(\Omega)^d)} := \sup_{0 < \delta \leq \|\mathbf{v}\|_{H_0^1}} \delta \inf_{\substack{\mathcal{T} \in \mathbb{T} \text{ s.t.} \\ \inf_{\mathbf{v} \in \mathbf{V}(\mathcal{T})} \|\mathbf{v} - \mathbf{V}\|_{H_0^1} \leq \delta}} (\#\mathcal{T} - \#\mathcal{T}_0)^s,$$

where  $\mathbb{T}$  stands for the set of all admissible partitions. Its associated norm  $\|\cdot\|_{\mathcal{A}_{\mathbf{V}}^s(H_0^1(\Omega)^d)} := |\cdot|_{\mathcal{A}_{\mathbf{V}}^s(H_0^1(\Omega)^d)} + \|\cdot\|_{H_0^1}$  is equivalent to  $\|\cdot\|_{\mathcal{A}_{\mathbf{V}}^s(H_0^1(\Omega)^d)}$  [22]. Recalling that the solution  $(\hat{\mathbf{u}}, \hat{p})$  are required to be approximated with a rate  $s$  only up to a certain threshold, it is natural to consider for any parameter  $\bar{\delta}$  the spaces

$$\mathcal{A}_{\mathbf{V}}^s(\bar{\delta}, \mathcal{T}_0, H_0^1(\Omega)^d) := \left\{ \mathbf{v} \in H_0^1(\Omega)^d \mid |\mathbf{v}|_{\mathcal{A}_{\mathbf{V}}^s(\bar{\delta}, H_0^1(\Omega)^d)} := \sup_{\bar{\delta} < \delta \leq \|\mathbf{v}\|_{H_0^1}} \delta \inf_{\substack{\mathcal{T} \in \mathbb{T} \text{ s.t.} \\ \inf_{\mathbf{v} \in \mathbf{V}(\mathcal{T})} \|\mathbf{v} - \mathbf{V}\|_{H_0^1} \leq \delta}} (\#\mathcal{T} - \#\mathcal{T}_0)^s < \infty \right\}.$$

Reproducing the argumentation leading to Lemma 4.2, one realize that if  $\mathbf{v} \in \mathcal{A}_{\mathbf{V}}^s(\mathcal{T}_0, H_0^1(\Omega)^d)$  and  $\hat{\mathbf{v}} \in H_0^1(\Omega)^d$  satisfy  $\|\mathbf{v} - \hat{\mathbf{v}}\|_{H_0^1} \leq \frac{\bar{\delta}}{2}$ , then  $\hat{\mathbf{v}} \in \mathcal{A}_{\mathbf{V}}^s(\bar{\delta}, \mathcal{T}_0, H_0^1(\Omega)^d)$  and

$$|\hat{\mathbf{v}}|_{\mathcal{A}_{\mathbf{V}}^s(\bar{\delta}, H_0^1(\Omega)^d)} \lesssim |\mathbf{v}|_{\mathcal{A}_{\mathbf{V}}^s(H_0^1(\Omega)^d)}.$$

As a consequence, requirement (4.5) on the STOKES procedure can then be replaced by

$$N(\hat{\mathbf{u}}, \hat{p}) \leq C \left( \|\hat{\mathbf{u}}\|_{\mathcal{A}_{\mathbf{V}}^s(\gamma\varepsilon, H_0^1(\Omega)^d)}^{1/s} + \|\hat{p}\|_{\mathcal{A}_{\mathbf{Q}}^s(\gamma\varepsilon, L_0^2(\Omega))}^{1/s} \right) \varepsilon^{-1/s},$$

where the function spaces  $\mathcal{A}_{\mathbf{Q}}^s(\bar{\delta}, \mathcal{T}_0, L_0^2(\Omega))$  associated to the pressure are defined similarly. The above condition is also fulfilled by the algorithm discussed in Section 5.3.

**4.4. Optimality of STOKES\_DISC.** We assume to be given routines RHS\_MOM, RHS\_MASS and VISC of class optimal performance for  $0 < s \leq S$ . Recall that the approximations of the viscosity  $\hat{\mu}$  obtained after a call of VISC satisfy

$$C_\mu^{-1} \mu_{\min} \leq \hat{\mu} \leq C_\mu \mu_{\max}, \quad \text{a.e. in } \Omega, \quad (4.6)$$

where  $C_\mu$  is an absolute constant. Regarding the approximation of the velocity and pressure, we assume that the STOKES routine satisfies the  $(\gamma\varepsilon)$ -**approximation optimality condition of order  $s$**  for some  $0 < \gamma \leq 1$  and for  $0 < s \leq S$ . Let  $q := \frac{2r}{r-2} < \infty$  be such that **Condition r** is satisfied with  $r > 2$  and recall that  $C_r$  denotes the continuity constant in (2.18). The main algorithm STOKES\_DISC depends on two parameters,  $0 < \xi < 1$  and  $\omega > 0$ . We restrict the latter as to satisfy

$$\omega \leq \frac{\gamma}{2D(2 + C_r(\|f\|_{W^{-1,r}} + \|g\|_{L^r}))} \leq 1, \quad (4.7)$$

where  $D$  is the constant in the perturbation estimate (2.10) explicitly provided in Remark 1 with  $\hat{\mu}_{\min} := C_\mu^{-1} \mu_{\min}$  and  $\hat{\mu}_{\max} := C_\mu \mu_{\max}$ .

The following result provides the optimality of the STOKES\_DISC algorithm described in Section 3.2.

**THEOREM 4.3.** (Optimality of STOKES\_DISC) *Let  $\mathcal{T}_0$  be an initial partition of  $\Omega$ ,  $0 < s \leq S$  and assume that **Condition r** holds for some  $r > 2$ . Let  $\mathbf{f} \in \mathcal{B}^s(H^{-1}(\Omega)^d)$ ,  $g \in \mathcal{B}^s(L_0^2(\Omega))$  and  $\mu \in \mathcal{B}^s(L^q(\Omega))$  satisfying the bounds (1.4) with  $q := \frac{2r}{r-2}$ . Moreover, assume that  $(\mathbf{u}, p) \in \mathcal{A}_{\mathcal{V}}^s(H_0^1(\Omega)^d) \times \mathcal{A}_{\mathcal{Q}}^s(L_0^2(\Omega))$ . If  $(\mathbf{U}_{k+1}, P_{k+1}) \in \mathcal{V}(\mathcal{T}_{k+1}) \times \mathcal{Q}(\mathcal{T}_{k+1})$  denotes the solution after the  $k$ th iteration of the STOKES\_DISC algorithm, then for  $k \geq 0$ , we have*

$$\|\mathbf{u} - \mathbf{U}_{k+1}\|_{H_0^1} + \|p - P_{k+1}\|_{L^2} \leq \varepsilon_k, \quad (4.8)$$

and

$$\#\mathcal{T}_{k+1} - \#\mathcal{T}_0 \lesssim \left( \|\mathbf{u}\|_{\mathcal{A}_{\mathcal{V}}^s(H_0^1(\Omega)^d)}^{1/s} + \|p\|_{\mathcal{A}_{\mathcal{Q}}^s(L_0^2(\Omega))}^{1/s} + |\mu|_{\mathcal{B}^s(L^q(\Omega))}^{1/s} + |\mathbf{f}|_{\mathcal{B}^s(H^{-1}(\Omega))}^{1/s} + |g|_{\mathcal{B}^s(L^2(\Omega))}^{1/s} \right) \varepsilon_k^{-1/s}. \quad (4.9)$$

*Proof.* [1] We first prove (4.8). Let  $(\mathbf{u}_k, p_k)$  denotes the solution of (1.3) with coefficients  $\mu_k$ ,  $\mathbf{f}_k$  and  $g_k$  delivered by the subroutines VISC, RHS\_MOM and RHS\_MASS at iteration  $k$ . In particular, there holds

$$\|\mu - \mu_k\|_{L^q} \leq \omega \varepsilon_k, \quad \|\mathbf{f} - \mathbf{f}_k\|_{H^{-1}} \leq \omega \varepsilon_k, \quad \|g - g_k\|_{L^2} \leq \omega \varepsilon_k.$$

and

$$C_\mu^{-1} \mu_{\min} \leq \mu_k \leq C_\mu \mu_{\max},$$

where  $C_\mu$  is the constant appearing in (4.6). Hence, the perturbation estimate (2.10) together with the  $L^r$  bound (2.18) yield

$$\|\mathbf{u} - \mathbf{u}_k\|_{H_0^1} + \|p - p_k\|_{L^2} \leq D(2\omega \varepsilon_k + \omega \varepsilon_k C_r(\|\mathbf{f}\|_{W^{-1,r}} + \|g\|_{L^r})) \leq \frac{\gamma \varepsilon_k}{2}, \quad (4.10)$$

where the last inequality follows from the restriction (4.7) on  $\omega$ . Inequality (4.8) follows noting that

$$\|\mathbf{u} - \mathbf{U}_{k+1}\|_{H_0^1} + \|p - P_{k+1}\|_{L^2} \leq \|\mathbf{u} - \mathbf{u}_k\|_{H_0^1} + \|p - p_k\|_{L^2} + \|\mathbf{u}_k - \mathbf{U}_{k+1}\|_{H_0^1} + \|p_k - P_{k+1}\|_{L^2} \leq \varepsilon_k,$$

where we have used  $\|\mathbf{u}_k - \mathbf{U}_{k+1}\|_{H_0^1} + \|p_k - P_{k+1}\|_{L^2} \leq \frac{\varepsilon_k}{2}$  and  $\gamma \leq 1$ .

[2] It remains to prove (4.9). At iteration  $j$  with  $0 \leq j \leq k$ , the STOKES\_DISC procedure creates a new partition  $\mathcal{T}_{j+1}$  by refining the marked cells and possibly others to ensure conformity. Let us denote by  $N_j(\mathbf{f})$ ,  $N_j(g)$ ,  $N_j(\mu)$  and  $N_j(\mathbf{u}_j, p_j)$  the cells selected for refinement at step  $j$  by the subroutines RHS\_MOM, RHS\_MASS, VISC and STOKES, respectively. Since the three first routines are assumed to be of class optimal performance of order  $0 < s < S$ , we have

$$N_j(\mathbf{f}) \lesssim |\mathbf{f}|_{\mathcal{B}^s(H^{-1}(\Omega))}^{1/s} (\omega \varepsilon_j)^{-1/s}, \quad N_j(g) \lesssim |g|_{\mathcal{B}^s(L_0^2(\Omega))}^{1/s} (\omega \varepsilon_j)^{-1/s} \text{ and } N_j(\mu) \lesssim |\mu|_{\mathcal{B}^s(L^q(\Omega))}^{1/s} (\omega \varepsilon_j)^{-1/s}. \quad (4.11)$$

In addition, the estimate (4.10) obtained in step 1 coupled with Lemma 4.2 ensure that  $(\mathbf{u}_j, p_j)$  is an  $(\gamma\varepsilon_j)$ -approximation of order  $s$  of  $(\mathbf{u}, p)$ . Hence, the  $(\gamma\varepsilon)$ -**approximation optimality condition of order  $s$**  assumed on STOKES implies that

$$N_j(\mathbf{u}_j, p_j) \lesssim \left( \|\mathbf{u}\|_{\mathcal{A}_{\mathfrak{V}}^s(H_0^1(\Omega))}^{1/s} + \|p\|_{\mathcal{A}_{\mathfrak{Q}}^s(L_0^2(\Omega))}^{1/s} \right) \varepsilon_j^{-1/s}. \quad (4.12)$$

At this point, only the number of marked triangles are taken into account. However, in view of estimate (3.1), we deduce a bound on the total number of elements

$$\#\mathcal{T}_{k+1} - \#\mathcal{T}_0 \leq \Lambda_0 \sum_{j=0}^k (N_j(\mathbf{f}) + N_j(g) + N_j(\mu) + N_j(\mathbf{u}_j, p_j)). \quad (4.13)$$

Gathering the estimates (4.11), (4.12) and (4.13) and recalling that  $\omega \leq 1$ , we get

$$\#\mathcal{T}_{k+1} - \#\mathcal{T}_0 \lesssim \left( \|\mathbf{u}\|_{\mathcal{A}_{\mathfrak{V}}^s(H_0^1(\Omega)^d)}^{1/s} + \|p\|_{\mathcal{A}_{\mathfrak{Q}}^s(L_0^2(\Omega))}^{1/s} + \|\mathbf{f}\|_{\mathcal{B}^s(H^{-1}(\Omega)^d)}^{1/s} + \|\mu\|_{\mathcal{B}^s(L^q(\Omega))}^{1/s} + \|g\|_{\mathcal{B}^s(L_0^2(\Omega))}^{1/s} \right) \sum_{j=0}^k (\omega\varepsilon_j)^{-1/s}.$$

It remains to use the properties  $\varepsilon_j := \xi^{j-k}\varepsilon_k$  and  $0 < \xi < 1$  to derive

$$\sum_{j=0}^k \varepsilon_j^{-1/s} \leq \frac{\varepsilon_k^{-1/s}}{1 - \xi^{1/s}}.$$

Estimate (4.9) follows gathering the last two relations.  $\square$

**5. Instances of RHS\_MOM, RHS\_MASS, VISC and STOKES.** The optimality of STOKES\_DISC provided by Theorem 4.3 relies on the class optimal performance assumption on RHS\_MOM, RHS\_MASS and VISC and the  $(\gamma\varepsilon)$ -approximation optimality condition for the STOKES subroutine; see Section 4. We present in this section algorithms that fulfill these conditions.

**5.1. Greedy Algorithm to Approximate General Functions.** The greedy procedure [4, 6, 7, 11, 12] adaptively constructs approximations of functions in different norms and can be used for RHS\_MOM, RHS\_MASS and as a basis for VISC - recall that the output of the latter is required to satisfy the extra condition (4.6). We briefly discuss a greedy algorithm to approximate functions in a  $L^q$  norm and refer the reader to [12, 22] for the case of  $H^{-1}$  data. For completeness, we mention that other algorithms are available such as procedures based on tree approximations [5].

Let  $f \in L^q(\Omega)$  and  $\mathcal{T}$  be a refinement of the initial partition  $\mathcal{T}_0$ . Given a polynomial degree  $m \geq 0$  and an exponent  $q \geq 2$  for the approximation norm, we define the local  $L^q$  error of  $f$  on  $T \in \mathcal{T}$  as

$$E(T) := E(f, T)_{m,q} := \inf_{S_T \in \mathbb{P}_m(T)} \|f - S_T\|_{L^q(T)}, \quad (5.1)$$

where  $\mathbb{P}_m(T)$  is the set of polynomials of degree at most  $m$  on  $T$ . Based on (5.1), the best approximation of  $f \in L^q(\Omega)$  on  $\mathcal{T}$  by piecewise polynomial functions of degree at most  $m$  is given by

$$S_{\mathcal{T}} := \sum_{T \in \mathcal{T}} S_T \chi_T,$$

where  $S_T \in \mathbb{P}_m(T)$  is such that  $E(T) = \|f - S_T\|_{L^q(T)}$  and  $\chi_T$  is the characteristic function over  $T$ . The total error between  $f$  and  $S_{\mathcal{T}}$  is then

$$\mathcal{E}(\mathcal{T}) := \mathcal{E}(f, \mathcal{T})_{m,q} := \left( \sum_{T \in \mathcal{T}} E(f, T)_{m,q}^q \right)^{1/q}.$$

The greedy procedure is based on the local error defined in (5.1) and selects for refinement the element where the error is the largest. Given a target tolerance  $\varepsilon > 0$  and a refinement  $\mathcal{T}$  of  $\mathcal{T}_0$ , it reads

```

GREEDY( $\mathcal{T}, f, \varepsilon$ )
  WHILE  $\mathcal{E}(\mathcal{T}) > \varepsilon$ 
     $T = \operatorname{argmax} \{E(T) \mid T \in \mathcal{T}\};$ 
     $\mathcal{T} = \text{BISECT}(T, \mathcal{T});$ 
  END WHILE
   $\mathcal{T} \leftarrow \text{MAKECONF}(\mathcal{T});$ 
END GREEDY

```

where  $\text{BISECT}(T, \mathcal{T})$  is a function that bisects  $T$  into its two children and returns the grid containing the children. It inputs a mesh and the element of this mesh that needs to be refined. The  $\text{MAKECONF}$  function is a subroutine that makes a triangulation conform. Both procedures have been discussed in Section 3.1. In particular, we have seen that it is possible to control the number of cells refined by imposing conformity invoking (3.1).

The greedy is not optimal on the entire class  $\mathcal{B}^s(L^q(\Omega))$  but on subclasses. The following proposition shows the optimality of **GREEDY** on certain Besov spaces. It is initially due to Binev *et al.* [4] when the algorithm starts from the initial partition  $\mathcal{T}_0$  and Bonito *et al.* [8] have extended this result to general refinement  $\mathcal{T}$  of  $\mathcal{T}_0$ . In what follows,  $B_\tau^s(L^\tau(\Omega))$   $s, \tau > 0$  denote Besov spaces; see [14, 24].

**PROPOSITION 5.1.** (Optimality of the **GREEDY**) *Let  $f \in B_\tau^s(L^\tau(\Omega))$  with  $\frac{s}{d} > \frac{1}{\tau} - \frac{1}{q}$  and  $0 < s \leq m+1$ . The number  $N(f)$  of cells marked for refinement by **GREEDY** to reach the tolerance  $\varepsilon > 0$  starting from  $\mathcal{T}$  satisfies*

$$N(f) \lesssim |f|_{B_\tau^s(L^\tau(\Omega))}^{d/s} \varepsilon^{-d/s}, \quad (5.2)$$

which implies that  $f \in \mathcal{B}^{s/d}(L^q(\Omega))$  with  $|f|_{\mathcal{B}^{s/d}(L^q(\Omega))} \lesssim |f|_{B_\tau^s(L^\tau(\Omega))}$ .

The proofs of this type of results usually assume that the input mesh of the algorithm is the initial partition  $\mathcal{T}_0$ . It is the case for the one presented in [4, 12]. However, since the **GREEDY** will be the basis of the AFEM used to approximate our data functions, we need to be able to call it from any refinement  $\mathcal{T}$  of  $\mathcal{T}_0$ . The following monotonicity property holds for the local error defined in (5.1)

$$E(T_1)^q + E(T_2)^q \leq E(T)^q, \quad T \in \mathcal{T}, \quad (5.3)$$

where  $T_1$  and  $T_2$  are the two children of  $T$  obtained by bisection. Based on (5.3), Bonito *et al.* [8] showed that (5.2) also holds starting from any refinement  $\mathcal{T}$  of  $\mathcal{T}_0$ . In particular, they proved that for any refinement  $\mathcal{T}$  of  $\mathcal{T}_0$ , we have

$$N(f, \mathcal{T}) \leq N(f, \mathcal{T}_0),$$

where  $N(f, \mathcal{T})$  and  $N(f, \mathcal{T}_0)$  are the number of cells marked by **GREEDY** to reach the prescribed tolerance starting from  $\mathcal{T}$  and  $\mathcal{T}_0$ , respectively.

**5.2. Algorithms to Approximate the Data.** We are now in position to introduce instances of subroutines **RHS\_MOM**, **RHS\_MASS** and **VISC**. All three are based on the **GREEDY** algorithm introduced in the previous section. For **RHS\_MASS**, since  $g \in L_0^2(\Omega)$ , the procedure is the same except that a projection onto the space of functions with zero mean value function is performed as a post-processing without affecting the complexity estimate (5.2). Concerning **RHS\_MOM**,  $\mathbf{f} \in L^2(\Omega)^d$  can be approximated directly by the above **GREEDY** procedure (with  $q = 2$  and approximating each component separately). We refer to [12, 22] for the  $H^{-1}(\Omega)^d$  case.

The function **VISC** requires in addition that the approximation  $\hat{\mu}$  satisfies the constraint

$$c\mu_{\min} \leq \hat{\mu}_{\min} \leq \hat{\mu}_{\max} \leq C\mu_{\max}, \quad (5.4)$$

for absolute constants  $c$  and  $C$ . In [8], Bonito, DeVore and Nocketto showed that one could take  $c = \frac{1}{2}$  by modifying the approximation  $\hat{\mu}$  defined on the partition given by the **GREEDY** algorithm on each cell  $\bar{T}$  by either setting (i)  $\hat{\nu}|_T = \hat{\mu}|_T$ ; (ii)  $\hat{\nu}|_T = \mu_{\min}$  or (iii)  $\hat{\nu} = \hat{\mu} + \alpha$ , for some  $\alpha > 0$ . It turns out that in that case the approximation of  $\mu$  is not deteriorated, i.e. for some absolute constant  $C$ ,  $\|\mu - \hat{\nu}\|_{L^q} \leq C\epsilon$  provided  $\|\mu - \hat{\mu}\|_{L^q} \leq \epsilon$ . Refer to [8] for a detailed discussion.

**5.3. The STOKES Procedure.** In this section, we discuss a possible instance of STOKES designed and proved to be of class optimal performance in  $\mathcal{A}_{\mathbf{V}}^s(H_0^1(\Omega)^d) \times \mathcal{A}_{\mathcal{Q}}^s(L_0^2(\Omega))$  by Kondratyuk and Stevenson in [19]. The finite element space  $\mathbf{V}$  is based on continuous piecewise polynomials of degree  $n \geq 1$  while  $\mathcal{Q}$  uses piecewise polynomials of degree  $n - 1$ . We describe this algorithm and show that it satisfies for some  $\gamma < 1$  the  $(\gamma\varepsilon)$ -**approximation optimality condition** introduced in Section 4 and critical to achieve the complexity estimate provided in Theorem 4.3.

The method is based on an Uzawa algorithm which takes advantage of the elliptic property of the Schur complement. Its adaptive strategy relies on the following sub-problems. Given a subdivision  $\mathcal{T}$ , we denote  $(\hat{\mathbf{u}}_{\mathcal{T}}, \hat{P}_{\mathcal{T}}) \in H_0^1(\Omega)^d \times \mathcal{Q}(\mathcal{T})$  the solution of the reduced Stokes problem

$$\int_{\Omega} \hat{\mu} \nabla \hat{\mathbf{u}}_{\mathcal{T}} \cdot \nabla \mathbf{v} - \int_{\Omega} \hat{P}_{\mathcal{T}} \operatorname{div}(\mathbf{v}) + \int_{\Omega} Q \operatorname{div}(\hat{\mathbf{u}}_{\mathcal{T}}) = \langle \hat{\mathbf{f}}, \mathbf{v} \rangle + \int_{\Omega} \hat{g} Q, \quad \forall (\mathbf{v}, Q) \in H_0^1(\Omega)^d \times \mathcal{Q}(\mathcal{T}), \quad (5.5)$$

which relaxes the divergence free constraint compared to (2.4). Given a pressure approximation  $\hat{P} \in \mathcal{Q}(\mathcal{T})$  subordinate to the subdivision  $\mathcal{T}$ , we also define the solution  $\hat{\mathbf{u}}^{\hat{P}} \in H_0^1(\Omega)^d$  of the elliptic problem

$$\int_{\Omega} \hat{\mu} \nabla \hat{\mathbf{u}}^{\hat{P}} \cdot \nabla \mathbf{v} = \langle \hat{\mathbf{f}}, \mathbf{v} \rangle + \int_{\Omega} \hat{P} \operatorname{div}(\mathbf{v}), \quad \forall \mathbf{v} \in H_0^1(\Omega)^d. \quad (5.6)$$

We denote by  $\hat{\mathbf{U}}^{\hat{P}}$ , the corresponding finite element solution subordinate to a refinement  $\mathcal{T}^* \succeq \mathcal{T}$ , i.e.  $\hat{\mathbf{U}}^{\hat{P}} \in \mathbf{V}(\mathcal{T}^*)$  solves

$$\int_{\Omega} \hat{\mu} \nabla \hat{\mathbf{U}}^{\hat{P}} \cdot \nabla \mathbf{V} = \langle \hat{\mathbf{f}}, \mathbf{V} \rangle + \int_{\Omega} \hat{P} \operatorname{div}(\mathbf{V}), \quad \forall \mathbf{V} \in \mathbf{V}(\mathcal{T}^*). \quad (5.7)$$

We are now in position to describe the adaptive algorithm  $\text{STOKES}(\mathcal{T}, \hat{\mu}, \hat{\mathbf{f}}, \hat{g}, \varepsilon)$  introduced and studied in [19] to approximate the solution  $(\hat{\mathbf{u}}, \hat{p})$  of the perturbed Stokes problem (2.4). It consists in three nested loops indexed by  $i, j, k$  generating subdivisions denoted by  $\mathcal{T}_{j,k}^i$ .

For fixed  $i$  and  $j$ , the loop indexed by  $k$  intends to approximate the solution  $\hat{\mathbf{u}}^{\hat{P}_j^i} \in H_0^1(\Omega)^d$  of (5.6) with  $\hat{P} = \hat{P}_j^i \in \mathcal{Q}(\mathcal{T}_{0,0}^i)$  given. The local residual error estimators on the cells  $T \in \mathcal{T}_{j,k}^i$

$$\operatorname{diam}(T)^2 \|\hat{\mathbf{f}} - \nabla \hat{P}_j^i + \operatorname{div}(\hat{\mu} \nabla \hat{\mathbf{U}}_{j,k}^i)\|_{L^2(T)}^2 + \operatorname{diam}(T) \|[\hat{P}_j^i \mathbf{n} - \hat{\mu} \nabla \hat{\mathbf{U}}_{j,k}^i \cdot \mathbf{n}]\|_{L^2(\partial T)}^2$$

is used to drive the Dörfler refinement strategy [15] and produce  $\mathcal{T}_{j,k+1}^i \succeq \mathcal{T}_{j,k}^i$ . Here  $[\![\cdot]\!]$  denotes the jump over the cell boundary,  $\mathbf{n}$  the outward unit normal vector and  $\hat{\mathbf{U}}_{j,k}^i \in \mathbf{V}(\mathcal{T}_{j,k}^i)$  is the solution of (5.7) with  $\hat{P} = \hat{P}_j^i$ . If  $\hat{\mathbf{U}}_{j,k+1}^i \in \mathbf{V}(\mathcal{T}_{j,k+1}^i)$  denotes the solution of (5.7) on the subdivision  $\mathcal{T}_{j,k+1}^i$  with  $\hat{P} = \hat{P}_j^i$ , the following contraction property holds

$$\|\hat{\mathbf{u}}^{\hat{P}_j^i} - \hat{\mathbf{U}}_{j,k+1}^i\|_{H_0^1} \leq \rho_1 \|\hat{\mathbf{u}}^{\hat{P}_j^i} - \hat{\mathbf{U}}_{j,k}^i\|_{H_0^1} \quad (5.8)$$

for some  $\rho_1 < 1$ . Moreover, the number of marked cells  $\#\mathcal{M}_{j,k}^i$  during this step satisfies

$$\#\mathcal{M}_{j,k}^i \lesssim \|\hat{\mathbf{u}}^{\hat{P}_j^i} - \hat{\mathbf{U}}_{j,k}^i\|_{H_0^1}^{-1/s} \left( (\|\hat{\mathbf{u}}\|_{\mathcal{A}_{\mathbf{V}}^s(H_0^1(\Omega)^d)} + \|\hat{\mathbf{u}}\|_{H_0^1})^{1/s} + (\|\hat{p}\|_{\mathcal{A}_{\mathcal{Q}}^s(L_0^2(\Omega))} + \|\hat{p}\|_{L^2})^{1/s} \right), \quad (5.9)$$

provided  $(\hat{\mathbf{u}}, \hat{p}) \in \mathcal{A}_{\mathbf{V}}^s(H_0^1(\Omega)^d) \times \mathcal{A}_{\mathcal{Q}}^s(L_0^2(\Omega))$ .

The loop in  $j$  corresponds to the Uzawa pressure update

$$\hat{P}_{j+1}^i = \hat{P}_j^i - Q_{\mathcal{T}_{0,0}^i}(\operatorname{div}(\hat{\mathbf{U}}_{j,k}^i) - \hat{g}),$$

which satisfies for some  $\rho_2 < 1$

$$\|\hat{P}_{\mathcal{T}_{0,0}^i} - \hat{P}_{j+1}^i\|_{L^2} \leq \rho_2 \|\hat{P}_{\mathcal{T}_{0,0}^i} - \hat{P}_j^i\|_{L^2}, \quad (5.10)$$



where  $Q_{\mathcal{T}_{0,0}^i}$  is the  $L^2$  projection onto  $\mathcal{Q}(\mathcal{T}_{0,0}^i)$  and  $\hat{P}_{\mathcal{T}_{0,0}^i} \in \mathcal{Q}(\mathcal{T}_{0,0}^i)$  is the pressure in (5.5). Note that no refinement takes place in this loop and there is therefore no contribution in the total of marked cells from this step.

In the process indexed by  $i$ ,  $\mathcal{T}_{j,k}^i$  is refined to yield  $\mathcal{T}_{0,0}^{i+1}$  using  $\|Q_{\mathcal{T}_{0,0}^i}(\operatorname{div}(\hat{\mathbf{U}}_{j,k}^i) - \hat{g})\|_{L^2(T)}$  as local error estimator over the cell  $T$  and the adaptive strategy provided in [5]. This case accounts for the pressure error and implies

$$\|\hat{p} - \hat{P}_{\underline{j}(i+1)}^{i+1}\|_{L^2} \leq \rho_3 \|\hat{p} - \hat{P}_{\underline{j}(i)}^i\|_{L^2}, \quad (5.11)$$

for some  $\rho_3 < 1$  and where  $\underline{j}(i)$  is the maximum values reached by  $j$  for a given  $i$  in the algorithm. For this loop again, the number of marked cells can also be estimated

$$\#\mathcal{M}^i \lesssim \|\hat{p} - \hat{P}_{\underline{j}(i)}^i\|_{L^2}^{-1/s} \left( |\hat{p}|_{\mathcal{A}_{\mathcal{Q}}^s(L_0^2(\Omega))} + \|\hat{p}\|_{L^2} \right)^{1/s}, \quad (5.12)$$

provided  $\hat{p} \in \mathcal{A}_{\mathcal{Q}}^s(L_0^2(\Omega))$ .

Due to the saddle point structure of the problem, a critical feature is to decide when stopping each of the infinite loops discussed above. For instance, notice that the pressure error may increase when performing refinements in loop  $k$  to decrease the error on the velocity. This issue is circumvented by Stevenson and Kondratyuk [19] by considering stopping criteria based on the relative magnitude of different error estimators. In particular, this guarantees that the algorithm does not over approximate each quantities by ensuring the existence of a constant  $C$  independent of  $i, j, k$  and  $\varepsilon$  such that

$$\|\hat{\mathbf{u}}_{j,k}^{\hat{P}_j^i} - \hat{\mathbf{U}}_{j,k}^i\| \geq C\varepsilon, \quad \|\hat{P}_{\mathcal{T}_{0,0}^i} - \hat{P}_j^i\|_{L^2} \geq C\varepsilon \quad \text{and} \quad \|\hat{p} - \hat{P}_{\underline{j}(i)}^i\|_{L^2} \geq C\varepsilon. \quad (5.13)$$

Contemplating the proof presented in [19], it transpires that the arguments leading to (5.9) and (5.12) do not require the full regularity  $(\hat{\mathbf{u}}, \hat{p}) \in \mathcal{A}_{\mathbf{V}}^s(H_0^1(\Omega)^d) \times \mathcal{A}_{\mathcal{Q}}^s(L_0^2(\Omega))$  but only that  $(\hat{\mathbf{u}}, \hat{p})$  can be approximated with desired rate up to the tolerance  $\gamma\|\hat{\mathbf{u}}_{j,k}^{\hat{P}_j^i} - \hat{\mathbf{U}}_{j,k}^i\|_{H_0^1}$  and  $\gamma\|\hat{p} - \hat{P}_{\underline{j}(i)}^i\|_{L^2}$ , where  $\gamma$  depends on the constants of the lower and upper bounds of the various error estimators as well as the parameters used for the marking strategies. Therefore, it suffices to require that  $(\hat{\mathbf{u}}, \hat{p})$  is a  $(\gamma\varepsilon)$ -approximation of order  $s$  of  $(\mathbf{u}, p) \in \mathcal{A}_{\mathbf{V}}^s(H_0^1(\Omega)^d) \times \mathcal{A}_{\mathcal{Q}}^s(L_0^2(\Omega))$  to deduce

$$\#\mathcal{M}_{j,k}^i \lesssim \|\hat{\mathbf{u}}_{j,k}^{\hat{P}_j^i} - \hat{\mathbf{U}}_{j,k}^i\|_{H_0^1}^{-1/s} \left( (\|\mathbf{u}\|_{\mathcal{A}_{\mathbf{V}}^s(H_0^1(\Omega)^d)} + \|\mathbf{u}\|_{H_0^1})^{1/s} + (|p|_{\mathcal{A}_{\mathcal{Q}}^s(L_0^2(\Omega))} + \|p\|_{L^2})^{1/s} \right),$$

and

$$\#\mathcal{M}^i \lesssim \|\hat{p} - \hat{P}_{\underline{j}(i)}^i\|_{L^2}^{-1/s} \left( |p|_{\mathcal{A}_{\mathcal{Q}}^s(L_0^2(\Omega))} + \|p\|_{L^2} \right)^{1/s},$$

provided  $(\mathbf{u}, p) \in \mathcal{A}_{\mathbf{V}}^s(H_0^1(\Omega)^d) \times \mathcal{A}_{\mathcal{Q}}^s(L_0^2(\Omega))$ . Hence, invoking estimate (5.13), we obtain the desired bound on the total number of marked cells  $N(\hat{\mathbf{u}}, \hat{p})$  to achieve the desired tolerance  $\varepsilon$  starting from the subdivision  $\mathcal{T}$ :

$$N(\hat{\mathbf{u}}, \hat{p}) \lesssim \left( \|\mathbf{u}\|_{\mathcal{A}_{\mathbf{V}}^s(H_0^1(\Omega)^d)}^{1/s} + \|p\|_{\mathcal{A}_{\mathcal{Q}}^s(L_0^2(\Omega))}^{1/s} \right) \varepsilon^{-1/s}.$$

We emphasize that the number of sub-iterations inside each of the three loops does not appear in the above estimate due to the contraction properties (5.8), (5.10) and (5.11).

It follows that the algorithm introduced in this section fulfills the  **$(\gamma\varepsilon)$ -approximation optimality condition of order  $s$** .

**6. Numerical Results.** We start by introducing the STOKES subroutine used in our main algorithm and provide numerical evidences of its optimality for a selected values of numerical parameters. Then, we discuss the performances of STOKES\_DISC in an insightful context. All computations presented in what follows were performed using the `deal.II` library [1].

**6.1. Convergence Rate for the STOKES Procedure.** The STOKES routines used in our numerical implementation is described now. We consider Taylor-Hood spaces to build  $\mathbf{V}(\mathcal{T})$  and  $\mathcal{Q}(\mathcal{T})$ ; see [16] for details. First, we solve (3.2) using a Schur complement method. Then, we compute the residual a posteriori error estimators for the Stokes system [25] and adopt a Dörfler marking strategy [15]. The domain is partitioned using quadrilaterals with one hanging node so that Conditions 3, 4 and 6 in [10] are fulfilled.

We consider the test case introduced by Verfürth [25]. Let  $\Omega := [-1, 1] \times [-1, 1] \setminus [0, 1] \times [-1, 1] \subset \mathbb{R}^2$  be the L-shaped domain and denote by  $(\rho, \theta)$  the polar coordinates about the origin. Let  $\alpha \approx 0.544$  and  $\omega = \frac{3\pi}{2}$  be given constants. The exact expression of the velocity and pressure are given by

$$\mathbf{u}(\rho, \theta) := \rho^\alpha \begin{pmatrix} \cos(\theta)\psi'(\theta) + (1 + \alpha)\sin(\theta)\psi(\theta) \\ \sin(\theta)\psi'(\theta) - (1 + \alpha)\cos(\theta)\psi(\theta) \end{pmatrix}, \quad (6.1a)$$

$$p(\rho, \theta) := -\rho^{1-\alpha} \frac{(1 + \alpha)^2 \psi'(\theta) + \psi'''(\theta)}{1 - \alpha} - p_m, \quad (6.1b)$$

where  $p_m \approx 0.01006$  is such that  $\int_{\Omega} p = 0$ , and

$$\psi(\theta) := \frac{\sin((1 + \alpha)\theta) \cos(\alpha\omega)}{(1 + \alpha)} - \cos((1 + \alpha)\theta) + \frac{\sin((\alpha - 1)\theta) \cos(\alpha\omega)}{(1 - \alpha)} + \cos((\alpha - 1)\theta).$$

The right-hand side and viscosity functions are  $\mathbf{f} = \mathbf{0}$ ,  $g = 0$  and  $\mu = 1$ , respectively, but non-vanishing Dirichlet boundary conditions for the velocity compatible with (6.1a) are imposed on  $\partial\Omega$ .

The target tolerance is  $\varepsilon = 0.02$  and the Dörfler marking parameter is  $\theta = 0.3$ . In Figure 6.1, the error

$$\|\mathbf{u} - \mathbf{U}\|_{H_0^1} + \|p - P\|_{L^2}$$

is plotted in log-log scale against the number of cells used in the subdivision and compared with a line of slope  $-1$  corresponding to the optimal convergence rate. The adaptive algorithm is able to deliver the optimal error decay rate.

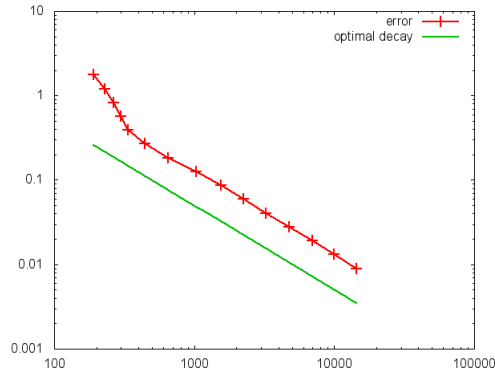


Fig. 6.1: Decay of the error in terms of degree of freedom invested in a singular case. The plot is based on 15 different meshes where the last one corresponds to a relative error of 0.07%. The optimal convergence rate 1 is achieved.

**6.2. A Singular Case for the STOKES\_DISC Procedure.** We illustrate the optimality of our algorithm STOKES\_DISC described in Section 3.2. We consider the same setting as in the previous example except that the data are chosen to be

$$\mu(\rho, \theta) := \begin{cases} 1 & \text{if } \rho \leq \rho_0, \\ \mu_2 & \text{otherwise,} \end{cases}$$

with  $\rho_0 = 0.5$ ,  $\mu_2 = 5$  and

$$\begin{aligned} \mathbf{f}_i(\rho, \theta) &:= \begin{cases} 0 & \text{if } \rho \leq \rho_0, \\ \frac{1}{\rho^2} (1 - \mu_2) \rho_0^\alpha \frac{\partial^2 \mathbf{v}_i}{\partial \theta^2}(\rho, \theta) & \text{otherwise,} \end{cases} \\ g(\rho, \theta) &:= \begin{cases} 0 & \text{if } \rho \leq \rho_0, \\ \frac{1}{\rho} \left(1 - \frac{1}{\mu_2}\right) \rho_0^\alpha \left(\cos(\theta) \frac{\partial \mathbf{v}_2}{\partial \theta}(\rho, \theta) - \sin(\theta) \frac{\partial \mathbf{v}_1}{\partial \theta}(\rho, \theta)\right) & \text{otherwise,} \end{cases} \end{aligned}$$

where the index  $i$  is used to denote the  $i$ th component of the vector valued function  $\mathbf{f}$ .

In this case, the exact pressure is still given by (6.1b) while the exact velocity reads

$$\mathbf{u}(\rho, \theta) := \begin{cases} \mathbf{v}(\rho, \theta) & \text{if } \rho \leq \rho_0, \\ \mathbf{v}(\rho_0, \theta) + \frac{1}{\mu_2} (\mathbf{v}(\rho, \theta) - \mathbf{v}(\rho_0, \theta)) & \text{otherwise,} \end{cases}$$

where  $\mathbf{v}(\rho, \theta)$  denotes the velocity given by (6.1a).

The parameters in **STOKES\_DISC** used for the computation are  $\varepsilon_0 = 2$ ,  $\xi = 0.7$  and  $\omega = 0.8$ . We now discuss the choice of  $r$  for which **Condition r** holds. In view of Remark 2, we decompose the domain  $\overline{\Omega} = \overline{\Omega}_1 \cup \overline{\Omega}_2$ , where  $\Omega_1 := \{(\rho, \theta) \in \Omega \mid \rho \leq \frac{\rho_0}{2}\}$ . In  $\Omega_1$ ,  $\mu$  is constant and so its approximation by a piecewise polynomial function is exact regardless the value of  $q_1$ . Hence the convergence rate of our method is only related to the value of  $q_2$  used to approximate  $\mu$  in  $\Omega_2$ . Since  $\mathbf{u} \in W^1(L^\infty(\Omega_2))^d$ , we have  $q_2 \in [2, \infty]$  in  $\Omega_2$ . Notice that the discontinuity of the viscosity takes place on a circular line and so can not be exactly captured by any subdivision of the initial mesh. Nevertheless, we have that  $\mathbf{u} \in \mathcal{A}^{1/2}(H_0^1(\Omega)^d)$ ,  $p \in \mathcal{A}^{1/2}(L_0^2(\Omega))$ ,  $\mathbf{f} \in \mathcal{B}^{1/2}(L^2(\Omega)^d) \subset \mathcal{B}^{1/2}(H^{-1}(\Omega)^d)$  and  $g \in \mathcal{B}^{1/2}(L_0^2(\Omega))$ . To approximate  $\mu$ , we use piecewise constant polynomials so that  $\mu \in \mathcal{B}^{1/q}(L^q(\Omega))$  for  $2 \leq q < \infty$ . Hence, choosing  $q = 2$ , we realize that Theorem 4.3 predicts an expected rate of convergence of  $s = 1/q$  with  $2 \leq q < \infty$ .

We test our procedure **STOKES\_DISC** for several values of  $q$ . The observed convergence rate

$$e_k := \frac{\log(\text{ERR}_k / \text{ERR}_{k-1})}{\log(\text{CELL}_k / \text{CELL}_{k-1})}$$

are computed at each iteration  $k$  of the algorithm, where  $\text{ERR}_k := \|\mathbf{u} - \mathbf{U}_k\|_{H_0^1} + \|p - P_k\|_{L^2}$  denotes the error after the  $k$ th iteration of the **STOKES\_DISC** algorithm and  $\text{CELL}_k$  represents the associated number of cells. The observed asymptotic values of  $e_k$  for different  $q$  are

$$-0.525 \text{ for } q = 2, \quad -0.344 \text{ for } q = 3, \quad -0.274 \text{ for } q = 4, \quad -0.193 \text{ for } q = 6.$$

We see that in all four cases, we obtain as expected  $e_k \approx -\frac{1}{q}$ . Notice that when  $q = \infty$  the algorithm does not converge since the error in the viscosity approximation saturate at 3.96, which is consistent with the jump in viscosity.

The meshes after several iterations of the **STOKES\_DISC** procedure with  $q = 2$  are presented in Figure 6.2. We see that during the first iterations, the refinement is mainly due to the lack of regularity of the solution. In a second step, the mesh is refined to account for the jump of the viscosity. Approximations of the pressure within **STOKES\_DISC** as well as the magnitude of the velocity field on the final subdivision are depicted in Figure 6.3.

## REFERENCES

- [1] W. Bangerth, R. Hartmann, and G. Kanschat. deal.II – a General Purpose Object Oriented Finite Element Library. *ACM Trans. Math. Softw.*, 33(4):24/1–24/27, 2007.
- [2] E. Bänsch, P. Morin, and R.H. Nochetto. An Adaptive Uzawa FEM for the Stokes Problem: Convergence Without the Inf-Sup Condition. *SIAM Journal on Numerical Analysis*, 40:1207–1229, 2002.
- [3] P. Binev, W. Dahmen, and R. DeVore. Adaptive Finite Element Methods with Convergence Rates. *Numerische Mathematik*, 97(2):219–268, 2004.
- [4] P. Binev, W. Dahmen, R.A. DeVore, and P. Petrushev. Approximation Classes for Adaptive Methods. *Serdica Mathematical Journal*, 28:391–416, 2002.
- [5] P. Binev and R.A. DeVore. Fast Computation in Adaptive Tree Approximation. *Numerische Mathematik*, 97:193–217, 2004.

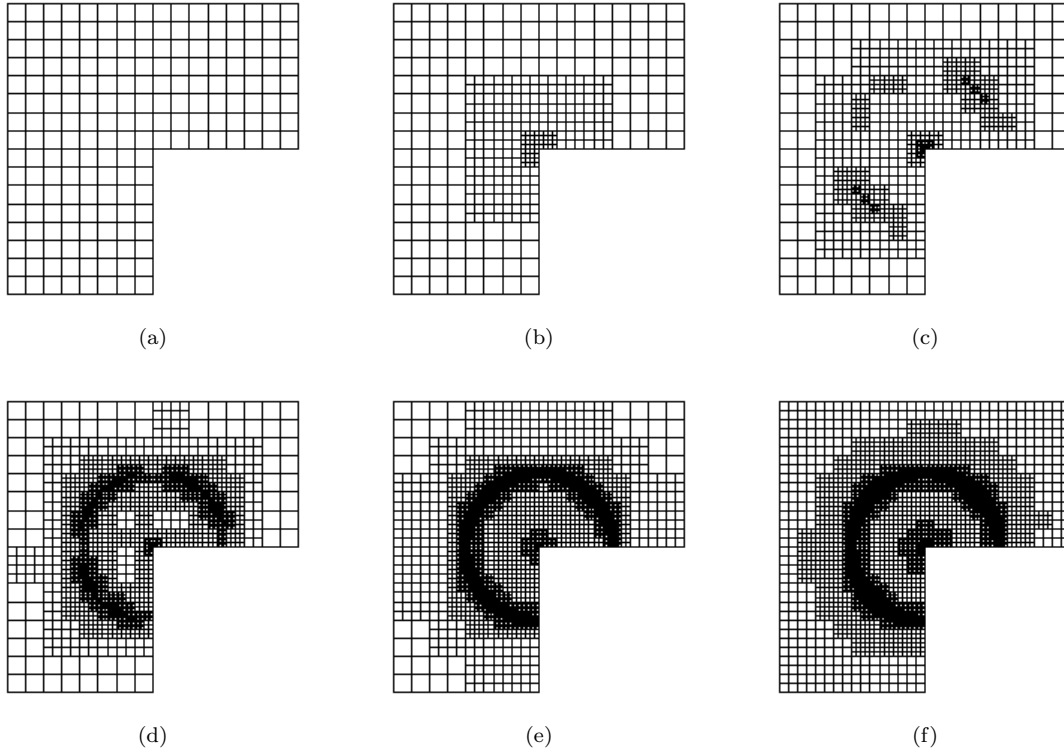


Fig. 6.2: Evolution of the subdivisions during the STOKES\_DISC algorithm with  $q = 2$ . The depicted grids corresponds to iterations (a)  $k = 0$ , (b)  $k = 2$ , (c)  $k = 4$ , (d)  $k = 6$ , (e)  $k = 8$  and (f)  $k = 10$ . At early stages, the refinements take place by the corner singularity and then by the location of the jump in viscosity.

- [6] M. Sh. Birman and M.Z. Solomyak. Piecewise-polynomial approximations of functions of the classes  $W_p^\alpha$ . *Mat. Sb. (N.S.)*, 73(115)(3):331–355, 1967.
- [7] A. Bonito, J.M. Cascón, P. Morin, and R.N. Nochetto. AFEM for geometric PDE: the Laplace-Beltrami operator. In *Analysis and Numerics of Partial Differential Equations*. Springer, 2013. to appear.
- [8] A. Bonito, R.A. DeVore, and R.H. Nochetto. Adaptive Finite Element Methods for Elliptic Problems with Discontinuous Coefficients. *arxiv preprint*, 2013.
- [9] A. Bonito, J.-L. Guermond, and F. Luddens. Regularity of the Maxwell Equations in Heterogeneous Media and Lipschitz Domain. submitted, 2013.
- [10] A. Bonito and R.H. Nochetto. Quasi-Optimal Convergence Rate of an Adaptive Discontinuous Galerkin Method. *SIAM Journal on Numerical Analysis*, 48(2):734–771, 2010.
- [11] A. Cohen, W. Dahmen, I. Daubechies, and R. DeVore. Tree approximation and optimal encoding. *Appl. Comput. Harmon. Anal.*, 11(2):192–226, 2001.
- [12] A. Cohen, R.A. DeVore, and R.H. Nochetto. Convergence Rates of AFEM with  $H^{-1}$  Data. *Foundations of Computational Mathematics*, 12:671–718, 2012.
- [13] S. Dahlke, W. Dahmen, and K. Urban. Adaptive Wavelet Methods for Saddle Point Problems—Optimal Convergence Rates. *SIAM Journal on Numerical Analysis*, 40(4):1230–1262, 2002.
- [14] R.A. DeVore. Nonlinear Approximation. In *Acta Numerica*, pages 51–150. Cambridge Univ Press, 1998.
- [15] W. Dörfler. A convergent adaptive algorithm for Poisson’s equation. *SIAM J. Numer. Anal.*, 33(3):1106–1124, 1996.
- [16] V. Girault and P.-A. Raviart. *Finite Element Methods for Navier-Stokes Equations*. Springer Series in Computational Mathematics, Springer-Verlag, 1986.
- [17] J. Hu and J. Xu. Convergence and Optimality of the Adaptive Nonconforming Linear Element Method for the Stokes Problem. *J. Sci. Comput.*, 55(1):125–148, 2013.
- [18] F. Jochmann. An  $H^s$ -regularity result for the gradient of solutions to elliptic equations with mixed boundary conditions. *J. Math. Anal. Appl.*, 238:429–450, 1999.
- [19] Y. Kondratyuk and R. Stevenson. An Optimal Adaptive Finite Element Method for the Stokes Problem. *SIAM Journal on Numerical Analysis*, 46(2):747–775, 2008.
- [20] V. Maz’ya and J. Rossmann. Lp Estimates of Solutions to Mixed Boundary Value Problems for the Stokes System in

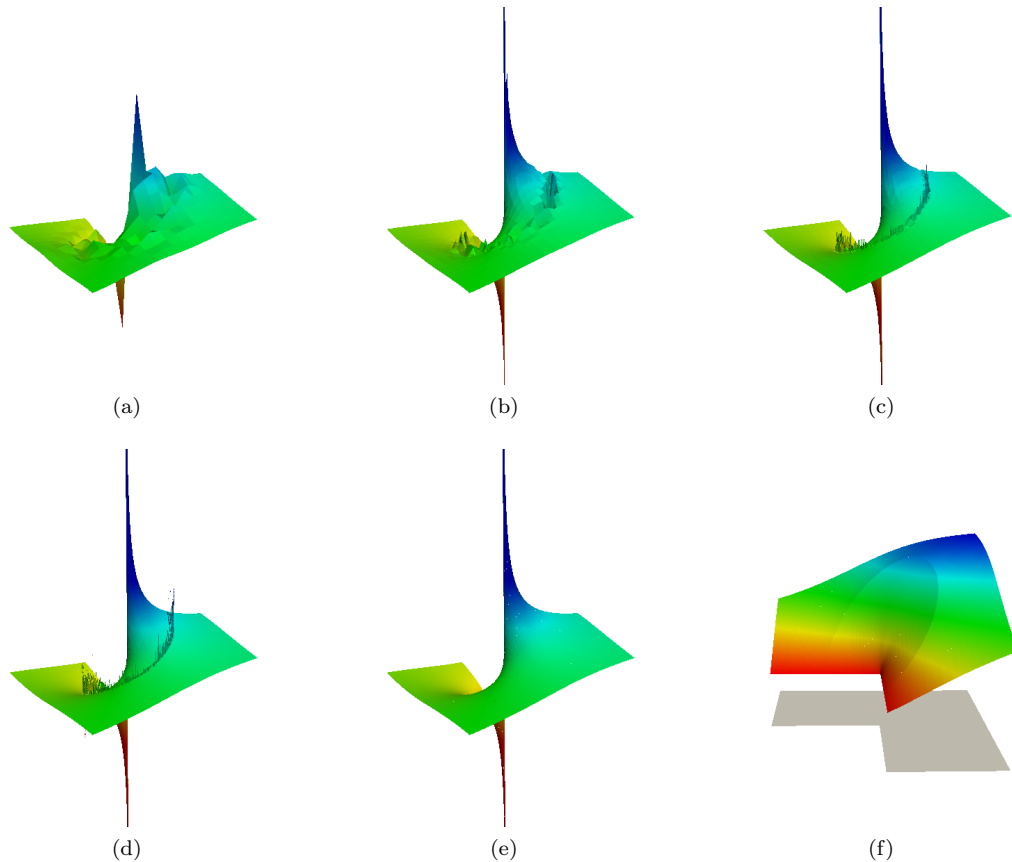


Fig. 6.3: Exact solution (e) and approximation of the pressure constructed by STOKES\_DISC after (a)  $k = 0$ , (b)  $k = 3$ , (c)  $k = 6$  and (d)  $k = 10$  iterations. The magnitude of the velocity field (f) is depicted on the final subdivision achieving a relative error of 2%.

- Polyhedral Domains. *Mathematische Nachrichten*, 280(7):751–793, 2007.
- [21] N.G. Meyers. An  $L^p$ -estimate for the gradient of solutions of second order elliptic divergence equations. *Ann. Scuola Norm. Sup. Pisa (3)*, 17:189–206, 1963.
  - [22] R. Stevenson. Optimality of a Standard Adaptive Finite Element Method. *Foundations of Computational Mathematics*, 7:245–269, 2007.
  - [23] R. Stevenson. The Completion of Locally Refined Simplicial Partitions Created by Bisection. *Mathematics of Computation*, 77(261):227–241, 2008.
  - [24] L. Tartar. *An Introduction to Sobolev Spaces and Interpolation Spaces*, volume 3. Springer, 2007.
  - [25] R. Verfürth. A Posteriori Error Estimators for the Stokes Equations. *Numerische Mathematik*, 55:309–325, 1989.