A MASS CONSERVING MIXED STRESS FORMULATION FOR STOKES FLOW WITH WEAKLY IMPOSED STRESS SYMMETRY*

JAY GOPALAKRISHNAN†, PHILIP L. LEDERER‡, AND JOACHIM SCHÖBERL‡

Abstract. We introduce a new discretization of a mixed formulation of the incompressible Stokes equations that includes symmetric viscous stresses. The method is built upon a mass conserving mixed formulation that we recently studied. The improvement in this work is a new method that directly approximates the viscous fluid stress σ, enforcing its symmetry weakly. The finite element space in which the stress is approximated consists of matrix-valued functions having continuous "normal-tangential" components across element interfaces. Stability is achieved by adding certain matrix bubbles that were introduced earlier in the literature on finite elements for linear elasticity. Like the earlier work, the new method here approximates the fluid velocity u using $H(\mathrm{div})$-conforming finite elements, thus providing exact mass conservation. Our error analysis shows optimal convergence rates for the pressure and the stress variables. An additional post processing yields an optimally convergent velocity satisfying exact mass conservation. The method is also pressure robust.

Key words. mixed finite elements; incompressible flows; Stokes equations; weak symmetry

AMS subject classifications. 35Q30, 65N12, 65N22, 65N30, 76D07, 76M10

1. Introduction.

In this work we introduce a new method for the discretization of steady incompressible Stokes system that includes symmetric viscous stresses. Let $\Omega \subset \mathbb{R}^d$ be a bounded domain with $d = 2$ or 3 having a Lipschitz boundary $\Gamma := \partial\Omega$. Let u and p be the velocity and the pressure, respectively. Given an external body force $f : \Omega \to \mathbb{R}^d$ and kinematic viscosity $\tilde{\nu} : \Omega \to \mathbb{R}$, the velocity-pressure formulation of the Stokes system is given by

$$\begin{cases} -\operatorname{div}(2\tilde{\nu}\varepsilon(u)) + \nabla p = f & \text{in } \Omega, \\ \operatorname{div}(u) = 0 & \text{in } \Omega, \\ u = 0 & \text{on } \Gamma, \end{cases} \tag{1.1}$$

where $\varepsilon(u) = (\nabla u + (\nabla u)^{\mathrm{T}})/2$. By introducing a new variable $\sigma = \nu\varepsilon(u)$ where $\nu := 2\tilde{\nu}$, equation (1.1) can be reformulated to

$$\frac{1}{\nu}\operatorname{dev}(\sigma) - \varepsilon(u) = 0 \qquad \text{in } \Omega, \tag{1.2a}$$

$$\operatorname{div}(\sigma) - \nabla p = -f \qquad \text{in } \Omega, \tag{1.2b}$$

$$\operatorname{div}(u) = 0 \qquad \text{in } \Omega, \tag{1.2c}$$

$$u = 0 \qquad \text{on } \Gamma. \tag{1.2d}$$

We shall call formulation (1.2) the **m***ass* **c***onserving mixed formulation with symmetric* **s***tresses*, or simply the MCS formulation. Although formulations (1.1) and (1.2) are formally equivalent, the MCS formulation (1.2) demands less regularity of the velocity field u. Many authors have studied this formulation previously [15, 14, 13, 12], including us [19]. In [19], following the others, we introduced a new variable $\sigma = \nu\nabla u$,

*Submitted to the editors DATE.

Philip L. Lederer has been funded by the Austrian Science Fund (FWF) through the research program "Taming complexity in partial differential systems" (F65) - project "Automated discretization in multiphysics" (P10).

†Portland State University, PO Box 751, Portland OR 97207,USA (gjay@pdx.edu)

‡Institute for Analysis and Scientific Computing, TU Wien, Austria (philip.lederer@tuwien.ac.at, joachim.schoeberl@tuwien.ac.at).

which is in general nonsymmetric, and considered an analogous formulation (which was also called an MCS formulation). The main novelty in [19] was that $\sigma = \nu\nabla u$ was set in a new function space $H(\operatorname{curl}\operatorname{div},\Omega)$ of matrix-valued functions whose divergence can continuously act on elements of $H_0(\operatorname{div},\Omega)$. Accordingly, the appropriate velocity space there was $H_0(\operatorname{div},\Omega)$, not $H_0^1(\Omega,\mathbb{R}^d)$ as in the classical velocity-pressure formulation.

In contrast to [19], in this work we set $\sigma = \nu\varepsilon(u)$, not $\nu\nabla u$. Our goal is to apply what we learnt in [19] to produce a new method that provides a direct approximation to the *symmetric* matrix function $\sigma = \nu\varepsilon(u)$. Being the viscous stress, this σ is practically more important than $\nu\nabla u$, especially when stress boundary conditions are involved [17]. We shall seek σ in the same function space $H(\operatorname{curl}\operatorname{div},\Omega)$ that we considered in [19]. We have shown in [19] that matrix-valued finite element functions with "normal-tangential" continuity across element interfaces are natural for approximating solutions in $H(\operatorname{curl}\operatorname{div},\Omega)$. We shall continue to use such finite elements here. It is interesting to note that in the HDG (hybrid discontinuous Galerkin) literature [11, 16] the potential importance of such normal-tangential continuity was noted and arrived at through a completely different approach.

The main point of departure in this work, stemming from that fact that the space $H(\operatorname{curl}\operatorname{div},\Omega)$ contains non-symmetric matrix-valued functions, is that we impose the symmetry of stress approximations weakly using Lagrange multipliers. This technique of imposing symmetry weakly is widely used in finite elements for linear elasticity [1, 2, 3, 14]. In particular, our analysis is inspired by the early work of Stenberg [31], who enriched the stress space by curls of local element bubbles. (In fact, this idea was even used in a Stokes mixed method [15], but their resulting method is not pressure robust.) These enrichment curls lie in the kernel of the divergence operator and are only "seen" by the weak-symmetry constraint allowing them to be used to prove discrete inf-sup stability. While in two dimensions – assuming a triangulation into simplices – this technique only increases the local polynomial order by 1, this is not the case in three dimensions. Years later [8, 18], it was realized that it is possible to retain the good convergence properties of Stenberg's construction and yet reduce the enrichment space. Introducing a "matrix bubble," these works added just enough extra curls needed to prove stability.

We shall see in later sections that the matrix bubble can also be used to enrich our discrete fluid stress space. This might seem astonishing at first. Indeed, an enrichment space for fluid stresses must map well when using a specific map that is natural to ensure normal-tangential continuity of the discrete stress space. Moreover, the enrichment functions must lie in the kernel of a realization of the distributional row-wise divergence used in MCS formulations (displayed in (3.4) below). It turns out that these properties are all fulfilled by an enrichment using a double curl involving matrix bubbles. Hence we are able to prove the discrete inf-sup condition. Stability then follows in the same type of norms used in [31] and is a key result of this work.

Some comments on the choice of the discrete velocity space and its implications are also in order here. As mentioned above, the velocity space within the MCS formulation is $V = H_0(\operatorname{div},\Omega)$. One of the main features of the first MCS method [19], as well the new version with weakly imposed symmetry of this paper, is that we can choose a discrete velocity space $V_h \subset V$ using $H(\operatorname{div})$-conforming finite elements. Therefore, our method is tailored to approximate the incompressibility constraint exactly, leading to pointwise and exactly divergence-free discrete velocity fields. The use of such $H(\operatorname{div})$-conforming velocities in Stokes flow is by no means new: for the standard velocity-pressure formulation, once can find it in [9, 10], and for the

Brinkman Problem in [21]. Therein, and also in the more recent works of [26, 25], the H^1-conformity is treated in a weak sense and a (hybrid) discontinuous Galerkin method is constructed. When employing $H(\mathrm{div})$-conforming finite elements, one has the luxury of choice. In [19], we used the $\mathcal{BDM}^{k+1}$ space [6] and added several local stress bubbles in order to guarantee stability. In contrast, in this paper, we have chosen to take the smaller Raviart-Thomas space [27] of order k, denoted by $\mathcal{RT}^k$. A similar choice was made also in the work of [16], where they presented a hybrid method for solving the Brinkman problem based off the work of [11]. Our current choice of the smaller space $\mathcal{RT}^k$ leads to a less accurate velocity approximation (compared to $\mathcal{BDM}^{k+1}$), so in order to recover the optimal convergence order of the velocity (measured in a discrete H^1-norm), we introduce a local element-wise post processing. Using the reconstruction operator of [22, 23] this post processing can be done retaining the exact divergence-free property.

The remainder of this paper is organized as follows. In Section 2, we define notation for common spaces used throughout this work and introduce an undiscretized formulation. Section 3 presents the MCS method for Stokes flow including symmetric viscous stresses. In Section 4, we present the new discrete method including the introduction of the matrix bubble. Section 5 proves a discrete inf-sup condition and develops a complete a priori error analysis of the discrete MCS system. In Section 6, we introduce a postprocessing for the discrete velocity. The concluding section (Section 7) reports various numerical experiments we performed to illustrate the theory.

2. Preliminaries. In this section, we introduce notation and present a weak formulation for Stokes flow that includes symmetric viscous stresses.

Let $\mathcal{D}(\Omega)$ or $\mathcal{D}(\Omega,\mathbb{R})$ denote the set of infinitely differentiable compactly supported real-valued functions on Ω and let $\mathcal{D}^*(\Omega)$ denote the space of distributions. To differentiate between scalar, vector and matrix-valued functions on Ω, we include the co-domain in the notation, e.g., $\mathcal{D}(\Omega,\mathbb{R}^d) = \{u : \Omega \to \mathbb{R}^d \,|\; u_i \in \mathcal{D}(\Omega)\}$. Let $\mathbb{M}$ denote the vector space of real $d \times d$ matrices. This notation scheme is similarly extended to other function spaces as needed. Thus, $L^2(\Omega) = L^2(\Omega,\mathbb{R})$ denotes the space of square integrable $\mathbb{R}$-valued functions on Ω, while analogous vector and matrix-valued function spaces are defined by $L^2(\Omega,\mathbb{R}^d) := \big\{u : \Omega \to \mathbb{R}^d \big|\; u_i \in L^2(\Omega)\big\}$ and $L^2(\Omega,\mathbb{M}) := \big\{\sigma : \Omega \to \mathbb{M} \big|\; \sigma_{ij} \in L^2(\Omega)\big\}$, respectively. Let $\mathbb{K}$ denote the vector space of $d \times d$ skew symmetric matrices, i.e., $\mathbb{K} = \mathrm{skw}(\mathbb{M})$, and let $L^2(\Omega,\mathbb{K}) := \big\{\sigma : \Omega \to \mathbb{K} \big|\; \sigma_{ij} \in L^2(\Omega)\big\}$.

Recall that the dimension d in this work is either 2 or 3. Accordingly, depending on the context, certain differential operators have different meanings. The "curl" operator, depending on the context, denotes one of the differential operators below.

$$\begin{aligned}
&\mathrm{curl}(\phi) = (-\partial_2\phi, \partial_1\phi)^{\mathrm{T}}, && \text{for } \phi \in \mathcal{D}^*(\Omega,\mathbb{R}), d = 2,\\
&\mathrm{curl}(\phi) = (\partial_2\phi_3 - \partial_3\phi_2, \partial_3\phi_1 - \partial_1\phi_3, \partial_1\phi_2 - \partial_2\phi_1)^{\mathrm{T}} && \text{for } \phi \in \mathcal{D}^*(\Omega,\mathbb{R}^3), d = 3,
\end{aligned}$$

where $(\cdot)^{\mathrm{T}}$ denotes the transpose and ∂_i abbreviates $\partial/\partial x_i$. For matrix-valued functions in both $d = 2$ and 3 cases, i.e., $\phi \in \mathcal{D}^*(\Omega,\mathbb{M})$, by $\mathrm{curl}(\phi)$ we mean the matrix obtained by taking curl row wise. Unfortunately, this still does not exhaust all the curl cases. In the $d = 2$ case, there are two possible definitions of $\mathrm{curl}(\phi)$ for $\phi \in \mathcal{D}^*(\Omega,\mathbb{R}^2)$,

$$\mathrm{curl}(\phi) = -\partial_2\phi_1 + \partial_1\phi_2, \qquad \text{or} \tag{2.1}$$

$$\mathrm{curl}(\phi) = \begin{pmatrix} \partial_2\phi_1 & -\partial_1\phi_1 \\ \partial_2\phi_2 & -\partial_1\phi_2 \end{pmatrix}, \tag{2.2}$$

and we shall have occasion to use both. The latter will not be used until (3.7) below, so until then, the reader may continue assuming we mean (2.1) whenever we consider curl of vector functions in $\mathbb{R}^2$. The operator ∇ is to be understood from context as an operator that results in either a vector whose components are $[\nabla\phi]_i = \partial_i\phi$ for $\phi \in \mathcal{D}^*(\Omega, \mathbb{R})$, a matrix whose entries are $[\nabla\phi]_{ij} = \partial_j\phi_i$ for $\phi \in \mathcal{D}^*(\Omega, \mathbb{R}^d)$, or a third-order tensor whose entries are $[\nabla\phi]_{ijk} = \partial_k\phi_{ij}$ for $\phi \in \mathcal{D}^*(\Omega, \mathbb{K})$. Finally, in a similar manner, we understand $\operatorname{div}(\phi)$ as either $\sum_{i=1}^d \partial_i\phi_i$ for vector-valued $\phi \in \mathcal{D}^*(\Omega, \mathbb{R}^d)$, or the row-wise divergence $\sum_{j=1}^d \partial_j\phi_{ij}$ for matrix-valued $\phi \in \mathcal{D}(\Omega, \mathbb{M})^*$.

Let $\tilde{d} = d(d-1)/2$ (so that $\tilde{d} = 1$ and 3 for $d = 2$ and 3, respectively). In addition to the standard Sobolev space $H^m(\Omega)$ for any $m \in \mathbb{R}$, we shall use the well-known space $H(\operatorname{div}, \Omega) = \{u \in L^2(\Omega, \mathbb{R}^d) : \operatorname{div}(u) \in L^2(\Omega)\}$. By its trace theorem, $H_0(\operatorname{div}, \Omega) = \{u \in H(\operatorname{div}, \Omega) : u \cdot n|_\Gamma = 0\}$ is a well-defined closed subspace, where n denotes the outward unit normal on Γ. Its dual space $[H_0(\operatorname{div}, \Omega)]^*$, as proved in [19, Theorem 2.1], satisfies

$$[H_0(\operatorname{div}, \Omega)]^* = H^{-1}(\operatorname{curl}, \Omega) = \{\phi \in H^{-1}(\Omega, \mathbb{R}^d) : \operatorname{curl}(\phi) \in H^{-1}(\Omega, \mathbb{R}^{\tilde{d}})\}. \tag{2.3}$$

In this work, the following space is important:

$$H(\operatorname{curl}\operatorname{div}, \Omega) := \{\sigma \in L^2(\Omega, \mathbb{M}) : \operatorname{div}(\sigma) \in [H_0(\operatorname{div}, \Omega)]^*\},$$

where the name results from (2.3): indeed a function $\sigma \in H(\operatorname{curl}\operatorname{div}, \Omega)$ fulfills $\operatorname{curl}\operatorname{div}(\sigma) \in H^{-1}(\Omega, \mathbb{R}^{\tilde{d}})$.

Next, let us derive a variational formulation of the system (1.2), which is based on the mixed stress formulation (MCS) introduced in chapter 3 in the work [19]. The method is based on a weaker regularity assumption of the velocity as compared to the standard velocity-pressure formulation (1.1). The velocity u and the pressure p now belong, respectively, to the spaces

$$V := H_0(\operatorname{div}, \Omega), \qquad Q := L^2_0(\Omega) := \{q \in L^2(\Omega) : \int_\Omega q \,\mathrm{d}x = 0\}.$$

Multiplying (1.2c) with a pressure test function $q \in Q$ and integrating over the domain Ω ends up in the familiar equation $(\operatorname{div}(u), q)_{L^2(\Omega)} = 0$, which we write as the last equation of the final Stokes system (2.5) written below. Here and throughout, the inner product of a space X is denoted by $(\cdot,\cdot)_X$. When X is the space of functions whose components are square integrable functions on Ω, we abbreviate $(\cdot,\cdot)_X$ to simply $(\cdot,\cdot)$, as done in (2.5) below. Similarly, while we generally denote the norm and seminorm on a Sobolev space X by $\|\cdot\|_X$ and $|\cdot|_X$, respectively, to simplify notation, we set $\|f\|_D^2 := (f, f)_D$, where $(f, g)_D$ denotes $L^2(D, \mathbb{V})$ inner product for any $\mathbb{V} \in \{\mathbb{R}, \mathbb{R}^d, \mathbb{K}, \mathbb{M}\}$ and any subset $D \subseteq \Omega$. Moreover, when $D = \Omega$, we omit the subscript and simply write $\|f\|$ for $\|f\|$.

To motivate the remaining equations of (2.5), let the deviatoric part of a matrix σ be defined by $\operatorname{dev}(\sigma) := \sigma - d^{-1}\operatorname{tr}(\sigma)\operatorname{Id}$, where Id denotes the identity matrix and $\operatorname{tr}(\sigma) := \sum_{i=1}^d \sigma_{ii}$ denotes the matrix trace. Since $\nu^{-1}\sigma = \varepsilon(u)$, due to the incompressibility constraint $\operatorname{div}(u) = 0$, we have the identity

$$\operatorname{dev}(\nu^{-1}\sigma) = \operatorname{dev}(\varepsilon(u)) = \varepsilon(u) - \frac{\nu}{d}\operatorname{tr}(\varepsilon(u))\operatorname{Id} = \varepsilon(u) - \frac{1}{d}\operatorname{div}(u)\operatorname{Id} = \varepsilon(u). \tag{2.4}$$

Since $\operatorname{tr}(\sigma) = 0$ and $\sigma = \sigma^{\mathrm{T}}$, we define the stress space as the following closed subspace of $H(\operatorname{curl}\operatorname{div}, \Omega)$:

$$\Sigma^{\mathrm{sym}} := \{\tau \in H(\operatorname{curl}\operatorname{div}, \Omega) : \operatorname{tr}(\tau) = 0,\ \tau = \tau^{\mathrm{T}}\}.$$

Testing equations (1.2a) with a test functions $\tau \in \Sigma^{\mathrm{sym}}$ and integrating over the domain, we have for the term including $\varepsilon(u)$ the identity

$$\begin{aligned}\int_\Omega \varepsilon(u) : \tau \,\mathrm{d}x &= \frac{1}{2}\int_\Omega \nabla u : \tau \,\mathrm{d}x + \frac{1}{2}\int_\Omega (\nabla u)^{\mathrm{T}} : \tau \,\mathrm{d}x \\ &= \frac{1}{2}\int_\Omega \nabla u : \tau \,\mathrm{d}x + \frac{1}{2}\int_\Omega \nabla u : \tau \,\mathrm{d}x = \int_\Omega \nabla u : \tau \,\mathrm{d}x\,.\end{aligned}$$

Using the knowledge that the velocity u should be in $H_0^1(\Omega)$, we obtain

$$(\nu^{-1}\operatorname{dev}(\sigma), \operatorname{dev}(\tau)) + \langle \operatorname{div}(\tau), u\rangle_{H_0(\operatorname{div},\Omega)} = 0,$$

which is the first equation in the system (2.5) below. Here and throughout, when working with elements f of the dual space X^* of a topological space X, we denote the action of f on an element $x \in X$ by $\langle f, x\rangle_X$, where we may omit the subscript X when its obvious from context. Finally we also test (1.2b) with $v \in V$ and integrate the pressure term by parts. This results in the remaining equation of (2.5).

Summarizing, the weak problem is to find $(\sigma, u, p) \in \Sigma^{\mathrm{sym}} \times V \times Q$ such that

$$\tag{2.5} \begin{cases} (\nu^{-1}\operatorname{dev}(\sigma), \operatorname{dev}(\tau)) + \langle \operatorname{div}(\tau), u\rangle_{H_0(\operatorname{div},\Omega)} = 0 & \text{for all } \tau \in \Sigma^{\mathrm{sym}}, \\ \langle \operatorname{div}(\sigma), v\rangle_{H_0(\operatorname{div},\Omega)} + (\operatorname{div}(v), p) = -(f, v) & \text{for all } v \in V, \\ (\operatorname{div}(u), q) = 0 & \text{for all } p \in Q. \end{cases}$$

In the ensuing section, we shall focus on a discrete analysis of a nonconforming scheme based on (2.5). Although wellposedness of (2.5) is an interesting question, we shall not comment further on it here since it is of no direct use in a nonconforming analysis.

3. The new method. In [19], we introduced an MCS method where σ was an approximation to (the generally non-symmetric) $\nu\nabla u$ instead of (the symmetric) $\nu\varepsilon(u)$ considered above. Since there was no symmetry requirement in [19], there we worked with the space $\Sigma := \{\tau \in H(\operatorname{curl}\operatorname{div}, \Omega) : \operatorname{tr}(\tau) = 0\}$ instead of Σ^{sym}. The finite element space for Σ designed there can be reutilized in the current symmetric case (with some modifications), once we reformulate the symmetry requirement as a constraint in a weak form.

To do so, we need further notation. Let $\kappa : \mathbb{R}^{\tilde{d}} \to \mathbb{K}$ be defined by

$$\tag{3.1} \kappa(v) = \frac{1}{2}\begin{pmatrix} 0 & -v \\ v & 0 \end{pmatrix} \text{ if } d = 2, \qquad \kappa(v) = \frac{1}{2}\begin{pmatrix} 0 & -v_3 & v_2 \\ v_3 & 0 & -v_1 \\ -v_2 & v_1 & 0 \end{pmatrix} \text{ if } d = 3.$$

When u represents the Stokes velocity, $\omega = \kappa(\operatorname{curl}(u))$ represents the *vorticity.* Since $\nabla u = \varepsilon(u) + \omega$, introducing ω as a new variable, and the symmetry condition $\sigma - \sigma^{\mathrm{T}} = 0$ as a new constraint, we obtain the boundary value problem

$$\tag{3.2a} \frac{1}{\nu}\operatorname{dev}(\sigma) - \nabla u + \omega = 0 \quad \text{in } \Omega,$$

$$\tag{3.2b} \operatorname{div}(\sigma) - \nabla p = -f \quad \text{in } \Omega,$$

$$\tag{3.2c} \sigma - \sigma^{\mathrm{T}} = 0 \quad \text{in } \Omega,$$

$$\tag{3.2d} \operatorname{div}(u) = 0 \quad \text{in } \Omega,$$

$$\tag{3.2e} u = 0 \quad \text{on } \Gamma.$$

In the remainder of this section, we introduce a discrete formulation approximating (3.2).

The method will be described on a subdivision (triangulation) $\mathcal{T}_h$ of Ω consisting of triangles in two dimensions and tetrahedra in three dimensions. For the analysis later, we shall assume that the $\mathcal{T}_h$ is quasiuniform. By h we denote the maximum of the diameters of all elements $T \in \mathcal{T}_h$. Quasiuniformity implies that $h \sim \operatorname{diam}(T)$ for all mesh elements T. Here and throughout, by $A \sim B$ we indicate that there exist two constants $c, C > 0$ independent of the mesh size h as well as the viscosity ν such $cA \le B \le cA$. Similarly, we use the notation $A \lesssim B$ if there exists a constant $C \neq C(h, \nu)$ such that $A \le CB$. All element interfaces and element boundaries on Γ are called facets and are collected into a set $\mathcal{F}_h$. This set is partitioned into facets on the boundary $\mathcal{F}_h^{\text{ext}}$ and interior facets $\mathcal{F}_h^{\text{int}}$. On each facet we denote by $[\![\cdot]\!]$ the standard jump operator. On a boundary facet the jump operator is just the identity. On all facets we denote by n a unit normal vector. When integrating over boundaries of d-dimensional domains, the orientation of n is assumed to be outward. On a facet with normal n adjacent to an mesh element T, the normal and tangential traces of a smooth function $\phi : T \to \mathbb{R}^d$ are defined by $\phi_n := \phi \cdot n$ and $\phi_t = \phi - \phi_n n$, respectively. Similarly, for a smooth $\psi : T \to \mathbb{M}$, the (scalar-valued) "normal-normal" and the (vector-valued) "normal-tangential" components are defined by $\psi_{nn} = \psi : (n \otimes n) = n^{\mathrm{T}} \psi n$ and $\psi_{nt} = \psi n - \psi_{nn} n$, respectively.

For any integers $m, k \ge 0$, the following "broken spaces" are viewed as consisting of functions on Ω without any continuity constraints across element interfaces:

$$H^m(\mathcal{T}_h) := \prod_{T \in \mathcal{T}_h} H^m(T), \qquad \mathbb{P}^k(\mathcal{T}_h) := \prod_{T \in \mathcal{T}_h} \mathbb{P}^k(T).$$

For $D \subset \Omega$ we use the notation $(\cdot,\cdot)_D$ for the inner product of $L^2(D)$ or its vector and tensor analogues such as $L^2(D, \mathbb{R}^d), L^2(D, \mathbb{M}), L^2(D, \mathbb{K})$. Also let $\|\cdot\|_D^2 = (\cdot,\cdot)_D$. Next for each element $T \in \mathcal{T}_h$ let $\mathbb{P}^k(T) \equiv \mathbb{P}^k(T, \mathbb{R})$ denote the set of polynomials of degree at most k on T. The vector and tensor analogues such as $\mathbb{P}^k(T, \mathbb{R}^d), \mathbb{P}^k(T, \mathbb{M}), \mathbb{P}^k(T, \mathbb{K})$ have their components in $\mathbb{P}^k(T)$. The broken spaces $\mathbb{P}^k(\mathcal{T}_h, \mathbb{R}^d), \mathbb{P}^k(\mathcal{T}_h, \mathbb{M})$, and $\mathbb{P}^k(\mathcal{T}_h, \mathbb{K})$ are defined similarly. We shall also use the conforming Raviart-Thomas space (see [4, 28]), $\mathcal{RT}^k := \{u_h \in H(\operatorname{div}, \Omega) : u_h|_T \in \mathbb{P}^k(T, \mathbb{R}^d) + x\mathbb{P}^k(T, \mathbb{R}) \text{ for all } T \in \mathcal{T}_h\}$.

3.1. Velocity, pressure, and vorticity spaces. For any $k \ge 1$, we use

$$V_h := V \cap \mathcal{RT}^k, \qquad Q_h := Q \cap \mathbb{P}^k(\mathcal{T}_h), \qquad W_h := \mathbb{P}^k(\mathcal{T}_h, \mathbb{K}),$$

for approximating the velocity, pressure, and vorticity, respectively.

Standard finite element mappings apply for these spaces. Let $\hat{T}$ be the unit simplex (for $d = 2$ and 3), which we shall refer to as the *reference element*, and let $T \in \mathcal{T}_h$. Let $\phi : \hat{T} \to T$ be an affine homeomorphism and set $F := \phi'$. By quasiuniformity, $\|F\|_{\ell^\infty} \sim h$, $\|F^{-1}\|_{\ell^\infty} \sim h^{-1}$, and $|\det(F)| \sim h^d$, estimates that we shall use tacitly in our scaling arguments later. Such arguments proceed by mapping functions on $\hat{T}$ to and from $\hat{T}$. Given a scalar-valued $\hat{q}_h$, a vector-valued $\hat{v}_h$, and a skew-symmetric matrix-valued $\hat{\eta}_h$ on the reference element $\hat{T}$, we map them to T by

$$\mathcal{Q}(q_h) = \hat{q}_h \circ \phi^{-1}, \mathcal{P}(\hat{v}_h) := \det(F)^{-1} F(\hat{v}_h \circ \phi^{-1}), \mathcal{W}(\hat{\eta}_h) := F^{-\mathrm{T}}(\hat{\eta}_h \circ \phi^{-1}) F^{-1}, \tag{3.3}$$

respectively, i.e., these are our mappings for functions in the pressure, velocity, and vorticity spaces, respectively. The first is the inverse of the standard pullback, the second is the standard Piola map, and the third is designed to preserve skew symmetry.

3.2. Stress space. The definition of our stress space and the equations of our method are motivated by the following result proved in [19, Section 4].

THEOREM 3.1. *Suppose τ is in $H^1(\mathcal{T}_h, \mathbb{M})$ and $v \in H^1(\mathcal{T}_h, \mathbb{R}^d)$.*
1. *If the normal-tangential trace τ_{nt} is continuous across element interfaces, and $\tau_{nn}|_{\partial T} \in H^{1/2}(\partial T)$ for all $T \in \mathcal{T}_h$, then τ is in $H(\operatorname{curl}\operatorname{div}, \Omega)$.*
2. *If $v \in H_0(\operatorname{div}, \Omega)$, then*

$$\langle \operatorname{div}(\tau), v\rangle_{H_0(\operatorname{div},\Omega)} = \sum_{T\in\mathcal{T}_h} \left[\int_T \operatorname{div}(\tau)\, v\, dx - \int_{\partial T} \tau_{nn} v_n \, ds \right]. \tag{3.4}$$

Clearly, matrix finite element subspaces having normal-tangential continuity are suggested by Theorem 3.1. Technically, the theorem's sufficient conditions for full conformity also include the condition $\sigma_{nn}|_{\partial T} \in H^{1/2}(\partial T)$. This condition is very restrictive as it would enforce continuity at vertices and edges in two and three dimensions respectively. If this constraint is relaxed, much simpler, albeit nonconforming, elements can be constructed. This was the approach we adopted in [19]. We continue in the same vein here and define the nonconforming stress space

$$\Sigma_h := \{\tau_h \in \mathbb{P}^k(\mathcal{T}_h, \mathbb{M}) : \operatorname{tr}(\tau_h) = 0, \; [\![(\tau_h)_{nt}]\!] = 0 \text{ for all } F \in \mathcal{F}_h^{\text{int}}\}. \tag{3.5}$$

As mentioned in the introduction, we must enrich the above stress space Σ_h to guarantee solvability of the resulting discrete system due to the additional weak symmetry constraints. We follow the approach of [31] and its later improvements [8, 18] to construct the needed enrichment space.

Define a cubic matrix-valued "bubble" function as follows. On a d-simplex T with vertices $a_0, \ldots, a_d$, let F_i denote the face opposite to a_i, and let λ_i denote the unique linear function that vanishes on F_i and equals one on a_i, i.e., the ith barycentric coordinate of T. Following [8, 18], we define $B \in \mathbb{P}^3(T, \mathbb{M})$ by

$$B = \sum_{i=0}^{3} \lambda_{i-3}\lambda_{i-2}\lambda_{i-1} \, \nabla\lambda_i \otimes \nabla\lambda_i \qquad \text{if } d = 3, \tag{3.6a}$$

$$B = \lambda_0\lambda_1\lambda_2 \qquad \text{if } d = 2, \tag{3.6b}$$

where the indices on the barycentric coordinates are calculated mod 4 in (3.6a). Let $\mathbb{P}_\perp^k(T, \mathbb{V})$ denote the L^2-orthogonal complement of $\mathbb{P}^{k-1}(T, \mathbb{V})$ in $\mathbb{P}^k(T, \mathbb{V})$ for $\mathbb{V} \in \{\mathbb{R}, \mathbb{K}\}$, and let $\mathbb{P}_\perp^k(\mathcal{T}_h, \mathbb{V}) = \prod_{T\in\mathcal{T}_h} \mathbb{P}_\perp^k(T, \mathbb{V})$. For any $k \geq 1$, define

$$\delta\Sigma_h := \left\{ \operatorname{dev}(\operatorname{curl}(\operatorname{curl}(r_h B))) : \; r_h \in \mathbb{P}_\perp^k(\mathcal{T}_h, \mathbb{K}) \right\}, \tag{3.7}$$

for $d = 2$ and 3, with the understanding that in $d = 2$ case, the outer curl is defined by (2.2), not (2.1). The total stress space is given by

$$\Sigma_h^+ := \Sigma_h \oplus \delta\Sigma_h, \qquad k \geq 1.$$

That functions in this space have normal-tangential continuity is a consequence of the following property proved in [8, Lemma 2.3].

LEMMA 3.2. *Let $q \in \mathbb{M}$ and $T \in \mathcal{T}_h$. The products qB and Bq have vanishing tangential trace on ∂T, so the function* $\operatorname{curl}(qB)$ *has vanishing normal trace on ∂T.*

LEMMA 3.3. *Any $\sigma \in \delta\Sigma_h$ has vanishing σ_{nt} and $[\![\sigma_{nt}]\!]$ on all facets $F \in \mathcal{F}_h$.*

Proof. Since $(\operatorname{dev}(\sigma))_{nt} = \sigma_{nt}$, this is a direct consequence of Lemma 3.2. ∎

We also need a proper mapping for functions in Σ_h^+ that preserves normal-tangential continuity. We continue to use the following map, first introduced in [19]:

$$\mathcal{M}(\hat\sigma_h) := \frac{1}{\det(F)} F^{-\mathrm{T}}(\hat\sigma_h \circ \phi^{-1}) F^{\mathrm{T}}. \tag{3.8}$$

As shown in [19, Lemma 5.3], on each facet, $(\mathcal{M}(\hat\sigma_h))_{nt}$ is a scalar multiple of $(\hat\sigma_h)_{nt}$ and $\operatorname{tr}(\hat\sigma_h) = 0$ if and only if $\operatorname{tr}(\mathcal{M}(\hat\sigma_h)) = 0$. Degrees of freedom are discussed in §3.4.

Remark 3.4. Note that in (3.6), B was given using barycentric coordinates as an expression that holds on any simplex. Let $\hat B$ be defined by the same expression on the reference element $\hat T$ after replacing λ_i by reference element barycentric coordinates $\hat\lambda_i$. Considering the obvious map that transforms $\hat\nabla\hat\lambda_i \otimes \hat\nabla\hat\lambda_i$ to $\nabla\lambda_i \otimes \nabla\lambda_i$, we find that the matrix bubble B on any simplex is given by

$$B := F^{-\mathrm{T}}(\hat B \circ \phi^{-1}) F^{-1}. \tag{3.9}$$

3.3. Equations of the method. For the derivation of the discrete variational formulation we turn our attention back to the weak formulation (2.5) and identify these forms:

$$\begin{aligned} &a : L^2(\Omega,\mathbb{M}) \times L^2(\Omega,\mathbb{M}) \to \mathbb{R}, && b_1 : V \times Q \to \mathbb{R},\\ &a(\sigma,\tau) := (\nu^{-1}\operatorname{dev}(\sigma), \operatorname{dev}(\tau)), && b_1(u,p) := (\operatorname{div}(u), p). \end{aligned}$$

The definition of the remaining bilinear form is motivated by the definition of the "distributional divergence" given by (3.4). To this end we define $b_2 : \{\tau \in H^1(\mathcal{T}_h,\mathbb{M}) : [\![\tau_{nt}]\!] = 0\} \times \big(\{v \in H^1(\mathcal{T}_h,\mathbb{R}^d) : [\![v_n]\!] = 0\} \times L^2(\Omega,\mathbb{M})\big) \to \mathbb{R}$ by

$$b_2(\tau,(v,\eta)) := \sum_{T\in\mathcal{T}_h} \int_T \operatorname{div}(\tau)\cdot v \,\mathrm{d}x + \sum_{T\in\mathcal{T}_h} \int_T \tau : \eta \,\mathrm{d}x - \sum_{F\in\mathcal{F}_h} \int_F [\![\tau_{nn}]\!] v_n \,\mathrm{d}s\,. \tag{3.10}$$

Integrating the first integral by parts, we find the equivalent representation

$$b_2(\tau,(v,\eta)) = -\sum_{T\in\mathcal{T}_h} \int_T \tau : (\nabla v - \eta)\,\mathrm{d}x + \sum_{F\in\mathcal{F}_h} \int_F \tau_{nt}\cdot [\![v_t]\!] \,\mathrm{d}s\,. \tag{3.11}$$

Using these forms, we state the method. For any $k \ge 1$, the *discrete MCS method with weakly imposed symmetry* finds $\sigma_h, u_h, \omega_h, p_h \in \Sigma_h^+ \times V_h \times W_h \times Q_h$ such that

$$\begin{cases} a(\sigma_h,\tau_h) + b_2(\tau_h,(u_h,\omega_h)) = 0 & \text{for all } \tau_h \in \Sigma_h^+,\\ b_2(\sigma_h,(v_h,\eta_h)) + b_1(v_h,p_h) = (-f, v_h) & \text{for all } (v_h,\eta_h) \in U_h := V_h \times W_h,\\ b_1(u_h,q_h) = 0 & \text{for all } q_h \in Q_h. \end{cases} \tag{3.12}$$

Since V_h and Q_h fulfills $\operatorname{div}(V_h) = Q_h$, the discrete velocity solution component u_h satisfies $\operatorname{div}(u_h) = 0$ point wise, providing exact mass conservation.

3.4. Degrees of freedom of the new stress space. We need degrees of freedom (d.o.f.s) for the stress space that are well-suited for imposing normal-tangential continuity across element interfaces. Since the bubbles in $\delta\Sigma_h$ have zero normal-tangential continuity, we ignore them for this discussion and focus on d.o.f.s that control Σ_h.

Consider $\Sigma_T = \{\tau|_T : \tau \in \Sigma_h\}$ on any mesh element T. Letting $\mathbb{D}$ denote the subspace of matrices $M \in \mathbb{M}$ satisfying $M : \mathrm{Id} = 0$, we may identify Σ_T with $\mathbb{P}^k(T,\mathbb{D})$.

Let us recall a basis for $\mathbb{D}$ that was given in [19]. Define the following two sets of constant matrix functions, for $d = 2$ and $d = 3$ cases, respectively, by

$$S^i := \operatorname{dev}\big(\nabla\lambda_{i+1} \otimes \operatorname{curl}(\lambda_{i+2})\big), \tag{3.13a}$$

$$S_0^i := \operatorname{dev}\big(\nabla\lambda_{i+1} \otimes (\nabla\lambda_{i+2} \times \nabla\lambda_{i+3})\big), \quad S_1^i := \operatorname{dev}\big(\nabla\lambda_{i+2} \otimes (\nabla\lambda_{i+3} \times \nabla\lambda_{i+1})\big), \tag{3.13b}$$

taking the indices mod 3 and mod 4, respectively. We proved in [19, Lemma 5.1] that the sets $\{S^i : i = 0, 1, 2\}$ and $\{S_q^i : i = 0, 1, 2, 3,\ q = 0, 1\}$ form a basis of $\mathbb{D}$ when $d = 2$ and 3, respectively.

Our d.o.fs for $\Sigma_T \equiv \mathbb{P}^k(T, \mathbb{D})$ are grouped into two. The first group is associated to the set of element facets ($d-1$ subsimplices of T), namely, for each facet $F \in \partial T$, we define the set of d.o.f.s

$$\Phi^F(\tau) := \int_F \tau_{nt} \cdot r \,\mathrm{d}s$$

for each r in any fixed basis for $\mathbb{P}^k(F, \mathbb{R}^{d-1})$. The next group is the set of interior d.o.f.s, defined by

$$\Phi^0(\tau) := \int_T \tau : \varsigma \,\mathrm{d}x$$

for all ς in any basis of $\mathbb{P}^{k-1}(T, \mathbb{D})$. We proceed to prove that the set of these d.o.f.s, $\Phi(T) := \Phi^0(\tau) \cup \{\Phi^F : F \subset \partial T\}$, is unisolvent.

THEOREM 3.5. *The set $\Phi(T)$ is a set of unisolvent d.o.f.s for $\Sigma_T \equiv \mathbb{P}^k(T, \mathbb{D})$.*

Proof. Suppose $\tau \in \Sigma_T$ satisfies $\phi(\tau) = 0$ for all d.o.f.s $\phi \in \Phi(T)$. We need to show that $\tau = 0$. From the facet d.o.f.s we conclude that τ_{nt} vanishes on ∂T. By [19, Lemma 5.2], τ may be expressed as

$$\tau = \sum_{i=0}^{2} \mu_i \lambda_i S^i \quad \text{or} \quad \tau = \sum_{q=0}^{1}\sum_{i=0}^{3} \mu_i^q \lambda_i S_q^i, \tag{3.14}$$

when $d = 2$ or 3, respectively, where $\mu_i, \mu_i^0, \mu_i^1 \in \mathbb{P}^{k-1}(T)$. The interior d.o.f.s imply that $\int_T \tau : s \,\mathrm{d}x = 0$ for any $s \in \mathbb{P}^{k-1}(\hat{T}, \mathbb{D})$. Choosing for s the expression on the right hand side in (3.14) omitting the λ_i, say for the $d = 2$ case, we obtain

$$\int_T \sum_{i=0}^{2} \mu_i \lambda_i S^i : \sum_{i=0}^{2} \mu_i S^i \,\mathrm{d}x = \int_T \lambda_i \Big| \sum_{i=0}^{2} \mu_i S^i \Big|^2 \mathrm{d}x = 0,$$

yielding $\mu_i = 0$, and thus $\tau = 0$. A similar argument in $d = 3$ case yields the same conclusion that $\tau = 0$.

To complete the proof, it now suffices to prove that $\dim(\Sigma_T)$ equals the number of d.o.f.s, i.e., $\#\Phi(T)$. Obviously, $\dim(\Sigma_T) = \dim \mathbb{P}^k(T, \mathbb{D}) = (d^2 - 1) \dim \mathbb{P}^k(T)$. The cardinality of $\Phi(T)$ equals the sum of the number of facet d.o.f.s $(d+1)(d-1) \dim \mathbb{P}^k(T)$ and the number of interior d.o.f.s $(d^2 - 1) \dim \mathbb{P}^{k-1}(T)$, which simplifies to $(d^2 - 1)\big(\dim \mathbb{P}^{k-1}(T) + \dim \mathbb{P}^k(F)\big)$, equalling $\dim(\Sigma_T)$. $\square$

Using these d.o.f.s, a canonical local interpolant $I_T(\tau)$ in Σ_T can be defined as usual, by requiring that $\psi(\tau - I_T\tau) = 0$, for all $\psi \in \Phi(T)$.

LEMMA 3.6. *For any $\tau \in H^1(T,\mathbb{D})$, we have $\mathcal{M}^{-1}(I_T\tau) = I_{\hat{T}}(\mathcal{M}^{-1}(\tau))$.*

Proof. This proceeds along the same lines as the proof of [19, Lemma 5.4]. ∎

The global interpolant I_{Σ_h} is also defined as usual. On each element $T \in \mathcal{T}_h$ the global interpolant $(I_{\Sigma_h}\tau)|_T$ coincides with the local interpolant $I_T(\tau|_T)$.

THEOREM 3.7. *For any $m \geq 1$ and any $\sigma \in \{\tau \in H^m(\mathcal{T}_h,\mathbb{D}) : [\![\tau_{nt}]\!] = 0\}$, the global interpolation operator I_{Σ_h} satisfies for all $s \leq \min(k+1,m)$*

$$\|\sigma - I_{\Sigma_h}\sigma\|^2 + \sum_{F\in\mathcal{F}_h} h\|(\sigma - I_{\Sigma_h}\sigma)_{nt}\|_F^2 \lesssim h^{2s}\|\sigma\|^2_{H^s(\mathcal{T}_h)}.$$

Proof. This follows from a standard Bramble-Hilbert argument using Lemma 3.6. ∎

4. A priori error analysis. In this section we first show the stability of the MCS method with weakly imposed symmetry by proving a discrete inf-sup condition (Theorem 4.14). We then prove consistency (Theorem 4.18), optimal error estimates (Theorem 4.19), and pressure robustness (Theorem 4.21). For simplicity, the analysis from now on assumes that ν is a constant.

4.1. Norms. In addition to the previous notation for norms (established in Section 2), hereon we also use $\|\cdot\|_h^2$ to abbreviate $\sum_{T\in\mathcal{T}_h}\|\cdot\|_T^2$, a notation that also serves to indicate that certain seminorms are defined using differential operators applied element by element, not globally, e.g.,

$$\|\varepsilon(v)\|_h^2 := \sum_{T\in\mathcal{T}_h}\|\varepsilon(v)\|_T^2, \qquad \|\operatorname{curl}(\gamma)\|_h^2 := \sum_{T\in\mathcal{T}_h}\|\operatorname{curl}(\gamma)\|_T^2,$$

$$\|v\|_{1,h,\varepsilon}^2 := \|\varepsilon(v)\|_h^2 + \sum_{F\in\mathcal{F}_h}\frac{1}{h}\big\|[\![v_t]\!]\big\|_F^2,$$

for $v \in H^1(\mathcal{T}_h,\mathbb{R}^d)$ and $\gamma \in H^1(\mathcal{T}_h,\mathbb{M})$. Recall that $U_h = V_h \times W_h$. Our analysis is based on norms of the type used in [31]. Accordingly, we will need to use the following norms for $v_h \in V_h$ and $\eta_h \in W_h$:

$$\|v_h\|_{V_h}^2 = \|v_h\|_{1,h,\varepsilon}^2, \qquad \|(v_h,\eta_h)\|_{U_h}^2 := \|v_h\|_{1,h,\varepsilon}^2 + \|\kappa(\operatorname{curl} v_h) - \eta_h\|_h^2.$$

Lemma 4.8 below will show that the latter is indeed a norm.

On the discrete space U_h, we will also need another norm defined using the following projections. On any mesh element T, let Π_T^{k-1} denote the $L^2(T,\mathbb{V})$ orthogonal projection onto $\mathbb{P}^k(T,\mathbb{V})$ where $\mathbb{V}$ is determined from context to be an appropriate vector space such as $\mathbb{R}^d$, or $\mathbb{M}$. When the element T is clear from context, we shall drop the subscript T in Π_T^{k-1} and simply write Π^{k-1}. Also, on each facet $F \in \mathcal{F}_h$, we introduce a projection onto the tangent plane $n_F^\perp$: for any $v \in L^2(F, n_F^\perp)$, the projection $\Pi_F^1 v \in \mathbb{P}^1(F, n_F^\perp)$ is defined by $(\Pi_F^1 v, r)_F = (v,r)_F$ for all $r \in \mathbb{P}^1(F, n_F^\perp)$. Using these, define

$$\|(v_h,\eta_h)\|_{U_h,*}^2 := \sum_{T\in\mathcal{T}_h}\|\Pi_T^{k-1}\operatorname{dev}(\nabla v_h - \eta_h)\|_T^2 + \sum_{F\in\mathcal{F}_h}\frac{1}{h}\|\Pi_F^1[\![(v_h)_t]\!]\|_F^2. \tag{4.1}$$

Lemma 4.7 below will help us go between this norm and $\|(v_h,\eta_h)\|_{U_h}$.

The remaining spaces Σ_h^+ and Q_h are simply normed by the L^2 norm $\|\cdot\|$. The full discrete space is normed by

$$\|(v_h,\eta_h,\tau_h,q_h)\|_* := \sqrt{\nu}\,||(v_h,\eta_h)||_{U_h} + \frac{1}{\sqrt{\nu}}(\|\tau_h\| + ||q_h||) \tag{4.2}$$

for any $(v_h,\eta_h,\tau_h,q_h) \in V_h \times W_h \times \Sigma_h^+ \times Q_h$.

4.2. Norm equivalences. Next, we use the finite element mappings introduced earlier –see (3.3) and (3.8)– to show several norm equivalences.

LEMMA 4.1. *Let $\tau_h \in \Sigma_h^+$. Then*

$$h^d \|\tau_h\|_T^2 \sim \|\hat{\tau}_h\|_{\hat{T}}^2 \quad \text{for all} \quad T \in \mathcal{T}_h \tag{4.3}$$

$$h^{d+1} \|(\tau_h)_{nt}\|_F^2 \sim \|(\hat{\tau}_h)_{\hat{n}\hat{t}}\|_{\hat{F}}^2 \quad \text{for all} \quad F \in \mathcal{F}_h. \tag{4.4}$$

$$\|\tau_h\|^2 \sim \sum_{T \in \mathcal{T}_h} \|\tau_h\|_T^2 + \sum_{F \in \mathcal{F}_h} h \big\|(\tau_h)_{nt}\big\|_F^2. \tag{4.5}$$

Proof. The first two follow by a simple scaling argument. For the third, see the proof of [19, Lemma 6.1]. ∎

In the proof of the next lemma, we use the space of rigid displacements $\mathbb{E} = \mathbb{P}^0(T, \mathbb{R}^d) + \mathbb{P}^0(T, \mathbb{K})\, x$. For each element $T \in \mathcal{T}_h$, let $\Pi^{\mathbb{E}} : H^1(T) \to \mathbb{E}$ denote the projector defined in [5]. Then, for any $v_h \in V_h$, the projection $\Pi^{\mathbb{E}} v_h \in \mathbb{E}$ fulfills the properties (see [5, eq. (3.3), (3.11)])

$$\|\nabla(v_h - \Pi^{\mathbb{E}} v_h)\|_T \sim \|\varepsilon(v_h)\|_T \qquad \text{for all } T \in \mathcal{T}_h, \tag{4.6}$$

$$\big\|[\![v_h - \Pi^{\mathbb{E}} v_h]\!]\big\|_F^2 \lesssim \sum_{T : T \cap F \neq \emptyset} h \|\varepsilon(v_h)\|_T^2 \qquad \text{for all } F \in \mathcal{F}_h. \tag{4.7}$$

We shall also use a global discrete Korn inequality, implied by [5, Theorem 3.1]. Namely, there is an h-independent constant c_K such that

$$c_K^2 \|\nabla v\|_h^2 \le \|\varepsilon(v)\|_h^2 + \sum_{F \in \mathcal{F}_h} h^{-1} \big\|\Pi_F^1 [\![v]\!]\big\|_F^2, \qquad \text{for all } v \in H^1(\mathcal{T}_h, \mathbb{R}^d). \tag{4.8}$$

LEMMA 4.2. *For all $(v_h, \eta_h) \in U_h$,*

$$\|(v_h, \eta_h)\|_{U_h}^2 \sim \|\varepsilon(v_h)\|_h^2 + \|\kappa(\operatorname{curl} v_h) - \eta_h\|_h^2 + \sum_{F \in \mathcal{F}_h} \frac{1}{h} \big\|\Pi_F^1 [\![(v_h)_t]\!]\big\|_F^2.$$

Proof. One side of the equivalence is obvious by the continuity of the Π_F^1. For the other direction first note that

$$h^{-1} \|[\![(v_h)_t]\!]\|_F^2 \le 2h^{-1} \|\Pi_F^1 [\![(v_h)_t]\!]\|_F^2 + 2h^{-1} \|[\![(v_h - \Pi_F^1 v_h)_t]\!]\|_F^2.$$

As $\Pi^{\mathbb{E}} v_h \in \mathbb{P}^1(T, \mathbb{R}^d)$ we have again by the continuity of Π_F^1,

$$\|[\![(v_h - \Pi_F^1 v_h)_t]\!]\|_F^2 = \|(\operatorname{Id} - \Pi_F^1)[\![(v_h - \Pi^{\mathbb{E}} v_h)_t]\!]\|_F^2 \le \|[\![(v_h - \Pi^{\mathbb{E}} v_h)_t]\!]\|_F^2.$$

We conclude the proof using (4.7). ∎

The following well-known property of Raviart-Thomas spaces (see, e.g., [7, Lemma 3.1]) is needed at several points.

LEMMA 4.3. *Let $v \in \mathbb{P}^k(T, \mathbb{R}^d) + x\mathbb{P}^k(T, \mathbb{R})$ and $\operatorname{div}(v) = 0$. Then $v \in \mathbb{P}^k(T, \mathbb{R}^d)$.*

LEMMA 4.4. *For all $T \in \mathcal{T}_h$ and $v \in \mathbb{P}^k(T, \mathbb{R}^d) + x\mathbb{P}^k(T, \mathbb{R})$,*

$$\|\varepsilon(v)\|_T^2 \sim \|\Pi^{k-1} \operatorname{dev}(\varepsilon(v))\|_T^2 + \|\operatorname{div}(v)\|_T^2 \tag{4.9}$$

$$\|(\operatorname{Id} - \Pi^{k-1})\kappa(\operatorname{curl} v)\|_T^2 \lesssim \|\operatorname{div}(v)\|_T^2, \tag{4.10}$$

$$\|(\operatorname{Id} - \Pi^{k-1})\nabla v\|_T^2 \lesssim \|\operatorname{div}(v)\|_T^2. \tag{4.11}$$

Proof. One side of the equivalence of (4.9) is obvious by the continuity of the Π^{k-1}. For the other direction, we use the following equivalence on the reference element $\hat{T}$:

$$\|\hat{\nabla}(\hat{q}\hat{x})\|_{\hat{T}} \sim \|\hat{\operatorname{div}}(\hat{q}\hat{x})\|_{\hat{T}}, \qquad \text{for all } \hat{q} \in \mathbb{P}^k(\hat{T}, \mathbb{R}). \tag{4.12}$$

This follows by finite dimensionality, because by the Euler identity, if any one of the above two terms is zero, then $\hat{q} = 0$ (see e.g., [24]). Consequently, given any $v \in \mathbb{P}^k(T, \mathbb{R}^d) + x\mathbb{P}^k(T, \mathbb{R})$, setting $\hat{v} = \mathcal{P}^{-1}(v)$, the following problem is uniquely solvable: find $\hat{b} \in \mathbb{P}^k(\hat{T}, \mathbb{R})$ such that

$$\int_{\hat{T}} \hat{\operatorname{div}}(\hat{x}\hat{b})\, \hat{\operatorname{div}}(\hat{x}\hat{q})\, \mathrm{d}x = \int_{\hat{T}} \hat{\operatorname{div}}(\hat{v})\, \hat{\operatorname{div}}(\hat{x}\hat{q})\, \mathrm{d}x, \qquad \text{for all } \hat{q} \in \mathbb{P}^k(\hat{T}, \mathbb{R}). \tag{4.13}$$

Since $\hat{\operatorname{div}}(\hat{x}\mathbb{P}^k(\hat{T}, \mathbb{R})) = \mathbb{P}^k(\hat{T}, \mathbb{R})$, (4.13) implies that $\hat{\operatorname{div}}(\hat{x}\hat{b}) = \hat{\operatorname{div}}(\hat{v})$. Put $r = \mathcal{P}^{-1}(\hat{x}\hat{b})$. Then, due to the properties of the Piola map $\mathcal{P}$, r is a function in $\mathbb{P}^k(T, \mathbb{R}^d) + x\mathbb{P}^k(T, \mathbb{R})$ satisfying $\operatorname{div}(r) = \operatorname{div}(v)$ in T, and a scaling argument using (4.12) implies

$$\|\nabla r\|_T \sim \|\operatorname{div}(r)\|_T. \tag{4.14}$$

Let $a = v - r \in \mathbb{P}^k(T, \mathbb{R}^d) + x\mathbb{P}^k(T, \mathbb{R})$. Then $\operatorname{div}(a) = 0$ and $v = a + r$ in T. Then we have

$$\begin{aligned}
\|\varepsilon(v)\|_T = \|\varepsilon(a + r)\|_T &\lesssim \|\operatorname{dev}(\varepsilon(a + r))\|_T + \|\operatorname{div}(v)\|_T \\
&\le \|\operatorname{dev}(\varepsilon(a))\|_T + \|\nabla r\|_T + \|\operatorname{div}(v)\|_T \\
&\lesssim \|\operatorname{dev}(\varepsilon(a))\|_T + \|\operatorname{div}(v)\|_T && \text{by (4.14).}
\end{aligned}$$

Since Lemma 4.3 implies that $a \in \mathbb{P}^k(T, \mathbb{R}^d)$,

$$\begin{aligned}
\|\varepsilon(v)\|_T &= \|\Pi^{k-1} \operatorname{dev}(\varepsilon(a))\|_T + \|\operatorname{div}(v)\|_T \\
&\le \|\Pi^{k-1} \operatorname{dev}(\varepsilon(v))\|_T + \|\Pi^{k-1} \operatorname{dev}(\varepsilon(r))\|_T + \|\operatorname{div}(v)\|_T \\
&\lesssim \|\Pi^{k-1} \operatorname{dev}(\varepsilon(v))\|_T + \|\operatorname{div}(v)\|_T && \text{by (4.14),}
\end{aligned}$$

which proves (4.9).

To prove (4.10), first note that due to the definition of $\kappa(\cdot)$, the equivalence $\|\kappa(\operatorname{curl} v)\|_T \sim \|\operatorname{curl}(v)\|_T$ holds. Thus, using the same decomposition as above, namely, $v = a + r$,

$$\|(\operatorname{Id} - \Pi^{k-1})\kappa(\operatorname{curl}(v))\|_T \le \|(\operatorname{Id} - \Pi^{k-1})\kappa(\operatorname{curl}(a))\|_T + \|(\operatorname{Id} - \Pi^{k-1})\kappa(\operatorname{curl}(r))\|_T.$$

As $\operatorname{curl}(a) \in \mathbb{P}^{k-1}(T, \mathbb{R}^{\tilde{d}})$, the first term on the right vanishes. The last term satisfies

$$\|(\operatorname{Id} - \Pi^{k-1})\kappa(\operatorname{curl}(r))\|_T \lesssim \|\operatorname{curl}(r)\|_T \le \|\nabla r\|_T \lesssim \|\operatorname{div}(r)\|_T = \|\operatorname{div}(v)\|_T,$$

due to (4.14). Hence (4.10) is proved.

The proof of (4.11) uses the same technique:

$$\|(\operatorname{Id} - \Pi^{k-1})\nabla v\|_T \le \|(\operatorname{Id} - \Pi^{k-1})\nabla a\|_T + \|(\operatorname{Id} - \Pi^{k-1})\nabla r\|_T \lesssim \|\operatorname{div}(v)\|_T,$$

where we have used that $a \in \mathbb{P}^k(T, \mathbb{R}^d)$ and (4.14). $\square$

Remark 4.5. The same technique shows that $\|\nabla v\|_T^2 \sim \|\Pi^{k-1}[\operatorname{dev}(\nabla v)]\|_T^2 + \|\operatorname{div}(v)\|_T^2$ for all Raviart-Thomas functions $v \in \mathbb{P}^k(T,\mathbb{R}^d)+x\mathbb{P}^k(T,\mathbb{R})$. The technique allows one to control the gradient of the highest order terms of a Raviart-Thomas function v by $\operatorname{div}(v)$. The same estimate does *not* hold for all v in $\mathbb{P}^{k+1}(T,\mathbb{R}^d)$.

LEMMA 4.6. *For all $T \in \mathcal{T}_h$ and $\eta_h \in W_h$,*

$$\|\nabla \eta_h\|_T \sim \|\operatorname{curl} \eta_h\|_T.$$

Proof. The proof is based on a scaling argument and equivalence of norms on finite dimensional spaces on the reference element. Recall the map ϕ and $F = \phi'$. Calculations using the chain rule yield

$$\hat{\operatorname{curl}}\big[F^{\mathrm{T}}(\eta_h \circ \phi)F\big] = F^{\mathrm{T}}\big[\operatorname{curl}(\eta_h)\circ\phi\big]F^{-\mathrm{T}}\det F, \qquad \text{if } d=3, \tag{4.15a}$$

$$\hat{\operatorname{curl}}\big[F^{\mathrm{T}}(\eta_h \circ \phi)F\big] = F^{\mathrm{T}}\big[\operatorname{curl}(\eta_h)\circ\phi\big]\det F, \qquad \text{if } d=2. \tag{4.15b}$$

We continue with the $d=3$ case only (since $d=2$ case proceeds using (4.15b) analogously). With $\hat{\eta}_h = F^{\mathrm{T}}(\eta_h\circ\phi)F$, standard estimates for F yield

$$\|\operatorname{curl}(\eta_h)\|_T^2 \sim h^{-3}\|\hat{\operatorname{curl}}(\hat{\eta}_h)\|_{\hat{T}}^2. \tag{4.16}$$

Let $\hat{v} \in \mathbb{P}^k(\hat{T},\mathbb{R}^d)$ and $v \in \mathbb{P}^k(T,\mathbb{R}^d)$ be such that $\hat{\eta}_h = \kappa(\hat{v})$ and $\eta_h = \kappa(v)$, where κ is as defined in (3.1). Then,

$$\|\nabla\eta_h\|_T^2 \sim \|\nabla v\|_T^2 \sim h^{-3}\|\hat{\nabla}\hat{v}\|_{\hat{T}}^2 \sim h^{-3}\|\hat{\nabla}\hat{\eta}_h\|_{\hat{T}}^2 \tag{4.17}$$

In view of (4.16) and (4.17), to complete the proof, it suffices to establish the reference element estimate

$$\|\hat{\operatorname{curl}}(\kappa(\hat{v}))\|_{\hat{T}} \sim \|\hat{\nabla}\hat{v}\|_{\hat{T}} \tag{4.18}$$

by proving that one side is zero if and only if the other side is zero. Note these two identities: $\hat{\operatorname{curl}}\,\kappa(\hat{v}) = (\hat{\nabla}\hat{v})^{\mathrm{T}} - \hat{\operatorname{div}}(\hat{v})\,\mathrm{Id}$, and $\hat{\operatorname{curl}}\kappa(\hat{v}) : \mathrm{Id} = -2\operatorname{div}(\hat{v})$. If $\hat{\operatorname{curl}}\,\kappa(\hat{v}) = 0$, then the latter identity implies $\hat{\operatorname{div}}(\hat{v}) = 0$, which when used in the former identity, yields $\hat{\nabla}\hat{v} = 0$. Combined with the obvious converse, we have established (4.18). ∎

LEMMA 4.7. *For all $T \in \mathcal{T}_h$ and $(v_h, \eta_h) \in U_h$,*

$$\begin{aligned}&\|\varepsilon(v_h)\|_T^2 + \|\kappa(\operatorname{curl} v_h) - \eta_h\|_T^2\\ &\qquad \sim \|\Pi^{k-1}\operatorname{dev}(\nabla v_h - \eta_h)\|_T^2 + h^2\|\operatorname{curl}(\eta_h)\|_T^2 + \|\operatorname{div}(v_h)\|_T^2.\end{aligned}$$

Proof. Since the decomposition $\nabla v_h = \varepsilon(v_h) + \kappa(\operatorname{curl}(v_h))$ is orthogonal in the Frobenius inner product, so is $\nabla v_h - \eta_h = \varepsilon(v_h) + [\kappa(\operatorname{curl}(v_h) - \eta_h]$. Application of the deviatoric and Π^{k-1} preserves this orthogonality. Hence, by Pythagoras theorem,

$$\big\|\Pi^{k-1}\operatorname{dev}(\nabla v_h - \eta_h)\big\|_T^2 = \big\|\Pi^{k-1}\operatorname{dev}(\varepsilon(v_h))\big\|_T^2 + \big\|\Pi^{k-1}[\kappa(\operatorname{curl}(v_h)) - \eta_h]\big\|_T^2. \tag{4.19}$$

We shall now prove the result using (4.19) and Lemma 4.4.

Proof of "$\lesssim$": Since

$$\begin{aligned}\|\varepsilon(v_h)\|_T^2 &\lesssim \|\Pi^{k-1}\operatorname{dev}(\varepsilon(v_h))\|_T^2 + \|\operatorname{div}(v_h)\|_T^2 && \text{by Lemma 4.4,}\\ &\le \big\|\Pi^{k-1}\operatorname{dev}(\nabla v_h - \eta_h)\big\|_T^2 + \|\operatorname{div}(v_h)\|_T^2 && \text{by (4.19),}\end{aligned}$$

it suffices to prove that

$$\|\kappa(\operatorname{curl}(v_h)) - \eta_h\|_T^2 \lesssim \big\|\Pi^{k-1}\operatorname{dev}(\nabla v_h - \eta_h) + h^2\|\operatorname{curl}(\eta_h)\|_T^2 + \|\operatorname{div}(v_h)\|_T^2, \tag{4.20}$$

which we do next. Since the projection $r_1 = \Pi^{k-1}(\kappa(\operatorname{curl}(v_h)) - \eta_h)$ can be bounded using (4.19), we focus on the remainder $r_2 = (\operatorname{Id} - \Pi^{k-1})(\kappa(\operatorname{curl}(v_h)) - \eta_h)$.

$$\begin{aligned}
\|r_2\|_T^2 &\le \|(\operatorname{Id}-\Pi^{k-1})\kappa(\operatorname{curl}(v_h))\|_T^2 + \|(\operatorname{Id}-\Pi^{k-1})\eta_h\|_T^2 \\
&\le \|\operatorname{div}(v_h)\|_T^2 + h^2\|\nabla\eta_h\|_T^2 && \text{by (4.10), Lemma 4.4,} \\
&\lesssim \|\operatorname{div}(v_h)\|_T^2 + h^2\|\operatorname{curl}(\eta_h)\|_T^2 && \text{by Lemma 4.6.}
\end{aligned}$$

When this estimate for r_2 is used in $\|\kappa(\operatorname{curl}(v_h)) - \eta_h\|_T^2 = \|r_1\|_T^2 + \|r_2\|_T^2$, and r_1 is bounded using (4.19), we obtain (4.20).

Proof of "$\gtrsim$": The last term of the lemma obviously satisfies $\|\operatorname{div}(v_h)\|_T^2 \lesssim \|\varepsilon(v_h)\|_T^2$, while the first term satisfies (by (4.19)).

$$\|\Pi^{k-1}\operatorname{dev}(\nabla v_h - \eta_h)\|_T^2 \le \|\varepsilon(v_h)\|_T^2 + \|\kappa(\operatorname{curl}(v_h)) - \eta_h\|_T^2$$

It remains to bound $h^2\|\operatorname{curl}(\eta_h)\|_T^2$. As $\operatorname{curl}[\kappa(\operatorname{curl}(\Pi^{\mathbb{E}} v_h))] = 0$, we obtain using an inverse inequality for polynomials

$$\begin{aligned}
h^2\|\operatorname{curl}\eta_h\|_T^2 = h^2\|\operatorname{curl}(\eta_h - \kappa(\operatorname{curl}(\Pi^{\mathbb{E}} v_h)))\|_T^2 &\lesssim \|\eta_h - \kappa(\operatorname{curl}\Pi^{\mathbb{E}} v_h)\|_T^2 \\
&\le \|\eta_h - \kappa(\operatorname{curl}(v_h))\|_T^2 + \|\kappa(\operatorname{curl}(v_h)) - \kappa(\operatorname{curl}\Pi^{\mathbb{E}} v_h)\|_T^2 \\
&\sim \|\eta_h - \kappa(\operatorname{curl}(v_h))\|_T^2 + \|\operatorname{curl}(v_h - \Pi^{\mathbb{E}} v_h)\|_T^2 \\
&\lesssim \|\eta_h - \kappa(\operatorname{curl}(v_h))\|_T^2 + \|\varepsilon(v_h)\|_T^2,
\end{aligned}$$

where we used (4.6) in the last step. ∎

LEMMA 4.8. *For any $v_h \in V_h$ and $\gamma_h \in W_h$,*

$$h\|\nabla\gamma_h\|_h^2 \lesssim \inf_{v_h\in V_h} \|(v_h,\gamma_h)\|_{U_h} \le \|\gamma_h\|^2, \qquad \|v_h\|_{1,h,\varepsilon} = \inf_{\eta_h\in W_h} \|(v_h,\eta_h)\|_{U_h}. \tag{4.21}$$

While the first estimate in (4.21) involves only the local constants from Lemmas 4.6 and 4.7, using the global constant c_K, we also have

$$(1+c_K)^{-1}\|\gamma_h\| \le \inf_{v_h\in V_h} \|(v_h,\gamma_h)\|_{U_h}. \tag{4.22}$$

Proof. To prove the first estimate of (4.21),

$$\begin{aligned}
\|(v_h,\gamma_h)\| \ge \|\varepsilon(v_h)\|_h^2 + \|\kappa(\operatorname{curl} v_h) - \gamma_h\|_h^2 &\gtrsim h^2\|\operatorname{curl}\gamma_h\|_h^2 && \text{by Lemma 4.7} \\
&\gtrsim h^2\|\nabla\gamma_h\|_h^2 && \text{by Lemma 4.6.}
\end{aligned}$$

Taking infimum over $v_h \in V_h$, we obtain the lower estimate of (4.21). The upper bound of the first infimum obviously follows by choosing $v_h = 0$.

To prove the equality in (4.21), observe that the infimum over $\eta_h \in W_h$ cannot be larger than $\|v_h\|_{1,h,\varepsilon}$ because we may choose $\eta_h = \kappa(\operatorname{curl} v_h)$. The reverse inequality also holds since $\|(v_h,\eta_h)\|_{U_h} \ge \|v_h\|_{1,h,\varepsilon}$ for any $\eta_h \in W_h$, so the equality must hold.

Finally, to prove (4.22), we use triangle inequality to get

$$\|\gamma_h\| \le \|\kappa(\operatorname{curl} v_h) - \gamma_h\|_h + \|\operatorname{curl} v_h\|_h \le \|(v_h,\gamma_h)\|_{U_h} + \|\nabla v_h\|_h.$$

Applying the Korn inequality (4.8) and noting that the jump of the normal components are zero for functions in $v_h \in H_0(\operatorname{div},\Omega)$, the proof is complete. ∎

4.3. Stability analysis. The next three lemmas lead us to a discrete inf-sup condition.

LEMMA 4.9. *Let* $\mu \in \mathbb{P}^k(T, \mathbb{M})$ *for some arbitrary element* $T \in \mathcal{T}_h$ *and define* $\tau = (\det F)\,\mathrm{dev}(\mathrm{curl}(\mathrm{curl}(\mu)B))$. *Then for* $d = 3, 2$,

$$\|\tau\|_T \sim h^{3-d} \|\,\mathrm{curl}(\mu)\|_T.$$

Proof. If $\mathrm{curl}\,\mu = 0$, then obviously $\tau = 0$. We claim that the converse is also true. Indeed, if $\tau = 0$, then putting $s = d^{-1}\mathrm{tr}(\mathrm{curl}(\mathrm{curl}(\mu)B))$, we have

$$\mathrm{curl}(\mathrm{curl}(\mu)B) = s\ \mathrm{Id}\,. \tag{4.23}$$

Taking divergence on both sides, we find that $\nabla s = 0$, so s must be a constant on T. Then, taking normal components of both sides of (4.23) on each facet, we find that $sn = 0$, so $s = 0$. Hence $\mathrm{curl}(\mathrm{curl}(\mu)B) = 0$, which in turn implies that $0 = (\mathrm{curl}(\mathrm{curl}(\mu)B, \mu)_T = (\mathrm{curl}(\mu)B, \mathrm{curl}(\mu))_T = 0$. Therefore, by [8, Lemma 2.2], $\mathrm{curl}(\mu) = 0$. Thus $\tau = 0$ if and only if $\mathrm{curl}\,\mu = 0$.

Applying this on the reference element $\hat{T}$ for $\hat{\mu} = F^{\mathrm{T}}(\mu \circ \phi)F \in \mathbb{P}^k(T, \mathbb{M})$ and $\hat{\tau} = \mathrm{dev}(\hat{\mathrm{curl}}(\hat{\mathrm{curl}}(\hat{\mu})\hat{B}))$ where $\hat{B}$ is in Remark 3.4, by finite dimensionality, we have

$$\|\hat{\tau}\|_{\hat{T}} \sim \|\hat{\mathrm{curl}}(\hat{\mu})\|_{\hat{T}}. \tag{4.24}$$

We will now show that $\tau = (\det F)\,\mathrm{dev}(\mathrm{curl}(\mathrm{curl}(\mu)B))$ is related to $\hat{\tau}$ by

$$\tau = \mathcal{M}(\hat{\tau}). \tag{4.25}$$

By the definition of $\mathcal{M}$,

$$(\det F)\,\mathcal{M}(\hat{\tau}) \circ \phi = F^{-\mathrm{T}}\,\mathrm{dev}(\hat{\mathrm{curl}}(\hat{\mathrm{curl}}(\hat{\mu})\hat{B}))F^{\mathrm{T}} = \mathrm{dev}(F^{-\mathrm{T}}\hat{\mathrm{curl}}(\hat{\mathrm{curl}}(\hat{\mu})\hat{B})F^{\mathrm{T}})$$

as trace is preserved under similarity transformations. Focusing on the part of the last term inside the deviatoric, in the $d = 3$ case,

$$\begin{aligned}
F^{-\mathrm{T}}\hat{\mathrm{curl}}(\hat{\mathrm{curl}}(\hat{\mu})\hat{B})F^{\mathrm{T}} &= F^{-\mathrm{T}}\hat{\mathrm{curl}}\big[\hat{\mathrm{curl}}(F^{\mathrm{T}}(\mu \circ \phi)F)\,F^{\mathrm{T}}(B \circ \phi)F\big]F^{\mathrm{T}} && \text{by (3.9)},\\
&= F^{-\mathrm{T}}\hat{\mathrm{curl}}\big[F^{\mathrm{T}}[\mathrm{curl}(\mu) \circ \phi]\,F^{-\mathrm{T}}(\det F)\,F^{\mathrm{T}}(B \circ \phi)F\big]F^{\mathrm{T}} && \text{by (4.15)},\\
&= (\det F)F^{-\mathrm{T}}\hat{\mathrm{curl}}\big[F^{\mathrm{T}}[\mathrm{curl}(\mu)B] \circ \phi\,F\big]F^{\mathrm{T}}\\
&= (\det F)^2 F^{-\mathrm{T}}F^{\mathrm{T}}\big[\,\mathrm{curl}(\mathrm{curl}(\mu)B) \circ \phi\big]F^{-\mathrm{T}}F^{\mathrm{T}} && \text{by (4.15)}.
\end{aligned}$$

This proves that

$$F^{-\mathrm{T}}\hat{\mathrm{curl}}(\hat{\mathrm{curl}}(\hat{\mu})\hat{B})F^{\mathrm{T}} = (\det F)^2\,\mathrm{curl}(\mathrm{curl}(\mu)B) \circ \phi$$

when $d = 3$. The same identity holds in the $d = 2$ case: the argument is similar after changing the definitions of the curls and the mapping of B appropriately. Thus, $\mathcal{M}(\hat{\tau}) \circ \phi = (\det F)\,\mathrm{dev}(\mathrm{curl}(\mathrm{curl}(\mu)B)) \circ \phi$ and (4.25) is proved.

Finally, the result follows from (4.25) by scaling arguments: indeed (4.24) implies, by (4.3) and (4.15), that

$$\begin{aligned}
h^3\|\tau\|_T^2 &\sim h^3\|\,\mathrm{curl}\,\mu\|_T^2 && \text{if } d = 3,\\
h^2\|\tau\|_T^2 &\sim h^4\|\,\mathrm{curl}\,\mu\|_T^2 && \text{if } d = 2,
\end{aligned}$$

from which the result follows. ∎

LEMMA 4.10. *For any $\gamma_h \in W_h$, there is a $\tau_h \in \Sigma_h^+$ such that*

$$(\tau_h, \gamma_h)_\Omega \gtrsim h\| \operatorname{curl} \gamma_h\|_h \, \|\tau_h\|. \tag{4.26}$$

Furthermore, for any $v_h \in V_h$, the same γ_h, τ_h pair satisfies

$$b_2(\tau_h, (v_h, \gamma_h)) \gtrsim \Big[h\| \operatorname{curl}(\gamma_h)\|_h - \| \operatorname{div}(v_h)\|_h\Big] \|\tau_h\|. \tag{4.27}$$

Proof. Given a $\gamma_h \in W_h$, set τ_h element by element by

$$\tau_h|_T = (\det F)\ \operatorname{dev}(\operatorname{curl}(\operatorname{curl}(\gamma_h|_T)B)).$$

Clearly, $\operatorname{dev}(\operatorname{curl}(\operatorname{curl}(\Pi^{k-1}\gamma_h)B))$ is in Σ_h. Since $\operatorname{dev}(\operatorname{curl}(\operatorname{curl}(\gamma_h - \Pi^{k-1}\gamma_h)B))$ is in $\delta\Sigma_h$, we conclude that $\tau_h \in \Sigma_h^+$. Since γ_h is trace-free, there holds the equivalence $(\tau_h, \gamma_h)_T = (\operatorname{curl}(\operatorname{curl}(\gamma_h|_T)B), \gamma_h)_T\ \det F$, which in turn implies, after integrating by parts and applying Lemma 3.2, $(\tau_h, \gamma_h)_T = (\operatorname{curl}(\gamma_h)B, \operatorname{curl}\gamma_h)_T \det F$.

In the $d = 3$ case, this yields

$$(\tau_h, \gamma_h)_T = \det F \int_T \sum_{i=0}^{3} \lambda_{i-3}\lambda_{i-2}\lambda_{i-1}| \operatorname{curl}(\gamma_h)\nabla\lambda_i|^2 \,\mathrm{d}x \tag{4.28}$$

Noting that $\nabla\lambda_i = -n_i/h_i$, where h_i is the distance from the ith vertex to the facet of the simplex opposite to it, and that the ℓ^2-norm of any matrix $m \in \mathbb{M}$ is equivalent to the sum of ℓ^2-norms of mn_i, a local scaling argument with $m = \operatorname{curl}(\gamma_h)$ and (4.28) imply

$$(\tau_h, \gamma_h)_T \gtrsim (\det F)h^{-2}\| \operatorname{curl}(\gamma_h)\|_T^2.$$

Therefore, $(\tau_h, \gamma_h)_\Omega \gtrsim h\| \operatorname{curl}(\gamma_h)\|_h^2 \gtrsim h\| \operatorname{curl}(\gamma_h)\|_h \, \|\tau_h\|$, by Lemma 4.9. This proves (4.26) in the $d = 3$ case. In the $d = 2$ case, the analogue of (4.28) gives $(\tau_h, \gamma_h)_T \gtrsim (\det F)\ \| \operatorname{curl}(\gamma_h)\|_T^2 \gtrsim h^2\| \operatorname{curl}(\gamma_h)\|_T^2 \geq h\| \operatorname{curl}(\gamma_h)\|_T \, \|\tau_h\|$, where we have used Lemma 4.9 again. This completes the proof of (4.26).

To prove (4.27), we use (3.11). The last sum in

$$b_2(\tau_h, (v_h, \gamma_h)) = -\sum_{T\in\mathcal{T}_h} \int_T \tau_h : (\nabla v_h - \gamma_h) \,\mathrm{d}x + \sum_{F\in\mathcal{F}_h} \int_F (\tau_h)_{nt} \cdot [\![(v_h)_t]\!] \,\mathrm{d}s$$

vanishes due to Lemma 3.3. Hence by (4.26),

$$b_2(\tau_h, (v_h, \gamma_h)) \gtrsim h\| \operatorname{curl}\gamma_h\|_h \, \|\tau_h\| - \sum_{T\in\mathcal{T}_h} (\tau_h, \nabla v_h)_T. \tag{4.29}$$

To handle the last term, note that

$$\begin{aligned} \frac{1}{\det F}(\tau_h, \nabla v_h)_T &= (\operatorname{curl}(\operatorname{curl}(\gamma_h)B), \nabla v_h)_T - (d^{-1}\operatorname{tr}(\operatorname{curl}(\operatorname{curl}(\gamma_h)B))\operatorname{Id}, \nabla v_h)_T \\ &= -(d^{-1}\operatorname{tr}(\operatorname{curl}(\operatorname{curl}(\gamma_h)B)), \operatorname{div}(v_h))_T \end{aligned}$$

because $(\operatorname{curl}(\operatorname{curl}(\gamma_h)B), \nabla v_h)_T = 0$. This follows by integrating one of the curls by parts, observing that the resulting volume term is zero (since $\operatorname{curl}(\nabla v_h) = 0$) and so is the resulting boundary term (due to Lemma 3.2). Continuing, we apply Cauchy-Schwarz inequality and an inverse inequality to get

$$\begin{aligned} |(\tau_h, \nabla v_h)_T| &\lesssim |\det F|h^{-1}\|B\|_{L^\infty(T)}\| \operatorname{curl}(\gamma_h)\|_T\| \operatorname{div}(v_h)\|_T \\ &\lesssim \|\tau_h\|_T\| \operatorname{div}(v_h)\|_T \end{aligned}$$

by Lemma 4.9. Returning to (4.29) and using this estimate, the proof is complete. ∎

Remark 4.11. The message of Lemmas 4.9 and 4.10 is that it is possible to choose a τ_h in the form of a deviatoric of a curl of a bubble to bound (from below) the term arising from the weak symmetry constraint. If τ_h was just a curl, it would not be seen by the equilibrium equation and the bound in (4.27) would not have the $\|\operatorname{div}(v_h)\|$-term, but our τ_h is a deviatoric (of a curl), thus necessitating this term.

LEMMA 4.12. *For any $(v_h, \gamma_h) \in U_h$, there is a $\tau_h \in \Sigma_h$ such that*

$$b_2(\tau_h, (v_h, \gamma_h)) \gtrsim \|(v_h, \gamma_h)\|_{U_h,*} \|\tau_h\|.$$

Proof. We only present the proof in two dimensions, as the three dimensional case is similar. From the local element basis exhibited in (3.13) (see also [19, §5.5] for a more detailed discussion), its clear that on any facet $F \in \mathcal{F}_h$, there exists a constant trace-free function S^F with the property that $S^F_{nt} \in \mathbb{P}^0(F, n_F^\perp)$, $\|S^F_{nt}\|_2 = 1$ on the facet F, and S^F_{nt} equals $(0,0)$ on all other facets in $\mathcal{F}_h$. Given any $(v_h, \gamma_h) \in U_h$, define

$$\tau_h^0 := \sum_{T\in\mathcal{T}_h} \sum_{F\in\mathcal{F}_h} -(S^F : \Pi^{k-1} \operatorname{dev}(\nabla v_h - \gamma_h))\, \lambda_T^F\, S^F, \quad \tau_h^1 := \sum_{F\in\mathcal{F}_h} \frac{1}{\sqrt{h}} \Pi^1([\![(v_h)_t]\!])\, S^F,$$

where λ_T^F is the unique barycentric coordinate function on the element T opposite to the facet F (so that $\lambda_T^F S^F$ is an nt-bubble). Clearly, τ_h^0 and τ_h^1 are in Σ_h. Using the norm equivalences stated in (4.5) and the mappings for v_h and γ_h given in (3.3), a scaling argument yields

$$\|\tau_h^0\|^2 \lesssim \sum_{T\in\mathcal{T}_h} \|\Pi^{k-1} \operatorname{dev}(\nabla v_h - \gamma_h))\|_T^2 \quad \text{and} \quad \|\tau_h^1\|^2 \lesssim \sum_{F\in\mathcal{F}_h} \frac{1}{h} \|\Pi^1 [\![(v_h)_t]\!]\|_F^2.$$

Setting $\tau_h = \alpha_0 \tau_h^0 + \alpha_1 \tau_h^1$ and selecting the constants α_0, α_1 appropriately, the rest of the proof proceeds along the same lines as the proof of [19, Lemma 6.5]. ∎

Remark 4.13. It is interesting to contrast Lemma 4.12 with [19, Lemma 6.5], which also gives a similar LBB-condition. The differences are (i) the velocity space in [19] is $H_0(\operatorname{div}, \Omega) \cap \mathbb{P}^{k+1}(\mathcal{T}_h, \mathbb{R}^d)$, (ii) the velocity norm is a discrete H^1-norm defined using ∇ in place of $\varepsilon(\cdot)$, (iii) there is no weak symmetry constraint and no associated space W_h in [19], and (iv) the stress space in [19] equals the Σ_h in (3.5) plus certain nt-bubbles of degree $k+1$ (different from our $\delta\Sigma_h$ here). Lemma 4.12 shows that the inf-sup condition in [19, Lemma 6.5] continues to hold even *if the nt-bubbles there are removed* and $H_0(\operatorname{div}, \Omega) \cap \mathbb{P}^{k+1}(\mathcal{T}_h, \mathbb{R}^d)$ is replaced by our Raviart-Thomas velocity space V_h. This observation can be extended to prove the convergence of the MCS formulation in [19] with so modified spaces.

THEOREM 4.14 (Discrete LBB-condition). *Let $v_h \in V_h$ and $\gamma_h \in W_h$. Then,*

$$\sup_{(\tau_h, q_h)\in\Sigma_h^+\times Q_h} \frac{b_1(v_h, q_h) + b_2(\tau_h, (v_h, \gamma_h))}{\|\tau_h\| + \|q_h\|} \gtrsim \|(v_h, \gamma_h)\|_{U_h}. \tag{4.30}$$

If v_h is in the divergence-free subspace $V_h^0 := \{z_h \in V_h : \operatorname{div}(z_h) = 0\}$, then

$$\sup_{\tau_h\in\Sigma_h^+} \frac{b_2(\tau_h, (v_h, \gamma_h))}{\|\tau_h\|} \gtrsim \|(v_h, \gamma_h)\|_{U_h}. \tag{4.31}$$

Proof. By Lemmas 4.10 and 4.12, for any given $(v_h, \gamma_h) \in U_h$, there are $\tau_h^1, \tau_h^2 \in \Sigma_h^+$ satisfying

$$b_2(\tau_h^1, (v_h, \gamma_h)) \gtrsim \Big[h \| \operatorname{curl}(\gamma_h)\|_h - \| \operatorname{div}(v_h)\|\Big] \|\tau_h^1\|, \tag{4.32}$$

$$b_2(\tau_h^2, (v_h, \gamma_h)) \gtrsim \|(v_h, \gamma_h)\|_{U_h,*} \|\tau_h^2\|, \tag{4.33}$$

Clearly, the same inequalities hold when τ_h^1 and τ_h^2 are scaled by any nonzero factor, so we may assume without loss of generality, that they have been scaled so that $\|\tau_h^1\| = h \| \operatorname{curl} \gamma_h\|_h$ and $\|\tau_h^2\| = \|(v_h, \gamma_h)\|_{U_h,*}$. Set $\tau_h = \alpha\tau_h^1 + \tau_h^2$, where $\alpha \in \mathbb{R}$ is to be chosen shortly. It follows from (4.32) and (4.33) that

$$b_2(\tau_h, (v_h, \gamma_h)) \gtrsim \alpha h^2 \| \operatorname{curl} \gamma_h\|_h^2 - \alpha h \| \operatorname{div}(v_h)\|_h \| \operatorname{curl} \gamma_h\|_h + \|(v_h, \gamma_h)\|_{U_h,*}^2. \tag{4.34}$$

Next, we choose $q_h \in Q_h$ so that $q_h = \beta \operatorname{div}(v_h)$, where $\beta \in \mathbb{R}$ is another constant to be chosen shortly. Then (4.34) implies

$$\begin{aligned} b_1(v_h, q_h) + b_2(\tau_h, (v_h, \gamma_h)) = \beta \| \operatorname{div}(v_h)\|_h^2 + \alpha h^2 \| \operatorname{curl} \gamma_h\|_h^2 + \|(v_h, \gamma_h)\|_{U_h,*}^2 \\ - \alpha h \| \operatorname{div}(v_h)\|_h \| \operatorname{curl} \gamma_h\|_h. \end{aligned}$$

Choose any $\alpha > 1$ and $\beta > \alpha^2/2$. Then, using Young's inequality for the last term,

$$b_1(v_h, q_h) + b_2(\tau_h, (v_h, \gamma_h)) \gtrsim \| \operatorname{div}(v_h)\|_h^2 + h^2 \| \operatorname{curl} \gamma_h\|_h^2 + \|(v_h, \gamma_h)\|_{U_h,*}^2.$$

Recalling that we also have

$$\|\tau_h\|_{\Sigma_h^+}^2 + \|q_h\|^2 \lesssim \| \operatorname{div}(v_h)\|_h^2 + h^2 \| \operatorname{curl} \gamma_h\|_h^2 + \|(v_h, \gamma_h)\|_{U_h,*}^2,$$

we can now conclude the proof of (4.30) using the norm equivalence of Lemma 4.7. The proof of (4.31) is similar (and in fact simpler since all terms involving $\operatorname{div}(v_h)$ vanish). ◻

4.4. Error estimates. In this subsection we show that the error in the discrete MCS solution converges at optimal order. As we have chosen polynomials of degree k for the stress space Σ_h, the optimal rate of convergence for $\|\sigma - \sigma_h\|$ is $\mathcal{O}(h^{k+1})$. However, the optimal rate for the velocity error in our discrete H^1-like norm, namely, $\|u - u_h\|_{1,h,\varepsilon}$ is only $\mathcal{O}(h^k)$ (since the Raviart-Thomas velocity space V_h only contains $\mathbb{P}^k(T, \mathbb{R}^d)$ within each mesh element T). Nevertheless, we are still able to prove optimal convergence rate of the stress error by using an appropriate interpolation operator and deducing that the stress error is independent of the velocity error. Another important property we shall conclude in this subsection is the pressure-robustness of the method.

Lemma 4.15 (Continuity). *The bilinear forms a, b_1 and b_2 are continuous:*

$$\begin{aligned} a(\varsigma_h, \tau_h) &\lesssim \big(\nu^{-1/2}\|\varsigma_h\|\big)\big(\nu^{-1/2}\|\tau_h\|\big), && \text{for all } \varsigma_h, \tau_h \in \Sigma_h^+, \\ b_1(v_h, q_h) &\lesssim \|(v_h, 0)\|_{U_h} \|q_h\|, && \text{for all } v_h \in V_h, q_h \in Q_h, \\ b_2(\tau_h, (v_h, \eta_h)) &\lesssim \|\tau_h\| \, \|(v_h, \eta_h)\|_{U_h}, && \text{for all } \tau_h \in \Sigma_h^+, (v_h, \eta_h) \in U_h. \end{aligned}$$

Proof. The continuity of a and b_1 follow by the Cauchy Schwarz inequality. For b_2, we use (3.11) and $\nabla v_h = \varepsilon(v_h) + \kappa(\operatorname{curl} v_h)$ to get

$$b_2(\tau_h, (v_h, \eta_h)) = -\sum_{T \in \mathcal{T}_h} \int_T \tau : \big[\varepsilon(v_h) + (\kappa(\operatorname{curl} v_h) - \eta_h)\big] \,\mathrm{d}x + \sum_{F \in \mathcal{F}_h} \int_F \tau_{nt} \cdot [\![(v_h)_t]\!] \,\mathrm{d}s\,.$$

Now, Cauchy-Schwarz inequality and (4.5) of Lemma 4.1 finishes the proof. ◻

LEMMA 4.16 (Coercivity in the kernel). *For all* (τ_h, q_h) *in the kernel*

$$K_h := \{(\tau_h, q_h) \in \Sigma_h \times Q_h : b_1(v_h, q_h) + b_2(\tau_h, (v_h, \eta_h)) = 0 \textit{ for all } (v_h, \eta_h) \in U_h\},$$

we have $\nu^{-1}\big(\|\tau_h\| + \|q_h\|\big)^2 \lesssim a(\tau_h, \tau_h)$.

Proof. By [25, Theorem 2.2], for any $q_h \in Q_h$, there is a $v_h \in V_h$ such that $\|q_h\|^2 \lesssim (\operatorname{div}(v_h), q_h)$ and a discrete H^1-norm of v_h is bounded by $\|q_h\|$. The latter bound implies, in particular, that $\|v_h\|_{1,h,\varepsilon} \lesssim \|q_h\|$, and also that $\eta_h = \kappa(\operatorname{curl} v_h)$ satisfies $\|(v_h, \eta_h)\|_{U_h} \lesssim \|q_h\|$. This together with Lemma 4.15 implies

$$\|q_h\|^2 \lesssim b_1(v_h, q_h) = -b_2(\tau_h, (v_h, \eta_h)) \lesssim \|\tau_h\| \, \|(v_h, \eta_h)\|_{U_h} \lesssim \|\tau_h\| \, \|q_h\|$$

yielding the needed bound for $\|q_h\|$. ∎

We are now ready to conclude an inf-sup condition for the bilinear form

$$\begin{aligned} B(v_h, \eta_h, &\tau_h, q_h; \tilde v_h, \tilde \eta_h, \tilde \tau_h, \tilde q_h) \\ &:= a(\tau_h, \tilde\tau_h) + b_1(v_h, \tilde q_h) + b_1(\tilde v_h, q_h) + b_2(\tau_h, (\tilde v_h, \tilde\eta_h)) + b_2(\tilde\tau_h, (v_h, \eta_h)). \end{aligned}$$

COROLLARY 4.17. *Let* $\tau_h \in \Sigma_h^+$, $v_h \in V_h$, $\eta_h \in W_h$, *and* $q_h \in Q_h$. *There holds*

$$\|(v_h, \eta_h, \tau_h, q_h)\|_* \lesssim \sup_{\substack{\tilde v_h \in V_h,\ \tilde\eta_h \in W_h \\ \tilde\tau_h \in \Sigma_h^+,\ \tilde q_h \in Q_h}} \frac{B(v_h, \eta_h, \tau_h, q_h; \tilde v_h, \tilde\eta_h, \tilde\tau_h, \tilde q_h)}{\|(\tilde v_h, \tilde\eta_h, \tilde\tau_h, \tilde q_h)\|_*}, \tag{4.35}$$

so, in particular, there is a unique solution for the discrete MCS system (3.12). *Moreover, if* v_h *is restricted to* V_h^0, *we also have*

$$\|(v_h, \eta_h, \tau_h, 0)\|_* \lesssim \sup_{\tilde v_h \in V_h^0,\ \tilde\eta_h \in W_h,\ \tilde\tau_h \in \Sigma_h^+} \frac{B(v_h, \eta_h, \tau_h, 0; \tilde v_h, \tilde\eta_h, \tilde\tau_h, 0)}{\|(\tilde v_h, \tilde\eta_h, \tilde\tau_h, 0)\|_*}. \tag{4.36}$$

Proof. The first inf-sup condition follows from the standard theory of mixed methods [4], using Theorem 4.14 (the inf-sup condition for b_1 and b_2 given by (4.30)), Lemma 4.15 (continuity of forms), and Lemma 4.16 (coercivity in the kernel).

The second inf-sup condition also follows in a similar fashion, but now using the other inequality (4.31) of Theorem 4.14. ∎

THEOREM 4.18 (Consistency). *The MCS method with weakly imposed symmetry* (3.12) *is consistent in the following sense. If the exact solution of the Stokes problem* (3.2) *is such that* $u \in H^1(\Omega, \mathbb{R}^d)$, $\omega \in L^2(\Omega, \mathbb{M})$, $\sigma \in H^1(\Omega, \mathbb{D})$ *and* $p \in L^2_0(\Omega, \mathbb{R})$, *then*

$$B(u, \omega, \sigma, p; v_h, \eta_h, \tau_h, q_h) = (-f, v_h)_\Omega$$

for all $v_h \in V_h, \eta_h \in W_h, q_h \in Q_h$, *and* $\tau_h \in \Sigma_h$.

The proof of Theorem 4.18 is easy (see, e.g., the similar proof of [19, Theorem 6.2]), so we omit it. We now have all the ingredients to prove the following convergence result. Let I_{V_h} denote the standard Raviart-Thomas interpolator (see, e.g., [4]) and let $\|(u, \omega, \sigma, p)\|_{\nu,s} = \nu^{-1}\|\sigma\|_{H^s(\mathcal{T}_h,\mathbb{D})} + \nu^{-1}\|p\|_{H^s(\mathcal{T}_h,\mathbb{R})} + \|\omega\|_{H^s(\mathcal{T}_h,\mathbb{K})} + \|u\|_{H^{s+1}(\mathcal{T}_h,\mathbb{R}^d)}$.

THEOREM 4.19 (Optimal convergence). *Let* $u \in H^1(\Omega, \mathbb{R}^d) \cap H^m(\mathcal{T}_h, \mathbb{R}^d)$, $\sigma \in H^1(\Omega, \mathbb{D}) \cap H^{m-1}(\mathcal{T}_h, \mathbb{D})$, $p \in L^2_0(\Omega, \mathbb{R}) \cap H^{m-1}(\mathcal{T}_h, \mathbb{R})$ *and* $\omega \in L^2(\Omega, \mathbb{K}) \cap H^{m-1}(\mathcal{T}_h, \mathbb{K})$

be the exact solution of the mixed Stokes problem (3.2)*, let* u_h,σ_h, ω_h *and* p_h *solve* (3.12) *and let* $s = \min(m-1, k+1)$. *Then,*

(4.37)
$$\frac{1}{\nu}(\|\sigma - \sigma_h\| + \|p - p_h\|) + \|(\omega_h - \Pi^k\omega, u_h - I_{V_h}u)\|_{U_h} \lesssim h^s \|(0, \omega, \sigma, p)\|_{\nu,s}.$$

Proof. Let $e_h^\sigma = I_{\Sigma_h}\sigma - \sigma_h$, $e_h^u = I_{V_h}u - u_h$, $e_h^\omega = \Pi^k\omega - \omega_h$, $e_h^p = \Pi^k p - p_h$ (where the two occurrences of Π^k represent projections onto two different discrete spaces per our prior notation). Denoting the analogous approximation errors by $a^\sigma = I_{\Sigma_h}\sigma - \sigma$, $a^u = I_{V_h}u - u$, $a^\omega = \Pi^k\omega - \omega$, and $a^p = \Pi^k p - p$, observe that Theorem 4.18 implies

$$B(e_h^u, e_h^\omega, e_h^\sigma, e_h^p;\ v_h, \eta_h, \tau_h, q_h) = B(a^u, a^\omega, a^\sigma, a^p;\ v_h, \eta_h, \tau_h, q_h) \tag{4.38}$$

for any $v_h \in V_h, \eta_h \in W_h, \tau_h \in \Sigma_h^+$, and $q_h \in Q_h$. The right hand side above is a sum of five terms $(\nu^{-1}a^\sigma, \tau_h) + b_1(a^u, q_h) + b_1(v_h, a^p) + b_2(\tau_h, (a^u, a^\omega)) + b_2(a^\sigma, (v_h, \eta_h))$. The second term vanishes: $b_1(a^u, q_h) = (\operatorname{div}(I_{V_h}u - u), q_h) = (\Pi^k \operatorname{div}(u) - \operatorname{div}(u), q_h) = 0$ as $\operatorname{div}(u) = 0$. The third term also vanishes: $b_1(v_h, a^p) = (\operatorname{div}(v_h), \Pi^k p - p) = 0$ since $\operatorname{div}(v_h) \in \mathbb{P}^k(\mathcal{T}_h)$. The fourth term, due to (3.10), is

$$b_2(\tau_h, (a^u, a^\omega)) = (\tau, a^\omega) + \sum_{T\in\mathcal{T}_h} (\operatorname{div}(\tau_h), I_{V_h}u - u)_T - \sum_{E\in\mathcal{F}_h} ([\![(\tau_h)_{nn}]\!], (I_{V_h}u - u)\cdot n)_E$$

where the last two terms vanish by the properties of the Raviart-Thomas d.o.f.s that define I_{V_h}, i.e., $b_2(\tau_h, (a^u, a^\omega)) = (\tau_h, a^\omega)$. The fifth term, due to (3.11), is

$$b_2(a^\sigma, (v_h, \eta_h)) = (a^\sigma, \eta_h - \nabla v_h) + \sum_{E\in\mathcal{F}_h} (a^\sigma_{nt}, [\![(v_h)_t]\!])_E.$$

Writing $(a^\sigma, \eta_h - \nabla v_h) = (a^\sigma, \eta_h) + (a^\sigma, (\Pi^{k-1} - \mathrm{Id})\nabla v_h) - (a^\sigma, \Pi^{k-1}\nabla v_h)$, note that by the d.o.f.s of Theorem 3.5, the last term $(a^\sigma, \Pi^{k-1}\nabla v_h)$ is zero, and moreover, $(a^\sigma, \eta_h) = (a^\sigma, \eta_h - \Pi^0\eta_h)$. Incorporating these observations on each term into (4.38), we obtain

$$\begin{aligned} B(e_h^u, e_h^\omega, e_h^\sigma, e_h^p;\ v_h, \eta_h, \tau_h, q_h) = (\nu^{-1}a^\sigma, \tau_h) + (\tau_h, a^\omega) + \sum_{F\in\mathcal{F}_h} (a^\sigma_{nt}, [\![(v_h)_t]\!])_F \\ + (a^\sigma, \eta_h - \Pi^0\eta_h) + (a^\sigma, (\Pi^{k-1} - \mathrm{Id})\nabla v_h). \end{aligned} \tag{4.39}$$

We proceed with the right hand side of (4.39). By (4.21) and Lemma 4.4,

$$\begin{aligned} \|\eta_h - \Pi^0\eta_h\| &\lesssim h\|\nabla\eta_h\|_h \lesssim \inf_{\tilde{v}_h\in V_h} \|(\tilde{v}_h, \eta_h)\|_{U_h} \le \|(v_h, \eta_h)\|_{U_h}, \\ \|(\Pi^{k-1} - \mathrm{Id})\nabla v_h\|_h &\lesssim \|\operatorname{div}(v_h)\|^2 \lesssim \|\varepsilon(v_h)\|_h^2 \le \|(v_h, \eta_h)\|_{U_h}. \end{aligned}$$

Using these after an application of the Cauchy-Schwarz inequality, (4.39) yields

$$\begin{aligned} &B(e_h^u, e_h^\omega, e_h^\sigma, e_h^p;\ v_h, \eta_h, \tau_h, q_h) \\ &\quad\lesssim \left[\frac{1}{\nu}\left(\|a^\sigma\|^2 + \sum_{F\in\mathcal{F}_h} h\|a^\sigma_{nt}\|_F^2\right) + \nu\|a^\omega\|^2\right]^{1/2} \left(\frac{1}{\nu}\|\tau_h\|^2 + \nu\|(v_h, \eta_h)\|^2_{U_h}\right)^{1/2} \\ &\quad\lesssim \left(\frac{1}{\sqrt{\nu}} h^s\|\sigma\|_{H^s(\mathcal{T}_h)} + \sqrt{\nu}h^s\|\omega\|_{H^s(\mathcal{T}_h)}\right)\|(v_h, \eta_h, \tau_h, q_h)\|_*, \end{aligned} \tag{4.40}$$

where we have used Theorem 3.7 and the approximation property of Π^k.

To complete the proof, we apply triangle inequality starting from the left hand side of (4.37), to get

$$
\begin{aligned}
\frac{1}{\nu}\Big(\|\sigma-\sigma_h\|+\|p-p_h\|\Big)&+\|(e_h^u,e_h^\omega)\|_{U_h}\\
&\le \frac{1}{\nu}\big(\|a^\sigma\|+\|a^p\|+\|e_h^\sigma\|+\|e_h^p\|\big)+\|(e_h^u,e_h^\omega)\|_{U_h}\\
&\lesssim \frac{h^s}{\nu}\big(\|\sigma\|_{H^s(\mathcal{T}_h)}+\|p\|_{H^s(\mathcal{T}_h)}\big)+\frac{1}{\sqrt{\nu}}\|(e_h^u,e_h^\omega,e_h^\sigma,e_h^p)\|_* \qquad (4.41)
\end{aligned}
$$

again using Theorem 3.7. Bounding the last term above using (4.35) and (4.40), the proof is complete. ∎

Remark 4.20 (Convergence in standard norms). Using also Lemma 4.8's estimate (4.22), a consequence of the global discrete Korn inequality, (4.37) implies

$$
\frac{1}{\nu}\|\sigma-\sigma_h\|+\frac{1}{\nu}\|p-p_h\|+\|\omega-\omega_h\|+\|u_h-I_{V_h}u\|_{V_h}\lesssim h^s\|(0,\omega,\sigma,p)\|_{\nu,s} \qquad (4.42)
$$

under the assumptions of Theorem 4.19. Note that even though the optimal rate for $\|u-u_h\|_{1,h,\varepsilon}$ is only $\mathcal{O}(h^k)$, (4.42) gives a *superconvergent* rate of $\mathcal{O}(h^{k+1})$ for $\|u_h-I_{V_h}u\|_{1,h,\varepsilon}$ when the solution is regular enough.

THEOREM 4.21 (Pressure robustness). *Under the same assumptions as Theorem 4.19 there holds*

$$
\frac{1}{\nu}\|\sigma-\sigma_h\|+\|\omega-\omega_h\|+\|u_h-I_{V_h}u\|_{V_h}\lesssim h^s\|(0,\omega,\sigma,0)\|_{\nu,s}.
$$

Proof. Proceeding along the lines of the proof of Theorem 4.19, omitting the pressure error, we obtain, instead of (4.41),

$$
\frac{1}{\nu}\|\sigma-\sigma_h\|+\|(e_h^u,e_h^\omega)\|_{U_h}\lesssim\frac{h^s}{\nu}\|\sigma\|_{H^s(\mathcal{T}_h)}+\frac{1}{\sqrt{\nu}}\|(e_h^u,e_h^\omega,e_h^\sigma,0)\|_*.
$$

We may now complete the proof as before by using (4.36) instead of (4.35). ∎

5. Postprocessing. In this section we describe and analyze a postprocessing for the discrete velocity. While for the raw solution u_h, we may only expect $\|u-u_h\|_{1,h,\varepsilon}$ to go to zero at the rate $\mathcal{O}(h^k)$, we will show that a locally postprocessed velocity u_h^* has error $\|u-u_h^*\|_{1,h,\varepsilon}$ that converges to zero at the higher rate $\mathcal{O}(h^{k+1})$ for sufficiently regular solutions. The key to obtain this enhanced accuracy, as in [31], is the $O(h^{k+1})$-superconvergence of $\|u_h-I_{V_h}u\|_{1,h,\varepsilon}$ – see Remark 4.20. Finally, we shall also show that u_h^* retains the prized structure preservation properties of exact mass conservation and pressure robustness.

The crucial ingredient is a reconstruction operator (see [22, 23]) whose properties are summarized in the next lemma. Let

$$
\begin{aligned}
V_h^* &= H_0(\mathrm{div},\Omega)\cap\mathbb{P}^{k+1}(\mathcal{T}_h,\mathbb{R}^d),\ \text{and}\\
V_h^{*,-} &= \{v_h\in\mathbb{P}^{k+1}(\mathcal{T}_h,\mathbb{R}^d):\ \Pi^k[\![(v_h)_n]\!]=0,\ \text{for all } F\in\mathcal{F}_h\}
\end{aligned}
$$

denote the $\mathcal{BDM}$ space (one order higher) and its "relaxed" analogue, respectively. The next result is a consequence of [22, Lemmas 3.3 and 4.8] and the Korn inequality (4.8).

LEMMA 5.1. *There exists an operator $\mathcal{R} : V_h^{*,-} \to V_h^*$, whose application is computable element-by-element, satisfying*
1. $\|\mathcal{R}v_h\|_{1,h,\varepsilon} \lesssim \|v_h\|_{1,h,\varepsilon}$, *for al* $v_h \in V_h^{*,-}$,
2. $\mathcal{R}v_h^* = v_h^*$ *for all* $v_h^* \in V_h^*$, *and*
3. *whenever the local (element-wise) property* $\operatorname{div}(v_h|_T) = 0$ *holds for all* $T \in \mathcal{T}_h$ *and all* $v_h \in V_h^{*,-}$, *the global property* $\operatorname{div}(\mathcal{R}v_h) = 0$ *holds.*

A simple choice of $\mathcal{R}$ is given by a (DG) generalization of the classical $\mathcal{BDM}$ interpolant. This was used in [20]. Another choice of $\mathcal{R}$, given in [22], based on a simple averaging of coefficients, is significantly less expensive for high orders.

The postprocessed solution $u_h^* \in V_h^*$ is given in two steps as follows. First, using the computed σ_h and u_h, solve the *local* (see Remark 5.3) minimization problem

$$u_h^{*,-} := \underset{\substack{v_h^{*,-} \in V_h^{*,-} \\ I_{V_h}(v_h^{*,-}) = u_h}}{\operatorname{argmin}} \|\nu\varepsilon(v_h^{*,-}) - \sigma_h\|_T^2. \tag{5.1}$$

Second, apply the reconstruction and set $u_h^* := \mathcal{R}(u_h^{*,-})$.

THEOREM 5.2. *Suppose the assumptions of Theorem 4.19 hold. Then* $u_h^* \in V_h^*$, $\operatorname{div}(u_h^*) = 0$, *and for* $s = \min(m-1, k+1)$ *we have the pressure-robust error estimate*

$$\|u - u_h^*\|_{1,h,\varepsilon} \lesssim h^s \|(u, \omega, \sigma, 0)\|_{\nu,s}.$$

Proof. On any $T \in \mathcal{T}_h$, the condition $I_{V_h}(u_h^{*,-}) = u_h$ implies that the Raviart-Thomas d.o.f.s applied to $u_h^{*,-}$ and u_h coincide. Hence, for all $q_h \in \mathbb{P}^k(T, \mathbb{R})$,

$$\begin{aligned}(\operatorname{div}(u_h^{*,-}), q_h)_T &= -(u_h^{*,-}, \nabla q_h)_T + (u_h^{*,-} \cdot n, q_h)_{\partial T} \\ &= -(u_h, \nabla q_h)_T + (u_h \cdot n, q_h)_{\partial T} = (\operatorname{div}(u_h), q_h) = 0\end{aligned}$$

as $\operatorname{div}(u_h) = 0$. Thus, Lemma 5.1 implies that $u_h \in V_h^*$ and $\operatorname{div}(u_h^*) = 0$.

It only remains to prove the error estimate. Let $I_{V_h^*}$ be the standard $\mathcal{BDM}^{k+1}$ interpolator. Then, $u_h^* = \mathcal{R}u_h^{*,-}$ satisfies

$$\begin{aligned}\|u - u_h^*\|_{1,h,\varepsilon} &\leq \|u - I_{V_h^*}u\|_{1,h,\varepsilon} + \|\mathcal{R}(I_{V_h^*}u - u_h^{*,-})\|_{1,h,\varepsilon} && \text{by Lemma 5.1 (2)}, \\ &\lesssim \|u - I_{V_h^*}u\|_{1,h,\varepsilon} + \|u - u_h^{*,-}\|_{1,h,\varepsilon} && \text{by Lemma 5.1 (1)}.\end{aligned}$$

Since standard approximation estimates yield $\|u - I_{V_h^*}u\|_{1,h,\varepsilon} \lesssim h^s \|(u, 0, 0, 0)\|_{\nu,s}$, we focus on the last term. A triangle inequality (where we add and subtract different functions in the element and facet terms) yields

$$\begin{aligned}\|u - u_h^{*,-}\|_{1,h,\varepsilon}^2 \lesssim &\sum_{T \in \mathcal{T}_h} \frac{1}{\nu^2} \|\nu\varepsilon(u) - \sigma_h\|_T^2 + \sum_{T \in \mathcal{T}_h} \frac{1}{\nu^2} \|\sigma_h - \nu\varepsilon(u_h^{*,-})\|_T^2 \\ &+ \sum_{F \in \mathcal{F}_h} \frac{1}{h} \|[\![(u - I_{V_h^*}u)_t]\!]\|_F^2 + \sum_{F \in \mathcal{F}_h} \frac{1}{h} \|[\![(I_{V_h^*}u - u_h^{*,-})_t]\!]\|_F^2.\end{aligned} \tag{5.2}$$

Naming the four sums on the right as s_1, s_2, s_3 and s_4, respectively, we proceed to estimate each. Obviously $s_1 = \nu^{-1}\|\sigma - \sigma_h\| \lesssim h^s \|(0, \omega, \sigma, 0)\|_{\nu,s}$ by Theorem 4.21.

To bound s_2, note that for any w_h in the admissible set of the minimization problem (5.1), we have $s_2 \leq \nu^{-2}\|\sigma_h - \nu\varepsilon(w_h)\|^2$. We choose $w_h = I_{V_h^*}u + u_h - I_{V_h}u \in$

$V_h^* \subset V_h^{*,-}$. Since $I_{V_h} I_{V_h^*} u = I_{V_h} u$ implies $I_{V_h} w_h = u_h$, the chosen w_h is in the admissible set. Hence,

$$\begin{aligned} s_2 &\le \nu^{-2}\|\sigma_h - \nu\varepsilon(w_h)\|^2 \le \nu^{-2}\big(\|\sigma_h - \nu\varepsilon(I_{V_h^*}u)\| + \|\nu\varepsilon(u_h) - \nu\varepsilon(I_{V_h}u)\|\big)^2 \\ &\lesssim \nu^{-2}\|\sigma_h - \nu\varepsilon(u)\|^2 + \nu^{-2}\|\nu\varepsilon(u) - \nu\varepsilon(I_{V_h^*}u)\|^2 + \nu^{-2}\|\nu\varepsilon(u_h) - \nu\varepsilon(I_{V_h}u)\|^2 \\ &= \nu^{-2}\|\sigma_h - \sigma\|^2 + \|u - I_{V_h^*}u\|_{1,h,\varepsilon}^2 + \|u_h - I_{V_h}u\|_{1,h,\varepsilon}^2, \end{aligned}$$

so a standard approximation estimate and Theorem 4.21 yield $s_2 \lesssim h^s\|(u,\omega,\sigma,0)\|_{\nu,s}$.

The same standard approximation estimate for $I_{V_h^*}$ also gives the estimate $s_3 \le \|u - I_{V_h^*}u\|_{1,h,\varepsilon} \lesssim h^s\|(u,\omega,\sigma,0)\|_{\nu,s}$. Hence it only remains to bound s_4. Observe that $I_{V_h^*}u - u_h^{*,-} = I_{V_h}(I_{V_h^*}u - u_h^{*,-}) + (\mathrm{Id} - I_{V_h})(I_{V_h^*}u - u_h^{*,-}) = (I_{V_h}u - u_h) + (\mathrm{Id} - I_{V_h})(I_{V_h^*}u - u_h^{*,-})$, because $I_{V_h} I_{V_h^*} u = I_{V_h} u$ and $I_{V_h} u_h^{*,-} = u_h$. Defining $a = (\mathrm{Id} - I_{V_h})(\mathrm{Id} - \Pi^{\mathbb{E}})(I_{V_h^*}u - u_h^{*,-})$, the identity $I_{V_h^*}u - u_h^{*,-} = (I_{V_h}u - u_h) + a$ holds because $(\mathrm{Id} - I_{V_h})\mathbb{E} = 0$ (as $k \ge 1$). Hence

$$s_4 \lesssim \|I_{V_h}u - u_h\|_{1,h,\varepsilon}^2 + \sum_{F \in \mathcal{F}_h} h^{-1}\big\|[\![a_t]\!]\big\|_F^2. \tag{5.3}$$

Since the first term can be bounded by Theorem 4.21, let us study the last term. On any facet F adjacent to a mesh element T, a trace inequality yields $h^{-1}\big\|[\![a_t]\!]\big\|_F^2 \le h^{-1}\|a_t\|_{\partial T}^2 \lesssim \|\nabla a\|_T^2 + h^{-2}\|a\|_T^2$. Hence,

$$\begin{aligned} h^{-1}\big\|[\![a_t]\!]\big\|_F^2 &\lesssim \|\nabla(\mathrm{Id} - \Pi^{\mathbb{E}})(I_{V_h^*}u - u_h^{*,-})\|_T^2 + h^{-2}\|(\mathrm{Id} - \Pi^{\mathbb{E}})(I_{V_h^*}u - u_h^{*,-})\|_T^2 \\ &\lesssim \|\varepsilon(I_{V_h^*}u - u_h^{*,-})\|_T^2 \end{aligned}$$

where we have used the continuity properties of I_{V_h}, scaling arguments, (4.6), and an estimate analogous to (4.7). Using triangle inequality and returning to (5.3),

$$s_4 \lesssim \|I_{V_h}u - u_h\|_{1,h,\varepsilon}^2 + \|\varepsilon(I_{V_h^*}u - u)\|_h^2 + \nu^{-2}\|\nu\varepsilon(u) - \sigma_h\|_h^2 + \nu^{-2}\|\sigma_h - \nu\varepsilon(u_h^{*,-})\|_h^2.$$

The last two terms are s_1 and s_2, respectively. Hence the prior estimates, the standard approximation estimate for $I_{V_h^*}$, and Theorem 4.21 shows $s_4 \lesssim h^s\|(u,\omega,\sigma,0)\|_{\nu,s}$. $\square$

Remark 5.3. The restriction of the minimizer of (5.1) to an element T, which we denote by $u_T^{*,-} \equiv u_h^{*,-}|_T$, can be computed using the following Euler-Lagrange equations. Letting $\Lambda_h^*(T) = \{\lambda : \lambda|_F \in \mathbb{P}^k(F,\mathbb{R})$ on all facets $F \subset \partial T\}$, the function $u_T^{*,-}$ is the unique function in $\mathbb{P}^{k+1}(T,\mathbb{R}^d)$, which together with $\ell_h^* \in \mathbb{P}^{k-1}(T,\mathbb{R}^d)$ and $\lambda_h^* \in \Lambda_h^*(T)$, satisfies

$$\begin{aligned} (\nu\varepsilon(u_T^{*,-}), \varepsilon(v))_T + (\ell_h^*, v)_T + (\lambda_h^*, v\cdot n)_{\partial T} &= (\sigma_h, \varepsilon(v))_T, \\ (u_T^{*,-}, \wp)_T &= (u_h, \wp)_T, \\ (u_T^{*,-}\cdot n, \mu)_{\partial T} &= (u_h\cdot n, \mu)_{\partial T}, \end{aligned}$$

for all $v \in \mathbb{P}^{k+1}(T,\mathbb{R}^d)$, $\wp \in \mathbb{P}^{k-1}(T,\mathbb{R}^d)$ and $\mu \in \Lambda_h^*(T)$. The last two equations are another way to express the constraint $I_{V_h} u_h^{*,-} = u_h$ in (5.1).

6. Numerical examples. In this last section we present two numerical examples to verify our method. All examples were implemented within the finite element library NGSolve/Netgen, see [29, 30] and on www.ngsolve.org. The computational domain is given by $\Omega = [0,1]^d$ and the velocity field is driven by the volume force determined by $f = -\operatorname{div}(\sigma) + \nabla p$ with the exact solution given by

$$\sigma = \nu\varepsilon(\operatorname{curl}(\psi_2)), \quad \text{and} \quad p := x^5 + y^5 - \frac{1}{3} \quad \text{for } d = 2$$
$$\sigma = \nu\varepsilon(\operatorname{curl}(\psi_3, \psi_3, \psi_3)), \quad \text{and} \quad p := x^5 + y^5 + z^5 - \frac{1}{2} \quad \text{for } d = 3.$$

Here $\psi_2 := x^2(x-1)^2y^2(y-1)^2$ and $\psi_3 := x^2(x-1)^2y^2(y-1)^2z^2(z-1)^2$ defines a given potential in two and three dimensions respectively and we choose the viscosity $\nu = 10^{-3}$.

In Tables 1a and 1b we report the errors in all the computed solution components for varying polynomial orders $k = 1, 2, 3$ in the two and the three dimensional cases, respectively. As predicted by Theorem 4.19 and Theorem 5.2 the corresponding errors converge at optimal order. Furthermore, the L^2-norm of error of the (postprocessed) velocity error converges at one order higher. Note that in three dimensions the errors are already quite small already on the coarsest mesh. It appears that to get out of the preasymptotic regime and see the proper convergence rate, it takes several steps.

REFERENCES

[1] D. N. Arnold, F. Brezzi, and J. Douglas, Jr., *PEERS: a new mixed finite element for plane elasticity*, Japan J. Appl. Math., 1 (1984), pp. 347–367.

[2] D. N. Arnold, R. S. Falk, and R. Winther, *Mixed finite element methods for linear elasticity with weakly imposed symmetry*, Math. Comp., 76 (2007), pp. 1699–1723.

[3] D. Boffi, F. Brezzi, and M. Fortin, *Reduced symmetry elements in linear elasticity*, Commun. Pure Appl. Anal., 8 (2009), pp. 95–121.

[4] ———, *Mixed Finite Element Methods and Applications*, Springer Science & Business Media, 2013.

[5] S. C. Brenner, *Korn's inequalities for piecewise H^1 vector fields*, Math. Comp., 73 (2004), pp. 1067–1087.

[6] F. Brezzi, J. Douglas Jr., and L. D. Marini, *Two families of mixed finite elements for second order elliptic problems*, Numerische Mathematik, 47 (1985), pp. 217–235.

[7] B. Cockburn and J. Gopalakrishnan, *A characterization of hybridized mixed methods for the Dirichlet problem*, SIAM J. Numer. Anal., 42 (2004), pp. 283–301.

[8] B. Cockburn, J. Gopalakrishnan, and J. Guzmán, *A new elasticity element made for enforcing weak stress symmetry*, Math. Comp., 79 (2010), pp. 1331–1349.

[9] B. Cockburn, G. Kanschat, and D. Schötzau, *A locally conservative LDG method for the incompressible Navier-Stokes equations*, Mathematics of Computation, 74 (2005), pp. 1067–1095.

[10] ———, *A note on discontinuous Galerkin divergence-free solutions of the Navier–Stokes equations*, Journal of Scientific Computing, 31 (2007), pp. 61–73.

[11] B. Cockburn and F.-J. Sayas, *Divergence-conforming HDG methods for Stokes flows*, Math. Comp., 83 (2014), pp. 1571–1598.

[12] M. Farhloul, *Mixed and nonconforming finite element methods for the stokes problem*, Canadian Applied Mathematics Quarterly, 3 (Fall 1995).

[13] M. Farhloul and M. Fortin, *A new mixed finite element for the Stokes and elasticity problems*, SIAM J. Numer. Anal., 30 (1993), pp. 971–990.

[14] M. Farhloul and M. Fortin, *Dual hybrid methods for the elasticity and the Stokes problems: a unified approach*, Numer. Math., 76 (1997), pp. 419–440.

[15] M. Farhloul and M. Fortin, *Review and complements on mixed-hybrid finite element methods for fluid flows*, in Proceedings of the 9th International Congress on Computational and Applied Mathematics (Leuven, 2000), vol. 140, 2002, pp. 301–313.

[16] G. Fu, Y. Jin, and W. Qiu, *Parameter-free superconvergent h(div)-conforming HDG methods for the Brinkman equations*, IMA Journal of Numerical Analysis, (2018), p. dry001.

[17] M. Giacomini, A. Karkoulias, R. Sevilla, and A. Huerta, *A superconvergent HDG method for Stokes flow with strongly enforced symmetry of the stress tensor*, Journal of Scientific Computing, 77 (2018), pp. 1679–1702.

$\lvert\mathcal{T}\rvert$	$\lVert\nabla u-\nabla u_h^*\rVert_h$	(eoc)	$\lVert u-u_h^*\rVert$	(eoc)	$\lVert\sigma-\sigma_h\rVert$	(eoc)	$\lVert p-p_h\rVert$	(eoc)	$\lVert\omega-\omega_h\rVert$	(eoc)
					$k=1$					
20	$9.9\cdot 10^{-3}$	(–)	$8.4\cdot 10^{-4}$	(–)	$1.0\cdot 10^{-2}$	(–)	$3.4\cdot 10^{-2}$	(–)	$8.8\cdot 10^{-3}$	(–)
80	$3.5\cdot 10^{-3}$	(1.5)	$1.7\cdot 10^{-4}$	(2.3)	$3.6\cdot 10^{-3}$	(1.5)	$9.4\cdot 10^{-3}$	(1.9)	$3.2\cdot 10^{-3}$	(1.5)
320	$9.5\cdot 10^{-4}$	(1.9)	$2.4\cdot 10^{-5}$	(2.8)	$9.4\cdot 10^{-4}$	(1.9)	$2.4\cdot 10^{-3}$	(2.0)	$9.2\cdot 10^{-4}$	(1.8)
1280	$2.5\cdot 10^{-4}$	(1.9)	$3.4\cdot 10^{-6}$	(2.8)	$2.5\cdot 10^{-4}$	(1.9)	$6.0\cdot 10^{-4}$	(2.0)	$2.6\cdot 10^{-4}$	(1.8)
5120	$6.5\cdot 10^{-5}$	(1.9)	$4.6\cdot 10^{-7}$	(2.9)	$6.3\cdot 10^{-5}$	(2.0)	$1.5\cdot 10^{-4}$	(2.0)	$6.9\cdot 10^{-5}$	(1.9)
					$k=2$					
20	$2.2\cdot 10^{-3}$	(–)	$1.0\cdot 10^{-4}$	(–)	$1.8\cdot 10^{-3}$	(–)	$3.7\cdot 10^{-3}$	(–)	$1.5\cdot 10^{-3}$	(–)
80	$5.0\cdot 10^{-4}$	(2.1)	$1.1\cdot 10^{-5}$	(3.2)	$3.7\cdot 10^{-4}$	(2.3)	$5.3\cdot 10^{-4}$	(2.8)	$2.8\cdot 10^{-4}$	(2.4)
320	$6.7\cdot 10^{-5}$	(2.9)	$7.7\cdot 10^{-7}$	(3.8)	$5.1\cdot 10^{-5}$	(2.9)	$6.7\cdot 10^{-5}$	(3.0)	$4.1\cdot 10^{-5}$	(2.7)
1280	$8.4\cdot 10^{-6}$	(3.0)	$4.9\cdot 10^{-8}$	(4.0)	$6.4\cdot 10^{-6}$	(3.0)	$8.5\cdot 10^{-6}$	(3.0)	$5.2\cdot 10^{-6}$	(3.0)
5120	$1.0\cdot 10^{-6}$	(3.0)	$3.1\cdot 10^{-9}$	(4.0)	$8.0\cdot 10^{-7}$	(3.0)	$1.1\cdot 10^{-6}$	(3.0)	$6.4\cdot 10^{-7}$	(3.0)
					$k=3$					
20	$4.1\cdot 10^{-4}$	(–)	$1.4\cdot 10^{-5}$	(–)	$2.4\cdot 10^{-4}$	(–)	$7.2\cdot 10^{-5}$	(–)	$2.2\cdot 10^{-4}$	(–)
80	$4.8\cdot 10^{-5}$	(3.1)	$8.4\cdot 10^{-7}$	(4.1)	$2.7\cdot 10^{-5}$	(3.1)	$5.7\cdot 10^{-6}$	(3.7)	$2.6\cdot 10^{-5}$	(3.1)
320	$3.0\cdot 10^{-6}$	(4.0)	$2.6\cdot 10^{-8}$	(5.0)	$1.7\cdot 10^{-6}$	(4.0)	$3.6\cdot 10^{-7}$	(4.0)	$1.7\cdot 10^{-6}$	(3.9)
1280	$1.9\cdot 10^{-7}$	(4.0)	$8.3\cdot 10^{-10}$	(5.0)	$1.1\cdot 10^{-7}$	(4.0)	$2.3\cdot 10^{-8}$	(4.0)	$1.1\cdot 10^{-7}$	(3.9)
5120	$1.2\cdot 10^{-8}$	(4.0)	$2.6\cdot 10^{-11}$	(5.0)	$7.1\cdot 10^{-9}$	(4.0)	$1.4\cdot 10^{-9}$	(4.0)	$7.3\cdot 10^{-9}$	(4.0)

(a) The $d = 2$ example.

$\lvert\mathcal{T}\rvert$	$\lVert\nabla u-\nabla u_h^*\rVert_h$	(eoc)	$\lVert u-u_h^*\rVert$	(eoc)	$\lVert\sigma-\sigma_h\rVert$	(eoc)	$\lVert p-p_h\rVert$	(eoc)	$\lVert\omega-\omega_h\rVert$	(eoc)
					$k=1$					
28	$1.5\cdot 10^{-3}$	(–)	$1.4\cdot 10^{-4}$	(–)	$1.5\cdot 10^{-3}$	(–)	$7.5\cdot 10^{-2}$	(–)	$1.1\cdot 10^{-3}$	(–)
224	$8.1\cdot 10^{-4}$	(0.9)	$5.4\cdot 10^{-5}$	(1.3)	$8.1\cdot 10^{-4}$	(0.8)	$3.1\cdot 10^{-2}$	(1.3)	$6.7\cdot 10^{-4}$	(0.7)
1792	$3.2\cdot 10^{-4}$	(1.4)	$1.3\cdot 10^{-5}$	(2.0)	$3.2\cdot 10^{-4}$	(1.4)	$9.5\cdot 10^{-3}$	(1.7)	$3.2\cdot 10^{-4}$	(1.1)
14336	$9.2\cdot 10^{-5}$	(1.8)	$1.9\cdot 10^{-6}$	(2.8)	$9.0\cdot 10^{-5}$	(1.8)	$2.5\cdot 10^{-3}$	(1.9)	$9.1\cdot 10^{-5}$	(1.8)
114688	$2.4\cdot 10^{-5}$	(1.9)	$2.5\cdot 10^{-7}$	(3.0)	$2.3\cdot 10^{-5}$	(2.0)	$6.4\cdot 10^{-4}$	(2.0)	$2.3\cdot 10^{-5}$	(1.9)
					$k=2$					
28	$5.0\cdot 10^{-4}$	(–)	$4.3\cdot 10^{-5}$	(–)	$5.8\cdot 10^{-4}$	(–)	$6.7\cdot 10^{-3}$	(–)	$4.9\cdot 10^{-4}$	(–)
224	$2.1\cdot 10^{-4}$	(1.3)	$9.7\cdot 10^{-6}$	(2.2)	$1.6\cdot 10^{-4}$	(1.9)	$1.6\cdot 10^{-3}$	(2.1)	$1.4\cdot 10^{-4}$	(1.9)
1792	$5.7\cdot 10^{-5}$	(1.9)	$1.5\cdot 10^{-6}$	(2.7)	$3.9\cdot 10^{-5}$	(2.0)	$2.6\cdot 10^{-4}$	(2.6)	$3.6\cdot 10^{-5}$	(1.9)
14336	$7.9\cdot 10^{-6}$	(2.9)	$1.1\cdot 10^{-7}$	(3.8)	$5.4\cdot 10^{-6}$	(2.8)	$3.5\cdot 10^{-5}$	(2.9)	$5.2\cdot 10^{-6}$	(2.8)
114688	$1.0\cdot 10^{-6}$	(2.9)	$7.0\cdot 10^{-9}$	(3.9)	$7.1\cdot 10^{-7}$	(2.9)	$4.5\cdot 10^{-6}$	(3.0)	$7.0\cdot 10^{-7}$	(2.9)
					$k=3$					
28	$1.8\cdot 10^{-4}$	(–)	$1.3\cdot 10^{-5}$	(–)	$1.7\cdot 10^{-4}$	(–)	$2.4\cdot 10^{-3}$	(–)	$1.3\cdot 10^{-4}$	(–)
224	$5.8\cdot 10^{-5}$	(1.6)	$2.4\cdot 10^{-6}$	(2.4)	$4.4\cdot 10^{-5}$	(1.9)	$2.5\cdot 10^{-4}$	(3.2)	$3.0\cdot 10^{-5}$	(2.1)
1792	$6.8\cdot 10^{-6}$	(3.1)	$1.7\cdot 10^{-7}$	(3.8)	$5.0\cdot 10^{-6}$	(3.2)	$3.0\cdot 10^{-5}$	(3.1)	$3.6\cdot 10^{-6}$	(3.0)
14336	$5.7\cdot 10^{-7}$	(3.6)	$7.3\cdot 10^{-9}$	(4.5)	$4.1\cdot 10^{-7}$	(3.6)	$2.1\cdot 10^{-6}$	(3.9)	$3.0\cdot 10^{-7}$	(3.6)
114688	$4.0\cdot 10^{-8}$	(3.9)	$2.5\cdot 10^{-10}$	(4.9)	$2.8\cdot 10^{-8}$	(3.9)	$1.3\cdot 10^{-7}$	(4.0)	$2.0\cdot 10^{-8}$	(3.9)

(b) The $d = 3$ example.

Table 1: Convergence rates for the postprocessed velocity and all other solution components for $\nu = 10^{-3}$

[18] J. Gopalakrishnan and J. Guzmán, *A second elasticity element using the matrix bubble*, IMA J. Numer. Anal., 32 (2012), pp. 352–372.

[19] J. Gopalakrishnan, P. L. Lederer, and J. Schöberl, *A mass conserving mixed stress formulation for the Stokes equations*, Preprint arXiv:1806.07173, (2018).

[20] J. Guzmán, C.-W. Shu, and F. A. Sequeira, *H(div) conforming and DG methods for incompressible Euler's equations*, IMA Journal of Numerical Analysis, (2016), p. drw054.

[21] J. Könnö and R. Stenberg, *Numerical computations with H(div)-finite elements for the*

Brinkman problem, Computational Geosciences, 16 (2012), pp. 139–158.

[22] P. L. Lederer, C. Lehrenfeld, and J. Schöberl, *Hybrid Discontinuous Galerkin methods with relaxed $H(div)$-conformity for incompressible flows. Part I*, to appear in SIAM journal on numerical analysis (preprint arXiv:1707.02782), (2017).

[23] ———, *Hybrid Discontinuous Galerkin methods with relaxed $H(div)$-conformity for incompressible flows. Part II*, to appear in ESAIM: M2AN (preprint arXiv:1805.06787), (2018).

[24] P. L. Lederer, A. Linke, C. Merdon, and J. Schöberl, *Divergence-free Reconstruction Operators for Pressure-Robust Stokes Discretizations with Continuous Pressure Finite Elements*, SIAM J. Numer. Anal., 55 (2017), pp. 1291–1314.

[25] P. L. Lederer and J. Schöberl, *Polynomial robust stability analysis for $H(div)$-conforming finite elements for the Stokes equations*, IMA Journal of Numerical Analysis, (2017), p. drx051.

[26] C. Lehrenfeld and J. Schöberl, *High order exactly divergence-free hybrid discontinuous galerkin methods for unsteady incompressible flows*, Computer Methods in Applied Mechanics and Engineering, 307 (2016), pp. 339 – 361.

[27] P.-A. Raviart and J. M. Thomas, *A mixed finite element method for 2nd order elliptic problems*, in Mathematical aspects of finite element methods (Proc. Conf., Consiglio Naz. delle Ricerche (C.N.R.), Rome, 1975), Springer, Berlin, 1977, pp. 292–315. Lecture Notes in Math., Vol. 606.

[28] ———, *A mixed finite element method for 2nd order elliptic problems*, (1977), pp. 292–315. Lecture Notes in Math., Vol. 606.

[29] J. Schöberl, *NETGEN An advancing front 2D/3D-mesh generator based on abstract rules*, Computing and Visualization in Science, 1 (1997), pp. 41–52.

[30] J. Schöberl, *C++11 implementation of finite elements in NGSolve*, Tech. Rep. ASC-2014-30, Institute for Analysis and Scientific Computing, September 2014.

[31] R. Stenberg, *A family of mixed finite elements for the elasticity problem*, Numerische Mathematik, 53 (1988), pp. 513–538.