Stabilized Finite Element Formulations for Shells in a Bending Dominated State

Dominique Chapelle, Rolf Stenberg

INRIA
INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Stabilized finite element formulations for shells in a bending dominated state

Dominique Chapelle and Rolf Stenberg

N° 2941

Juillet 1996

THÈME 4

ISSN 0249-6399

INRIA
ROCQUENCOURT

Stabilized finite element formulations for shells in a bending dominated state

Dominique Chapelle * and Rolf Stenberg †

Thème 4 — Simulation et optimisation
de systèmes complexes
Projet Mostra

Rapport de recherche n ° 2941 — Juillet 1996 — 49 pages

Abstract: We consider the design of finite element methods for the Naghdi shell model in the case when the deformation is bending dominated. Two formulations based on stabilizing techniques are introduced and it is proved that they are stable, hence free from locking. The theoretical estimates are confirmed by numerical benchmark studies.

Key-words: finite element methods, stabilized methods, shells, Naghdi model

AMS(MOS) subject classifications (1991 revision): 65N30, 73K15, 73V05

(Résumé : tsvp)

This work was supported in part by the Human Capital and Mobility Program "Shells: Mathematical Modeling and Analysis, Scientific Computing" of the Commission of the European Communities (Contract # ERBCHRXCT940536).

* Laboratoire Central des Ponts et Chaussées – UMR 113 LCPC/CNRS – 2, allée Kepler – 77420 Champs-sur-Marne – e-mail: chapelle@inrets.fr.

† Laboratory for Strength of Materials – Faculty of Mechanical Engineering – Helsinki University of Technology – 02150 Esbo, Finland – e-mail: stenberg@hut.fi.

Méthodes d'éléments finis stabilisées pour les coques à flexion dominante

Résumé : On s'intéresse aux méthodes d'éléments finis appliquées au modèle de coque de Naghdi dans le cas où la déformation est "à flexion dominante", dans le but d'éliminer le verrouillage numérique. On propose deux formulations, bâties sur le principe des méthodes mixtes stabilisées, dont on démontre qu'elles sont stables, ce qui exclut effectivement le verrouillage. Ces résultats théoriques sont corroborés par des études numériques effectuées sur des cas-tests.

1 Introduction

During the last decade great progress has been obtained in the understanding of the locking of finite element methods for various "thin structures". For one-dimensional problems, i.e. beams and arches, it is now completely known how "locking-free" finite element methods should be constructed, cf. e.g. [1, 4]. For plates based on the Reissner-Mindlin theory considerable advances have also been achieved, and by now the problem with the shear locking can safely be claimed to be solved. In this respect we refer to the papers [3, 8, 10, 48] in which the optimal order of convergence is rigorously established for several methods and families of methods.

With regard to shells the present situation is unfortunately far from satisfactory. Even if the literature on the field is voluminous, it is commonly admitted that the elements presently in use are not completely reliable, cf. e.g. the survey [11]. From the viewpoint of numerical analysis, it is not surprising to encounter difficulties with shells. In fact, the general concept of "shell" covers whole families of problems with very marked differences in behaviour depending on e.g. the geometry of the mid-surface, the boundary conditions and the loading. Hence, the goal of developing "the shell element" may be too ambitious, at least at present.

There are two main classes of shell behaviour that can be clearly distinguished: the membrane and bending dominated cases. Mathematically, the membrane case is a singularly perturbated second order elliptic problem when the thickness of the shell is small. For this case the standard finite element method usually works quite well. Difficulties arise in the other case, i.e. when the deformation is bending dominated. Then, the limit problem leads to constraints which give rise to locking if they are exactly imposed in the finite element model.

The numerical analysis of the locking problems for shells is still in its infancy. The papers are in fact so few that they can very well be briefly reviewed here. Pitkäranta was the first to address this question in [40], where he considers hp finite element methods for a cylindrical shell in a bending dominated state. He shows that the standard h-version locks for low order methods, or if the finite element mesh is not aligned with the axis of the cylinder. In the paper it is also shown that the p-version with a fixed mesh is free from locking. Finally, a carefully designed hp-method is shown to be asymptotically convergent both with respect to h and p, but unfortunately only on an aligned rectangular mesh. Pitkäranta analyses the method directly from the (modified) energy expression. The "partial selective reduced integration" introduced in the more recent paper by Arnold and Brezzi [2] is based on the classical approach of writing the problem in mixed form. They first use a splitting of the energy (already used by Pitkäranta in [40]) to write the mixed system in such a form that the "Z-ellipticity" condition of the classical

saddle point theory is avoided [9]. Then they prove the "inf-sup" condition by constructing a Fortin operator. For this, the technique of "bubble functions" is used. In the construction of the Fortin operator, the authors need the assumption that the geometrical parameters (i.e. the fundamental forms and the Christoffel symbols) are piecewise constants, which severely restricts the applicability of their results (namely to circular cylinders). It seems to be non-trivial to extend their analysis to the general case. In [49] Suri studies the approach of Arnold and Brezzi as an *hp*-method for rectangular elements. The analysis of Suri is in the spirit of Pitkäranta, and the modified energy expression is analysed directly without the use of the equivalent mixed method. Also here the assumption of piecewise constant coefficients is used. Let us remark that, intuitively, this assumption seems to be even more restrictive for a "*p*-dominated extension" procedure as the elements then can be quite large and the variation in the geometry big. Let us finally mention the paper by Kirmse [28] in which a spherical surface is considered and a locking free method is designed.

The purpose of the present paper is to explore the application of stabilization techniques to shell problems. These stabilization techniques have earlier been shown to give methods free from locking for several related problems in continuum mechanics such as incompressible elasticity [26, 22, 21], beams [31], arches [30] and Reissner-Mindlin plates [39, 27, 47, 33]. The stabilization technique is quite simple. The equations are first written in a variational mixed form. Then, properly weighted least-squares-type expressions of the equilibrium and constitutive equations are added to the bilinear form. This is, indeed, how we here treat the equations of the Naghdi shell model [35, 5]. For the resulting finite element method we are able to prove the stability, hence the method is free from locking. For our result we do not need any restrictive assumptions on the geometry. Neither do we need any stabilizing bubble degrees of freedom, and we are able to use standard finite element spaces. We should also emphasize that the the auxiliary variables introduced in the mixed stabilized bilinear form are condensed in an implementation which then takes the standard displacement form. Our method is formulated in the *h*-version. In principle, the method can be formulated in a general *hp* setting, but as we extensively use inverse estimates the resulting method would not be uniformly stable with respect to the polynomial degree p. This lack of stability might, however, be compensated for by the better approximability properties of an *hp*-method, especially if a properly refined mesh is used. A general analysis would require that all technical results are verified also with respect to p. As the present paper is already rather technical, we have not tried to perform this.

In other respects our analysis suffers from the same shortcomings as the previous papers on the subject. In order to concentrate on the locking, we give our estimates assuming that the exact solution is sufficiently smooth. As it is known that the shell problems have boundary layers of different length scales [41], this assumption is not very realistic. But as we have a stable method, the treatment of the boundary layers should be done by a proper mesh refinement and this is a problem of approximation theory. The biggest shortcoming for the method we present, and also for the methods that have been mathematically studied in previous works [2, 49, 28], is that they do not perform properly when they are applied to membrane dominated shells. Hence, we are still far from having "the shell element".

An outline of the paper is as follows. In the next section we specify our notation regarding the differential geometry of shell surfaces. In Section 3 we recall the Naghdi shell model and give the equations in the bending dominated case that we focus on. In the next two sections we introduce and analyse two finite element formulations, corresponding to different choices of the stabilization weights. Then, in Section 6, we give numerical results obtained with these methods in benchmark computations. Finally, we close the paper with an appendix in which we prove some results of differential geometry that we have used in our analysis.

2 Shell geometry and notation

We use the classical representation of the shell geometry, as described in [24]. Thus, the shell mid-surface is characterized by a map $\vec{r}$ which is a one-to-one mapping from Ω, an open domain of $I\!R^2$, into $I\!R^3$. $\vec{r}$ will be assumed to be as smooth as required (in practice $\mathcal{C}^3(\bar{\Omega})$). The actual surface we denote by $\mathcal{S}$, i.e. $\mathcal{S} = \vec{r}(\bar{\Omega})$. We consider a shell of uniform thickness t.

Regarding the concepts and quantities of the differential geometry of surfaces that we use in this study, we again refer to [24] for both definitions and notation. However, with a view to making equations more compact, we also make use of the alternate notation by which surface tensors are represented by letters with a number of underbars equal to their order. In particular, a scalar is denoted by a simple letter. With this notation, a tensor is considered as independent of the type (covariant or contravariant) of its representations in the curvilinear coordinate system, and transforming one set of components into another is simply done by using the transformation formulas. Thus we will write

$$\underline{\underline{T}} = T^{\alpha\beta}\,\vec{a}_\alpha \otimes \vec{a}_\beta = T_{\alpha\beta}\,\vec{a}^\alpha \otimes \vec{a}^\beta = T_\alpha^{\cdot\beta}\,\vec{a}^\alpha \otimes \vec{a}_\beta = T_{\cdot\beta}^{\alpha}\,\vec{a}_\alpha \otimes \vec{a}^\beta.$$

Here $\vec{a}_\alpha = \partial\vec{r}/\partial\xi_\alpha$, $\alpha = 1,2$, are the covariant surface base vectors and $\vec{a}^\alpha$, $\alpha = 1,2$, is the contravariant base. A single dot will denote the simple contraction operator:

$$\underline{u} \cdot \underline{v} \stackrel{\text{def}}{=} u^\alpha v_\alpha = u_\alpha v^\alpha,$$

$$\underline{\underline{T}} \cdot \underline{v} \stackrel{\text{def}}{=} T^{\alpha\beta} v_\beta \, \vec{a}_\alpha = T_{\alpha\beta} v^\beta \, \vec{a}^\alpha = \ldots,$$

$$\underline{\underline{T}} \cdot \underline{\underline{X}} \stackrel{\text{def}}{=} T^{\alpha\beta} X_{\beta\gamma} \, \vec{a}_\alpha \otimes \vec{a}^\gamma = T_{\alpha\beta} X^{\beta\gamma} \, \vec{a}^\alpha \otimes \vec{a}_\gamma = \ldots,$$

whereas a colon denotes a double contraction:

$$\underline{\underline{T}} : \underline{\underline{X}} \stackrel{\text{def}}{=} T^{\alpha\beta} X_{\beta\alpha}.$$

The metric tensor for the surface we denote by $\underline{\underline{a}}$ and the second and third fundamental forms are denoted by $\underline{\underline{b}}$ and $\underline{\underline{c}}$ $(= \underline{\underline{b}} : \underline{\underline{b}})$, respectively.

For a second-order tensor, we define the transposition and symmetry operators as follows

$${}^t\underline{\underline{T}} \stackrel{\text{def}}{=} T^{\beta\alpha} \, \vec{a}_\alpha \otimes \vec{a}_\beta,$$

$$\underline{\underline{\Sigma}}(\underline{\underline{T}}) \stackrel{\text{def}}{=} \frac{1}{2}(\underline{\underline{T}} + {}^t\underline{\underline{T}}).$$

We next consider differential operators. The gradient of a tensor is obtained by taking the covariant derivative and adding one index

$$\underline{\nabla} w \stackrel{\text{def}}{=} w_{|\alpha} \, \vec{a}^\alpha = w_{,\alpha} \, \vec{a}^\alpha,$$

$$\underline{\underline{\nabla}} \, \underline{u} \stackrel{\text{def}}{=} u_{\alpha|\beta} \, \vec{a}^\alpha \otimes \vec{a}^\beta,$$

and the divergence by contracting the gradient on the last two indices

$$\operatorname{div} \underline{u} \stackrel{\text{def}}{=} u^\alpha{}_{|\alpha},$$

$$\underline{\operatorname{div}} \, \underline{\underline{T}} \stackrel{\text{def}}{=} T^{\alpha\beta}{}_{|\beta} \, \vec{a}_\alpha.$$

For a surface integral over ω, a subdomain of Ω, we use the following compact notation

$$\int_\omega w \, dS \stackrel{\text{def}}{=} \int_\omega w(\xi_1, \xi_2) \sqrt{a} \, d\xi_1 d\xi_2.$$

We also write

$$\int_C w \, ds \stackrel{\text{def}}{=} \int_C w(\xi_1, \xi_2) \, ds$$

for an integral along a curve C, with $(ds)^2 = a_{\alpha\beta} d\xi_\alpha d\xi_\beta$. With this notation, Gauss theorem reads

$$\int_\omega \operatorname{div} \underline{u} \, dS = \int_{\partial\omega} \underline{u} \cdot \underline{\nu} \, ds, \tag{2.1}$$

where $\underline{\nu}$ is the unit outward normal, in the tangential plane, to the boundary $\vec{r}(\partial\omega)$. From this the following Green formulas are derived

$$\int_\omega \underline{u} \cdot \underline{\nabla} w \, dS = \int_{\partial\omega} (\underline{u} \cdot \underline{\nu}) w \, ds - \int_\omega (\operatorname{div} \underline{u}) w \, dS, \tag{2.2}$$

$$\int_\omega {}^t\underline{\underline{X}} : \underline{\underline{\nabla}}\, \underline{u} \, dS = \int_{\partial\omega} (\underline{\underline{X}} \cdot \underline{\nu}) \cdot \underline{u} \, ds - \int_\omega (\underline{\operatorname{div}}\, \underline{\underline{X}}) \cdot \underline{u} \, dS. \tag{2.3}$$

For our analysis it will be convenient to use special norms for Sobolev spaces of tensors. If two zero-order tensors v and w are in $L^2(\omega)$ we define the "intrinsic" inner-product and norm

$$\langle v, w \rangle_\omega \stackrel{\text{def}}{=} \int_\omega v w \, dS,$$

$$\|w\|_{0,\omega} \stackrel{\text{def}}{=} \langle w, w \rangle_\omega^{\frac{1}{2}}.$$

This new norm is equivalent to the regular $L^2(\omega)$-norm, cf. [43]. We also extend this definition to higher order tensors and write

$$\langle \underline{u}, \underline{v} \rangle_\omega \stackrel{\text{def}}{=} \int_\omega \underline{u} \cdot \underline{v} \, dS,$$

$$\left\langle \underline{\underline{T}}, \underline{\underline{X}} \right\rangle_\omega \stackrel{\text{def}}{=} \int_\omega {}^t\underline{\underline{T}} : \underline{\underline{X}} \, dS.$$

We also denote the corresponding norms by $\|\cdot\|_{0,\omega}$ and these can be shown to be equivalent to the regular $L^2(\omega)$-norm of any of the representations of these tensors in covariant or contravariant components [43]. Finally, we introduce new $H^1(\omega)$-norms through

$$\|w\|_{1,\omega}^2 \stackrel{\text{def}}{=} \|w\|_{0,\omega}^2 + \|\underline{\nabla} w\|_{0,\omega}^2,$$

$$\|\underline{u}\|_{1,\omega}^2 \stackrel{\text{def}}{=} \|\underline{u}\|_{0,\omega}^2 + \|\underline{\underline{\nabla}}\, \underline{u}\|_{0,\omega}^2.$$

Recalling the relationship between covariant and regular derivatives, it is clear that this new norm is again equivalent to the standard $H^1(\omega)$-norm of the the tensors in any representation. Using higher-order covariant differentiation, there is no difficulty in similarly defining intrinsic H^k-norms ($k \geq 2$) and establishing the corresponding equivalence properties. For all norm and inner-product signs, we will omit the domain subscript when the domain considered is Ω itself.

3 The variational shell model

The shell model we consider, the so-called Naghdi model, is of the Reissner-Mindlin type, i.e. it includes the effect of shear deformation, cf. [35, 5, 50, 52, 19] and the references therein. The unknowns are $\vec{u} = (\underline{u}, u_3)$, the displacement at each point of the mid-surface decomposed into tangential and transverse parts (respectively a first-order and zero-order surface tensor), and $\underline{\theta}$ as the first-order tensor representing

the rotation of a fiber normal to the mid-surface $\mathcal{S}$ in the undeformed configuration. We let $\partial\Omega = \Gamma_0 \cup \Gamma_1$ and we assume that the boundary conditions on $\partial\mathcal{S} = \vec{r}(\Gamma_0) \cup \vec{r}(\Gamma_1)$ are given so that $\vec{r}(\Gamma_0)$ is the part of the boundary along which the shell is fully clamped (i.e. $\vec{u} = \vec{0}$, $\underline{\theta} = \underline{0}$) and $\vec{r}(\Gamma_1)$ the part where all displacements are left free. The problem is posed in the domain Ω and we define the displacement space $\mathcal{U}$ by

$$\mathcal{U} \stackrel{\text{def}}{=} \left\{ (\vec{v}, \underline{\eta}) \in [H^1(\Omega)]^3 \times [H^1(\Omega)]^2 \,|\, \vec{v} = \vec{0} \text{ and } \underline{\eta} = \underline{0} \text{ on } \Gamma_0 \right\}.$$

The variational formulation then reads [5]: find $(\vec{u}, \underline{\theta}) \in \mathcal{U}$ such

$$\begin{aligned} \frac{t^3}{12} \int_\Omega \underline{\underline{\kappa}}(\vec{u}, \underline{\theta}) : \underline{\underline{\underline{\underline{E}}}} : \underline{\underline{\kappa}}(\vec{v}, \underline{\eta}) \, dS + t \int_\Omega \underline{\underline{\varepsilon}}(\vec{u}) : \underline{\underline{\underline{\underline{E}}}} : \underline{\underline{\varepsilon}}(\vec{v}) \, dS \\ + t \int_\Omega \underline{\gamma}(\vec{u}, \underline{\theta}) \cdot \underline{\underline{G}} \cdot \underline{\gamma}(\vec{v}, \underline{\eta}) \, dS = \int_\Omega \vec{g} \cdot \vec{v} \, dS \qquad \forall (\vec{v}, \underline{\eta}) \in \mathcal{U}. \end{aligned} \tag{3.1}$$

Here the material properties are given by the two tensors

$$E^{\alpha\beta\lambda\mu} = \frac{E}{2(1+\nu)} \left[a^{\alpha\lambda} a^{\beta\mu} + a^{\alpha\mu} a^{\beta\lambda} + \frac{2\nu}{1-\nu} a^{\alpha\beta} a^{\lambda\mu} \right],$$

$$G^{\alpha\beta} = \frac{E}{2(1+\nu)} a^{\alpha\beta},$$

and $\underline{\underline{\kappa}}$, $\underline{\underline{\varepsilon}}$, $\underline{\gamma}$ denote the bending, membrane and shear strain tensors, respectively:

$$\underline{\underline{\kappa}}(\vec{u}, \underline{\theta}) \stackrel{\text{def}}{=} \underline{\underline{\Sigma}}(\underline{\underline{\nabla}} \, \underline{\theta}) - \underline{\underline{\Sigma}}(\underline{\underline{b}} \cdot \underline{\underline{\nabla}} \, \underline{u}) + \underline{\underline{c}} \, u_3,$$

$$\underline{\underline{\varepsilon}}(\vec{u}) \stackrel{\text{def}}{=} \underline{\underline{\Sigma}}(\underline{\underline{\nabla}} \, \underline{u}) - \underline{\underline{b}} \, u_3,$$

$$\underline{\gamma}(\vec{u}, \underline{\theta}) \stackrel{\text{def}}{=} \underline{\nabla} u_3 + \underline{\underline{b}} \cdot \underline{u} + \underline{\theta}.$$

We also recall that $\underline{\underline{b}}$ and $\underline{\underline{c}}$ are the symmetric tensors corresponding to the second and third fundamental forms of the surface, respectively.

We also write equation (3.1) in the shorter form:

$$t^3 A(\vec{u}, \underline{\theta}; \vec{v}, \underline{\eta}) + t D(\vec{u}, \underline{\theta}; \vec{v}, \underline{\eta}) = (\vec{g}, \vec{v}), \tag{3.2}$$

with

$$A(\vec{u}, \underline{\theta}; \vec{v}, \underline{\eta}) \stackrel{\text{def}}{=} \int_\Omega \underline{\underline{\kappa}}(\vec{u}, \underline{\theta}) : \underline{\underline{\underline{\underline{E}}}} : \underline{\underline{\kappa}}(\vec{v}, \underline{\eta}) \, dS \tag{3.3}$$

and

$$D(\vec{u}, \underline{\theta}; \vec{v}, \underline{\eta}) \stackrel{\text{def}}{=} \int_\Omega \underline{\underline{\varepsilon}}(\vec{u}) : \underline{\underline{\underline{\underline{E}}}} : \underline{\underline{\varepsilon}}(\vec{v}) \, dS + \int_\Omega \underline{\gamma}(\vec{u}, \underline{\theta}) \cdot \underline{\underline{G}} \cdot \underline{\gamma}(\vec{v}, \underline{\eta}) \, dS. \tag{3.4}$$

Implicit in the use of a shell model is the assumption that the shell is "thin". Hence, it is essential to study how the solution behaves in the limit when the thickness $t \to 0$. This limit behavior is very different depending on the geometry

of the shell and the boundary conditions and depends in a crucial way (cf. [44, 45, 15, 16, 36, 37]) on the subspace

$$\mathcal{U}^0 \stackrel{\text{def}}{=} \left\{(\vec{v},\underline{\eta}) \in \mathcal{U} | D(\vec{v},\underline{\eta};\vec{v},\underline{\eta}) = 0\right\}$$

consisting of the displacement fields with vanishing shear and membrane strains.

In this paper we only consider the case when the shell is said to be in a *bending dominated state of deformation*. This is the case when (cf. [44, 16, 37])

$$\mathcal{U}^0 \neq \{(\vec{0},\underline{0})\}.$$

Then the limit problem is obtained by assuming the loading to be given by $\vec{g} = t^3\vec{f}$ with $\vec{f}$ independent of the thickness t. In the sequel we therefore study the following problem.

$\mathcal{P}_t$: *Find* $(\vec{u},\underline{\theta}) \in \mathcal{U}$ *such that*

$$A(\vec{u},\underline{\theta};\vec{v},\underline{\eta}) + t^{-2}D(\vec{u},\underline{\theta};\vec{v},\underline{\eta}) = F(\vec{v}) \qquad \forall(\vec{v},\underline{\eta}) \in \mathcal{U}. \tag{3.5}$$

Here we introduced the notation $F(\cdot) \stackrel{\text{def}}{=} (\vec{f},\cdot)$.

From [6] and [18] we know that, for any $t > 0$, $\mathcal{P}_t$ has a unique solution, provided that the distributed force field $\vec{f}$ is in an appropriate space, say $[L^2(\Omega)]^3$. Furthermore, as t tends to zero, $\mathcal{P}_t$ can be seen as a standard penalty problem. The limit problem $\mathcal{P}_0$ is defined as:

$\mathcal{P}_0$: *Find* $(\vec{u}_0,\underline{\theta}_0)$ *in* $\mathcal{U}^0$ *such that*

$$A(\vec{u}_0,\underline{\theta}_0;\vec{v},\underline{\eta}) = F(\vec{v}) \qquad \forall(\vec{v},\underline{\eta}) \in \mathcal{U}^0. \tag{3.6}$$

The term "limit problem" is justified since, as $t \to 0$, it can be shown that the solution $(\vec{u},\underline{\theta})$ of $\mathcal{P}_t$ converges strongly in $\mathcal{U}$ to the solution $(\vec{u}_0,\underline{\theta}_0)$ of $\mathcal{P}_0$ (cf. [13, 40]).

In the approximation of plate problems it has often turned out to be advantageous to use a mixed formulation, cf. [9, 10, 20]. This is the approach that we use here as well. Hence, we define as new unknowns stress variables dual to the strains in the variational formulation (3.5). By $\underline{\underline{n}}$ we denote the symmetric membrane force tensor, Here we depart from the usual notation, cf. Remark 3.1 below. The shear force we denote by $\underline{q}$. These are connected to the membrane and shear strains through the constitutive equations

$$\underline{\underline{n}} = \frac{1}{t^2}\underline{\underline{\underline{\underline{E}}}} : \underline{\underline{\varepsilon}}(\vec{u}), \tag{3.7}$$

$$\underline{q} = \frac{1}{t^2}\underline{\underline{\underline{G}}} \cdot \underline{\gamma}(\vec{u},\underline{\theta}). \tag{3.8}$$

The symmetric bending moment tensor is not taken as an independent unknown, but we use the following abbreviation based on the bending constitutive equation:

$$\underline{\underline{m}}(\vec{u},\underline{\theta}) \stackrel{\text{def}}{=} \frac{1}{12}\underline{\underline{\underline{\underline{E}}}} : \underline{\underline{\kappa}}(\vec{u},\underline{\theta}). \tag{3.9}$$

By introducing

$$(\underline{\underline{n}},\underline{q}) \in \mathcal{Q} \stackrel{\text{def}}{=} \{\, \underline{\underline{p}} \in [L^2(\Omega)]^{2\times 2} \mid {}^t\underline{\underline{p}} = \underline{\underline{p}} \,\} \times [L^2(\Omega)]^2$$

as new unknowns and by writing the constitutive equations (3.8) and (3.9) in a weak form, we obtain an equivalent "mixed" formulation of the problem:

$\mathcal{M}_t$: *Find* $(\vec{u},\underline{\theta},\underline{\underline{n}},\underline{q}) \in \mathcal{U}\times\mathcal{Q}$ *such that*

$$B(\vec{u},\underline{\theta},\underline{\underline{n}},\underline{q};\vec{v},\underline{\eta},\underline{\underline{p}},\underline{r}) = F(\vec{v}) \quad \forall (\vec{v},\underline{\eta},\underline{\underline{p}},\underline{r}) \in \mathcal{U}\times\mathcal{Q}, \tag{3.10}$$

with the new bilinear forms defined through

$$\begin{aligned} B(\vec{u},\underline{\theta},\underline{\underline{n}},\underline{q};\vec{v},\underline{\eta},\underline{\underline{p}},\underline{r}) \stackrel{\text{def}}{=}\ & A(\vec{u},\underline{\theta};\vec{v},\underline{\eta}) + M(\vec{v},\underline{\eta};\underline{\underline{n}},\underline{q}) + M(\vec{u},\underline{\theta};\underline{\underline{p}},\underline{r}) \\ & - N(\underline{\underline{n}},\underline{q};,\underline{\underline{p}},\underline{r}), \end{aligned}$$

$$M(\vec{v},\underline{\eta};\underline{\underline{p}},\underline{r}) \stackrel{\text{def}}{=} \int_\Omega (\underline{\underline{\varepsilon}}(\vec{v}) : \underline{\underline{p}} + \underline{\gamma}(\vec{v},\underline{\eta})\cdot\underline{r})\, dS$$

and

$$N(\underline{\underline{n}},\underline{q};,\underline{\underline{p}},\underline{r}) \stackrel{\text{def}}{=} t^2 \int_\Omega (\underline{\underline{n}} : \underline{\underline{\underline{\underline{\breve{E}}}}} : \underline{\underline{p}} + \underline{q}\cdot\underline{\underline{\breve{G}}}\cdot\underline{r})\, dS.$$

Here we use the notation

$$\breve{E}_{\alpha\beta\lambda\mu} = \frac{1+\nu}{2E}\left[a_{\alpha\lambda}a_{\beta\mu} + a_{\alpha\mu}a_{\beta\lambda} - \frac{2\nu}{1+\nu}a_{\alpha\beta}a_{\lambda\mu}\right],$$

$$\breve{G}_{\alpha\beta} = \frac{2(1+\nu)}{E}a_{\alpha\beta}.$$

The second order tensor $\underline{\underline{\breve{G}}}$ is the inverse of $\underline{\underline{G}}$ (i.e. $\underline{\underline{G}}\cdot\underline{\underline{\breve{G}}} = \underline{\underline{\breve{G}}}\cdot\underline{\underline{G}} = \underline{\underline{a}}$), whereas $\underline{\underline{\underline{\underline{\breve{E}}}}}$ is such that, for any two symmetric tensors $\underline{\underline{T}}$ and $\underline{\underline{X}}$:

$$\underline{\underline{T}} = \underline{\underline{\underline{\underline{E}}}} : \underline{\underline{X}} \quad \Longleftrightarrow \quad \underline{\underline{X}} = \underline{\underline{\underline{\underline{\breve{E}}}}} : \underline{\underline{T}}. \tag{3.11}$$

Finally, we recall that using the symmetry of the second order tensors, and the Green formulas (2.2) and (2.3), the following integration by parts formulas are obtained

$$\begin{aligned} \int_\omega \underline{\underline{m}} : \underline{\underline{\kappa}}(\vec{v},\underline{\eta})\, dS = & -\int_\omega \underline{\operatorname{div}}\,\underline{\underline{m}}\cdot\underline{\eta}\, dS + \int_\omega \underline{\operatorname{div}}\,(\underline{\underline{b}}\cdot\underline{\underline{m}})\cdot\underline{v}\, dS \\ & + \int_\omega (\underline{\underline{b}} : (\underline{\underline{b}}\cdot\underline{\underline{m}}))\, v_3\, dS + \int_{\partial\omega} (\underline{\underline{m}}\cdot\underline{\nu})\cdot\underline{\eta}\, ds - \int_{\partial\omega} ((\underline{\underline{b}}\cdot\underline{\underline{m}})\cdot\underline{\nu})\cdot\underline{v}\, ds, \end{aligned} \tag{3.12}$$

$$\int_\omega \underline{\underline{n}} : \underline{\underline{\varepsilon}}(\vec{v})\, dS = -\int_\omega \underline{\mathrm{div}}\ \underline{\underline{n}} \cdot \underline{v}\, dS - \int_\omega (\underline{\underline{b}} : \underline{\underline{n}})\, v_3\, dS + \int_{\partial\omega} (\underline{\underline{n}} \cdot \underline{\nu}) \cdot \underline{v}\, ds \tag{3.13}$$

and

$$\int_\omega \underline{q} : \underline{\gamma}(\vec{v}, \underline{\eta})\, dS = \int_\omega \underline{q} \cdot \underline{\eta}\, dS + \int_\omega (\underline{\underline{b}} \cdot \underline{q}) \cdot \underline{v}\, dS - \int_\omega \mathrm{div}\, \underline{q}\, v_3\, dS + \int_{\partial\omega} (\underline{q} \cdot \underline{\nu})\, v_3\, ds. \tag{3.14}$$

Collecting these formulas gives

$$\begin{aligned}
&\int_\omega \underline{\underline{m}} : \underline{\underline{\kappa}}(\vec{v}, \underline{\eta})\, dS + \int_\omega \underline{\underline{n}} : \underline{\underline{\varepsilon}}(\vec{v})\, dS + \int_\omega \underline{q} : \underline{\gamma}(\vec{v}, \underline{\eta})\, dS = \\
&\quad - \int_\omega (\underline{\mathrm{div}}\ \underline{\underline{m}} - \underline{q}) \cdot \underline{\eta}\, dS - \int_\omega (\underline{\mathrm{div}}\ (\underline{\underline{n}} - \underline{\underline{b}} \cdot \underline{\underline{m}}) - \underline{\underline{b}} \cdot \underline{q}) \cdot \underline{v}\, dS \\
&\quad - \int_\omega (\mathrm{div}\, \underline{q} + \underline{\underline{b}} : (\underline{\underline{n}} - \underline{\underline{b}} \cdot \underline{\underline{m}}))\, v_3\, dS + \int_{\partial\omega} (\underline{\underline{m}} \cdot \underline{\nu}) \cdot \underline{\eta}\, ds \\
&\quad + \int_{\partial\omega} ((\underline{\underline{n}} - \underline{\underline{b}} \cdot \underline{\underline{m}}) \cdot \underline{\nu}) \cdot \underline{v}\, ds + \int_{\partial\omega} (\underline{q} \cdot \underline{\nu})\, v_3\, ds.
\end{aligned} \tag{3.15}$$

Hence, recalling the constitutive equations (3.7)–(3.9), problem $\mathcal{P}_t$ gives the following equations of equilibrium

$$\underline{\mathrm{div}}\ \underline{\underline{m}} - \underline{q} = 0, \tag{3.16}$$

$$\underline{\mathrm{div}}\ (\underline{\underline{n}} - \underline{\underline{b}} \cdot \underline{\underline{m}}) - \underline{\underline{b}} \cdot \underline{q} + \underline{f} = 0, \tag{3.17}$$

$$\mathrm{div}\, \underline{q} + \underline{\underline{b}} : (\underline{\underline{n}} - \underline{\underline{b}} \cdot \underline{\underline{m}}) + f_3 = 0, \tag{3.18}$$

in the domain Ω, and the natural boundary conditions

$$\underline{\underline{m}} \cdot \underline{\nu} = 0, \tag{3.19}$$

$$(\underline{\underline{n}} - \underline{\underline{b}} \cdot \underline{\underline{m}}) \cdot \underline{\nu} = 0, \tag{3.20}$$

$$\underline{q} \cdot \underline{\nu} = 0, \tag{3.21}$$

on the free boundary Γ_1.

Remark 3.1 In the above equations the quantity

$$\underline{\underline{\tilde{n}}} = \underline{\underline{n}} - \underline{\underline{b}} \cdot \underline{\underline{m}} \tag{3.22}$$

is the tensor obtained by integrating the three-dimensional membrane stresses over the shell thickness, cf. e.g. [29, 24, 35, 52]. Usually this is called the membrane force. For designing a stable method it seems more practical to introduce the variable $\underline{\underline{n}}$, and not $\underline{\underline{\tilde{n}}}$, as an new unknown. For simplicity we refer to $\underline{\underline{n}}$ as the membrane force. ■

4 A first approximation scheme

We use a finite element partitioning $\mathcal{C}_h$ of $\bar{\Omega}$ into straight-sided triangles or quadrilaterals. Naturally, the partitioning is assumed to satisfy the usual regularity and compatibility conditions, cf. [14, 7, 42]. The diameter of an element $K \in \mathcal{C}_h$ we denote by h_K, and we let $h = \max_{K \in \mathcal{C}_h} h_K$. By Γ_h we denote the collection of edges in the mesh, and by h_E the length of an edge $E \in \Gamma_h$. The letters C and c are henceforth used to denote generic strictly positive constants, independent of both t and h, which are allowed to take different values at different occurrences except when appearing with indices.

Let $\mathcal{U}^h$ and $\mathcal{Q}^h$ be finite element subspaces of $\mathcal{U}$ and $\mathcal{Q}$, respectively. A mixed finite element method of the "classical" saddle point type would be based on the formulation $\mathcal{M}_t$, viz.

$\mathcal{M}_t^h$: *Find* $(\vec{u}^h, \underline{\theta}^h, \underline{\underline{n}}^h, \underline{q}^h) \in \mathcal{U}^h \times \mathcal{Q}^h$ *such that*

$$B(\vec{u}^h, \underline{\theta}^h, \underline{\underline{n}}^h, \underline{q}^h; \vec{v}, \underline{\eta}, \underline{\underline{p}}, \underline{r}) = F(\vec{v}, \underline{\eta}, \underline{\underline{p}}, \underline{r}) \quad \forall (\vec{v}, \underline{\eta}, \underline{\underline{p}}, \underline{r}) \in \mathcal{U}^h \times \mathcal{Q}^h. \tag{4.1}$$

It appears difficult to design a stable finite element method directly based on this formulation. Hence, we follow an approach that has turned out to be fruitful for the related beam, arch and plate problems, cf. [22, 21, 47]. The approximation scheme we propose is derived from (4.1) by adding weighted least-squares-type terms of the equilibrium equations (3.16)–(3.18) and the constitutive relations (3.7)–(3.8). In order to simplify the presentation of our method, we introduce the weighted inner-product

$$\langle v, w \rangle_h \stackrel{\text{def}}{=} \sum_{K \in \mathcal{C}_h} h_K^2 \, \langle v, w \rangle_K \,, \tag{4.2}$$

with analogous definitions for higher order tensors. Further, in accordance with (3.9) and (3.22) we introduce the notation

$$\tilde{\underline{\underline{n}}}(\vec{v}, \underline{\eta}, \underline{\underline{p}}) \stackrel{\text{def}}{=} \underline{\underline{p}} - \underline{\underline{b}} \cdot \underline{\underline{m}}(\vec{v}, \underline{\eta}). \tag{4.3}$$

Our first stabilized method will now be defined as follows:

$\mathcal{S}_t^h$: *Find* $(\vec{u}^h, \underline{\theta}^h, \underline{\underline{n}}^h, \underline{q}^h) \in \mathcal{U}^h \times \mathcal{Q}^h$ *such that*

$$B_h(\vec{u}^h, \underline{\theta}^h, \underline{\underline{n}}^h, \underline{q}^h; \vec{v}, \underline{\eta}, \underline{\underline{p}}, \underline{r}) = F_h(\vec{v}, \underline{\eta}, \underline{\underline{p}}, \underline{r}) \quad \forall (\vec{v}, \underline{\eta}, \underline{\underline{p}}, \underline{r}) \in \mathcal{U}^h \times \mathcal{Q}^h, \tag{4.4}$$

where

$$\begin{aligned} B_h(\vec{w}, \underline{\tau}, \underline{\underline{k}}, \underline{s}; \vec{v}, \underline{\eta}, \underline{\underline{p}}, \underline{r}) \stackrel{\text{def}}{=} & \; B(\vec{w}, \underline{\tau}, \underline{\underline{k}}, \underline{s}; \vec{v}, \underline{\eta}, \underline{\underline{p}}, \underline{r}) + S_h^e(\vec{w}, \underline{\tau}, \underline{\underline{k}}, \underline{s}; \vec{v}, \underline{\eta}, \underline{\underline{p}}, \underline{r}) \\ & + S_h^c(\vec{w}, \underline{\tau}, \underline{\underline{k}}, \underline{s}; \vec{v}, \underline{\eta}, \underline{\underline{p}}, \underline{r}), \end{aligned} \tag{4.5}$$

with a stabilizing term originating from the equilibrium equations

$$\begin{aligned} S_h^e(\vec{w},\underline{\tau},\underline{\underline{k}},\underline{s};\vec{v},\underline{\eta},\underline{\underline{p}},\underline{r}) & \\ \stackrel{\text{def}}{=} & -\alpha_1 \left\langle \underline{\text{div}}\ \underline{\underline{m}}(\vec{w},\underline{\tau}) - \underline{s}, \underline{\text{div}}\ \underline{\underline{m}}(\vec{v},\underline{\eta}) - \underline{r} \right\rangle_h \\ & -\alpha_2 \left\langle \underline{\text{div}}\ \underline{\tilde{\underline{n}}}(\vec{w},\underline{\tau},\underline{\underline{k}}) - \underline{b}\cdot\underline{s}, \underline{\text{div}}\ \underline{\tilde{\underline{n}}}(\vec{v},\underline{\eta},\underline{\underline{p}}) - \underline{b}\cdot\underline{r} \right\rangle_h \\ & -\alpha_3 \left\langle \text{div}\,\underline{s} + \underline{\underline{b}} : \underline{\tilde{\underline{n}}}(\vec{w},\underline{\tau},\underline{\underline{k}}), \text{div}\,\underline{r} + \underline{\underline{b}} : \underline{\tilde{\underline{n}}}(\vec{v},\underline{\eta},\underline{\underline{p}}) \right\rangle_h \end{aligned} \tag{4.6}$$

and one from the constitutive equations

$$\begin{aligned} S_h^c(\vec{w},\underline{\tau},\underline{\underline{k}},\underline{s};\vec{v},\underline{\eta},\underline{\underline{p}},\underline{r}) & \\ \stackrel{\text{def}}{=} & \alpha_4 \int_\Omega \left[\underline{\underline{\varepsilon}}(\vec{w}) - t^2 \breve{\underline{\underline{\underline{\underline{E}}}}} : \underline{\underline{k}} \right] : \underline{\underline{\underline{\underline{E}}}} : \left[\underline{\underline{\varepsilon}}(\vec{v}) - t^2 \breve{\underline{\underline{\underline{\underline{E}}}}} : \underline{\underline{p}} \right] \, dS \\ & + \alpha_5 \int_\Omega \left[\underline{\gamma}(\vec{w},\underline{\tau}) - t^2 \breve{\underline{\underline{G}}} \cdot \underline{s} \right] \cdot \underline{\underline{G}} \cdot \left[\underline{\gamma}(\vec{v},\underline{\eta}) - t^2 \breve{\underline{\underline{G}}} \cdot \underline{r} \right] \, dS. \end{aligned} \tag{4.7}$$

The right hand side is defined as

$$\begin{aligned} F_h(\vec{v},\underline{\eta},\underline{\underline{p}},\underline{r}) \stackrel{\text{def}}{=} & F(\vec{v}) + \alpha_2 \left\langle \underline{f}, \underline{\text{div}}\ \underline{\tilde{\underline{n}}}(\vec{v},\underline{\eta},\underline{\underline{p}}) - \underline{b}\cdot\underline{r} \right\rangle_h \\ & + \alpha_3 \left\langle f_3, \text{div}\,\underline{r} + \underline{\underline{b}} : \underline{\tilde{\underline{n}}}(\vec{v},\underline{\eta},\underline{\underline{p}}) \right\rangle_h . \end{aligned} \tag{4.8}$$

The norm used for the displacement is the modified $H^1(\Omega)$-norm that we previously defined, i.e.

$$\|\vec{v},\underline{\eta}\|_1 \stackrel{\text{def}}{=} \left(\|\underline{v}\|_1^2 + \|v_3\|_1^2 + \|\underline{\theta}\|_1 \right)^{1/2}. \tag{4.9}$$

For the stresses, we use a discrete norm defined by

$$\|\underline{\underline{p}},\underline{r}\|_{t,h} \stackrel{\text{def}}{=} \left(t^2 \|\underline{\underline{p}},\underline{r}\|_0^2 + |\underline{\underline{p}},\underline{r}|_I^2 + |\underline{\underline{p}},\underline{r}|_J^2 \right)^{1/2}, \tag{4.10}$$

with "interior" terms

$$|\underline{\underline{p}},\underline{r}|_I \stackrel{\text{def}}{=} \Big(\sum_{K\in\mathcal{C}_h} h_K^2 \|\underline{r}\|_{0,K}^2 + \sum_{K\in\mathcal{C}_h} h_K^2 \|\underline{\text{div}}\ \underline{\underline{p}}\|_{0,K}^2 + \sum_{K\in\mathcal{C}_h} h_K^2 \|\text{div}\,\underline{r} + \underline{\underline{b}} : \underline{\underline{p}}\|_{0,K}^2 \Big)^{1/2}, \tag{4.11}$$

and "jump" terms

$$|\underline{\underline{p}},\underline{r}|_J \stackrel{\text{def}}{=} \Big(\sum_{E\in\Gamma_h\setminus\Gamma_0} h_E \int_E \left[[\![\underline{\underline{p}}\cdot\underline{\nu}]\!]^2 + [\![\underline{r}\cdot\underline{\nu}]\!]^2 \right] \, ds \Big)^{1/2}, \tag{4.12}$$

where $[\![\cdot]\!]$ stands for the jump of the quantity if the edge is in the interior, and for the quantity itself if it is on Γ_1. Finally, we let

$$\|\vec{v},\underline{\eta},\underline{\underline{p}},\underline{r}\|_{t,h} \stackrel{\text{def}}{=} \left(\|\vec{v},\underline{\eta}\|_1^2 + \|\underline{\underline{p}},\underline{r}\|_{t,h}^2 \right)^{1/2}. \tag{4.13}$$

We begin our analysis by noting the consistency of formulation $\mathcal{S}_t^h$ with respect to the continuous problem $\mathcal{M}_t$.

Lemma 4.1 *Suppose that* $\vec{f} \in [L^2(\Omega)]^3$. *Then the solution* $(\vec{u}, \underline{\theta}, \underline{\underline{n}}, \underline{q}) \in \mathcal{U} \times \mathcal{Q}$ *of* $\mathcal{M}_t$ *satisfies*

$$B_h(\vec{u}, \underline{\theta}, \underline{\underline{n}}, \underline{q}; \vec{v}, \underline{\eta}, \underline{\underline{p}}, \underline{r}) = F_h(\vec{v}, \underline{\eta}, \underline{\underline{p}}, \underline{r}) \quad \forall (\vec{v}, \underline{\eta}, \underline{\underline{p}}, \underline{r}) \in \mathcal{U}^h \times \mathcal{Q}^h. \tag{4.14}$$

Proof: First we note that, due to the assumption $\vec{f} \in [L^2(\Omega)]^3$, the quantity $B_h(\vec{u}, \underline{\theta}, \underline{\underline{n}}, \underline{q}; \vec{v}, \underline{\eta}, \underline{\underline{p}}, \underline{r})$ is well defined for all $(\vec{v}, \underline{\eta}, \underline{\underline{p}}, \underline{r}) \in \mathcal{U}^h \times \mathcal{Q}^h$. Therefore, using the continuous formulation $\mathcal{M}_t$, the equilibrium equations (3.16)–(3.18) and the constitutive equations (3.7)–(3.8), we directly obtain

$$\begin{aligned}
B_h(\vec{u}, \underline{\theta}, \underline{\underline{n}}, \underline{q}; \vec{v}, \underline{\eta}, \underline{\underline{p}}, \underline{r}) &= B(\vec{u}, \underline{\theta}, \underline{\underline{n}}, \underline{q}; \vec{v}, \underline{\eta}, \underline{\underline{p}}, \underline{r}) \\
&\quad -\alpha_1 \left\langle \underline{\operatorname{div}}\, \underline{\underline{m}}(\vec{u}, \underline{\theta}) - \underline{q}, \underline{\operatorname{div}}\, \underline{\underline{m}}(\vec{v}, \underline{\eta}) - \underline{r} \right\rangle_h \\
&\quad -\alpha_2 \left\langle \underline{\operatorname{div}}\, \underline{\underline{\tilde{n}}}(\vec{u}, \underline{\theta}, \underline{\underline{n}}) - \underline{\underline{b}} \cdot \underline{q}, \underline{\operatorname{div}}\, \underline{\underline{\tilde{n}}}(\vec{v}, \underline{\eta}, \underline{\underline{p}}) - \underline{\underline{b}} \cdot \underline{r} \right\rangle_h \\
&\quad -\alpha_3 \left\langle \operatorname{div} \underline{q} + \underline{\underline{b}} : \underline{\underline{\tilde{n}}}(\vec{u}, \underline{\theta}, \underline{\underline{n}}), \operatorname{div} \underline{r} + \underline{\underline{b}} : \underline{\underline{\tilde{n}}}(\vec{v}, \underline{\eta}, \underline{\underline{p}}) \right\rangle_h \\
&\quad +\alpha_4 \int_\Omega \left[\underline{\underline{\varepsilon}}(\vec{u}) - t^2 \underline{\underline{\underline{\breve{E}}}} : \underline{\underline{n}} \right] : \underline{\underline{\underline{E}}} : \left[\underline{\underline{\varepsilon}}(\vec{v}) - t^2 \underline{\underline{\underline{\breve{E}}}} : \underline{\underline{p}} \right] dS \\
&\quad +\alpha_5 \int_\Omega \left[\underline{\gamma}(\vec{u}, \underline{\theta}) - t^2 \underline{\underline{\breve{G}}} \cdot \underline{q} \right] \cdot \underline{\underline{G}} \cdot \left[\underline{\gamma}(\vec{v}, \underline{\eta}) - t^2 \underline{\underline{\breve{G}}} \cdot \underline{r} \right] dS \\
&= F(\vec{v}) + \alpha_2 \left\langle \underline{f}, \underline{\operatorname{div}}\, \underline{\underline{\tilde{n}}}(\vec{v}, \underline{\eta}, \underline{\underline{p}}) - \underline{\underline{b}} \cdot \underline{r} \right\rangle_h \\
&\quad +\alpha_3 \left\langle f_3, \operatorname{div} \underline{r} + \underline{\underline{b}} : \underline{\underline{\tilde{n}}}(\vec{v}, \underline{\eta}, \underline{\underline{p}}) \right\rangle_h \\
&= F_h(\vec{v}, \underline{\eta}, \underline{\underline{p}}, \underline{r}).
\end{aligned}$$

■

In order to obtain a stable method we have to specify the finite element spaces. To this end we let, for $l \geq 0$,

$$R_l(K) \stackrel{\text{def}}{=} \begin{cases} P_l(K) & \text{if } K \text{ is a triangle,} \\ Q_l(K) & \text{if } K \text{ is a quadrilateral,} \end{cases} \tag{4.15}$$

and for $k \geq 1$ we now define

$$\begin{aligned}
\mathcal{U}^h \stackrel{\text{def}}{=} \{ (\vec{v}, \underline{\eta}) \in \mathcal{U} \quad | \quad & v_i|_K \in R_{k+1}(K),\ i = 1, 2, 3, \\
& \text{and } \eta_\alpha|_K \in R_k(K),\ \alpha = 1, 2,\ \forall K \in \mathcal{C}_h \},
\end{aligned} \tag{4.16}$$

$$\begin{aligned}
\mathcal{Q}^h \stackrel{\text{def}}{=} \{ (\underline{\underline{p}}, \underline{r}) \in \mathcal{Q} \quad | \quad & p^{\alpha\beta}|_K \in R_{k-1}(K),\ \alpha, \beta = 1, 2, \\
& \text{and } r^\alpha|_K \in R_{k-1}(K),\ \alpha = 1, 2,\ \forall K \in \mathcal{C}_h \}.
\end{aligned} \tag{4.17}$$

Remark 4.1 With the choice of finite elements spaces that we made, the stresses are interpolated discontinuously between elements. This implies that they can be

eliminated at the element level to obtain a purely displacement-based numerical scheme, cf. equation (6.3) below. ■

Remark 4.2 Our choice of covariant components for the displacement variables, and of contravariant ones for the stresses, enables us to establish Lemma 4.4 (using Lemma A.2), which we need for proving the stability of the method. ■

We are now ready to state our main stability result. We will assume that the shell thickness is in the range $0 < t \leq t_0$, with t_0 fixed.

Lemma 4.2 *Assume that:*

(S1) $0 < \alpha_i < C_i^I$, *for* $i = 1, 2$, *where the positive constants* C_i^I *are fixed, derived from inverse inequalities.*

(S2) $\alpha_3 > 0$.

(S3) $0 < \alpha_i < t_0^{-2}$, *for* $i = 4, 5$.

(S4) h *is sufficiently small.*

Then the stabilized formulation $\mathcal{S}_t^h$ *is stable, i.e. there is a positive constant* C *such that, for all* $(\vec{w}, \underline{\tau}, \underline{\underline{k}}, \underline{s}) \in \mathcal{U}^h \times \mathcal{Q}^h$, *there exist* $(\vec{z}, \underline{\eta}, \underline{\underline{p}}, \underline{r}) \in \mathcal{U}^h \times \mathcal{Q}^h$ *with*

$$B_h(\vec{w}, \underline{\tau}, \underline{\underline{k}}, \underline{s}; \vec{z}, \underline{\eta}, \underline{\underline{p}}, \underline{r}) \geq C \|\vec{w}, \underline{\tau}, \underline{\underline{k}}, \underline{s}\|_{t,h} \tag{4.18}$$

and

$$\|\vec{z}, \underline{\eta}, \underline{\underline{p}}, \underline{r}\|_{t,h} \leq 1. \tag{4.19}$$

Proof: Let $(\vec{w}, \underline{\tau}, \underline{\underline{k}}, \underline{s}) \in \mathcal{U}^h \times \mathcal{Q}^h$ be arbitrary. The proof we will now divide into three steps.

Step 1. First, a direct calculation gives

$$\begin{aligned}
&B_h(\vec{w}, \underline{\tau}, \underline{\underline{k}}, \underline{s}; \vec{w}, \underline{\tau}, -\underline{\underline{k}}, -\underline{s}) \\
&= A(\vec{w}, \underline{\tau}; \vec{w}, \underline{\tau}) + \alpha_4 \int_\Omega \underline{\underline{\varepsilon}}(\vec{w}) : \underline{\underline{\underline{\underline{E}}}} : \underline{\underline{\varepsilon}}(\vec{w})\, dS + \alpha_5 \int_\Omega \underline{\gamma}(\vec{w}, \underline{\tau}) \cdot \underline{\underline{G}} \cdot \underline{\gamma}(\vec{w}, \underline{\tau})\, dS \\
&\quad + t^2(1 - \alpha_4 t^2) \int_\Omega \underline{\underline{k}} : \underline{\underline{\underline{\underline{\breve{E}}}}} : \underline{\underline{k}}\, dS + t^2(1 - \alpha_5 t^2) \int_\Omega \underline{s} \cdot \underline{\underline{\breve{G}}} \cdot \underline{s}\, dS \\
&\quad + \alpha_1 \sum_{K \in \mathcal{C}_h} h_K^2 \left(\|\underline{s}\|_{0,K}^2 - \|\underline{\text{div}}\, \underline{\underline{m}}(\vec{w}, \underline{\tau})\|_{0,K}^2 \right) \\
&\quad + \alpha_2 \sum_{K \in \mathcal{C}_h} h_K^2 \left(\|\underline{\text{div}}\, \underline{\underline{k}} - \underline{b} \cdot \underline{s}\|_{0,K}^2 - \|\underline{\text{div}}\, \left[\underline{b} \cdot \underline{\underline{m}}(\vec{w}, \underline{\tau}) \right] \|_{0,K}^2 \right) \\
&\quad + \alpha_3 \sum_{K \in \mathcal{C}_h} h_K^2 \left(\|\text{div}\, \underline{s} + \underline{b} : \underline{\underline{k}}\|_{0,K}^2 - \|\underline{c} : \underline{\underline{m}}(\vec{w}, \underline{\tau})\|_{0,K}^2 \right).
\end{aligned} \tag{4.20}$$

We will use the following inverse estimates

$$\sum_{K\in\mathcal{C}_h} h_K^2 \|\underline{\mathrm{div}}\, \underline{\underline{m}}(\vec{w},\underline{\tau})\|_{0,K}^2 \le C_1(A(\vec{w},\underline{\tau};\vec{w},\underline{\tau}) + h^2\|\vec{w},\underline{\tau}\|_{0,K}^2), \tag{4.21}$$

$$\sum_{K\in\mathcal{C}_h} h_K^2 \|\underline{\mathrm{div}}\, (\underline{b}\cdot\underline{\underline{m}}(\vec{w},\underline{\tau}))\|_{0,K}^2 \le C_2(A(\vec{w},\underline{\tau};\vec{w},\underline{\tau}) + h^2\|\vec{w},\underline{\tau}\|_0^2). \tag{4.22}$$

For clarity, we will establish (4.21) and (4.22) as a separate result in Lemma 4.3 below. We also have

$$\sum_{K\in\mathcal{C}_h} h_K^2 \|\underline{\underline{c}} : \underline{\underline{m}}(\vec{w},\underline{\tau})\|_{0,K}^2 \le Ch^2\|\underline{\underline{m}}(\vec{w},\underline{\tau})\|_0^2, \tag{4.23}$$

and from Lemma A.1:

$$A(\vec{w},\underline{\tau};\vec{w},\underline{\tau}) = 12\int_\Omega \underline{\underline{m}}(\vec{w},\underline{\tau}) : \breve{\underline{\underline{\underline{\underline{E}}}}} : \underline{\underline{m}}(\vec{w},\underline{\tau})\, dS \ge C\|\underline{\underline{m}}(\vec{w},\underline{\tau})\|_0^2, \tag{4.24}$$

so that we get

$$\sum_{K\in\mathcal{C}_h} h_K^2 \|\underline{\underline{c}} : \underline{\underline{m}}(\vec{w},\underline{\tau})\|_{0,K}^2 \le Ch^2 A(\vec{w},\underline{\tau};\vec{w},\underline{\tau}). \tag{4.25}$$

Combining equations (4.21), (4.22) and (4.25) with (4.20), we obtain

$$\begin{aligned} B_h(\vec{w},\underline{\tau},\underline{\underline{k}},\underline{s};\vec{w},\underline{\tau},-\underline{\underline{k}},-\underline{s}) \\ \ge\ & (1-\alpha_1 C_1 - \alpha_2 C_2 - Ch^2)A(\vec{w},\underline{\tau};\vec{w},\underline{\tau}) - Ch^2\|\vec{w},\underline{\tau}\|_0^2 \\ & +\alpha_4\int_\Omega \underline{\underline{\varepsilon}}(\vec{w}) : \underline{\underline{\underline{\underline{E}}}} : \underline{\underline{\varepsilon}}(\vec{w})\, dS + \alpha_5\int_\Omega \underline{\gamma}(\vec{w},\underline{\tau})\cdot\underline{\underline{G}}\cdot\underline{\gamma}(\vec{w},\underline{\tau})\, dS \\ & +t^2(1-\alpha_4 t^2)\int_\Omega \underline{\underline{k}} : \breve{\underline{\underline{\underline{\underline{E}}}}} : \underline{\underline{k}}\, dS + t^2(1-\alpha_5 t^2)\int_\Omega \underline{s}\cdot\breve{\underline{\underline{G}}}\cdot\underline{s}\, dS \\ & +\alpha_1\sum_{K\in\mathcal{C}_h} h_K^2\|\underline{s}\|_{0,K}^2 + \alpha_2\sum_{K\in\mathcal{C}_h} h_K^2\|\underline{\mathrm{div}}\,\underline{\underline{k}} - \underline{\underline{b}}\cdot\underline{s}\|_{0,K}^2 \\ & +\alpha_3\sum_{K\in\mathcal{C}_h} h_K^2\|\mathrm{div}\,\underline{s} + \underline{\underline{b}} : \underline{\underline{k}}\|_{0,K}^2. \end{aligned} \tag{4.26}$$

Since $\alpha_i > 0$, $i = 4, 5$, the ellipticity result proved in [6, 18] gives

$$\begin{aligned} (1-\alpha_1 C_1 - \alpha_2 C_2 - Ch^2)A(\vec{w},\underline{\tau};\vec{w},\underline{\tau}) \\ +\alpha_4\int_\Omega \underline{\underline{\varepsilon}}(\vec{w}) : \underline{\underline{\underline{\underline{E}}}} : \underline{\underline{\varepsilon}}(\vec{w})\, dS + \alpha_5\int_\Omega \underline{\gamma}(\vec{w},\underline{\tau})\cdot\underline{\underline{G}}\cdot\underline{\gamma}(\vec{w},\underline{\tau})\, dS \ge C\|\vec{w},\underline{\tau}\|_1^2, \end{aligned} \tag{4.27}$$

when choosing

$$\alpha_i < C_i^I \stackrel{\mathrm{def}}{=} (2C_i)^{-1},\ i = 1, 2, \tag{4.28}$$

and for h small enough. Next, Lemma A.1 implies

$$\int_\Omega \underline{\underline{k}} : \breve{\underline{\underline{\underline{\underline{E}}}}} : \underline{\underline{k}}\, dS \ge C\|\underline{\underline{k}}\|_0^2. \tag{4.29}$$

Further, it holds

$$\int_\Omega \underline{s}\cdot\breve{\underline{\underline{G}}}\cdot\underline{s}\, dS = \frac{2(1+\nu)}{E}\|\underline{s}\|_0^2. \tag{4.30}$$

Since

$$1 - \alpha_i t^2 \geq 1 - \alpha_i t_0^2 > 0, \ i = 4, 5, \tag{4.31}$$

a combination of the above relations gives

$$\begin{aligned} & B_h(\vec{w}, \underline{\tau}, \underline{\underline{k}}, \underline{s}; \vec{w}, \underline{\tau}, -\underline{\underline{k}}, -\underline{s}) \\ & \quad \geq \ C_3 \left(\|\vec{w}, \underline{\tau}\|_1^2 + t^2 \|\underline{\underline{k}}, \underline{s}\|_0^2 \right) \\ & \qquad + \alpha_1 \sum_{K \in \mathcal{C}_h} h_K^2 \|\underline{s}\|_{0,K}^2 + \alpha_2 \sum_{K \in \mathcal{C}_h} h_K^2 \|\underline{\mathrm{div}}\, \underline{\underline{k}} - \underline{\underline{b}} \cdot \underline{s}\|_{0,K}^2 \\ & \qquad + \alpha_3 \sum_{K \in \mathcal{C}_h} h_K^2 \|\mathrm{div}\, \underline{s} + \underline{\underline{b}} : \underline{\underline{k}}\|_{0,K}^2, \end{aligned} \tag{4.32}$$

again when h is small enough, to take care of the term $-Ch^2 \|\vec{w}, \underline{\tau}\|_0^2$ in (4.26). Finally, using the arithmetic-geometric mean inequality, and the boundedness of $\underline{\underline{b}}$, we get, for $0 < \eta < 1$,

$$\sum_{K \in \mathcal{C}_h} h_K^2 \|\underline{\mathrm{div}}\, \underline{\underline{k}} - \underline{\underline{b}} \cdot \underline{s}\|_{0,K}^2 \geq (1 - \eta) \sum_{K \in \mathcal{C}_h} h_K^2 \|\underline{\mathrm{div}}\, \underline{\underline{k}}\|_{0,K}^2 - C_4 (\frac{1}{\eta} - 1) \sum_{K \in \mathcal{C}_h} h_K^2 \|\underline{s}\|_{0,K}^2. \tag{4.33}$$

Hence, by setting η close enough to one, we obtain

$$B_h(\vec{w}, \underline{\tau}, \underline{\underline{k}}, \underline{s}; \vec{w}, \underline{\tau}, -\underline{\underline{k}}, -\underline{s}) \geq C_5 \left(\|\vec{w}, \underline{\tau}\|_1^2 + t^2 \|\underline{\underline{k}}, \underline{s}\|_0^2 + |\underline{\underline{k}}, \underline{s}|_I^2 \right) \tag{4.34}$$

Step 2. Next, we use the result that there exists $(\vec{v}, \underline{0}) \in \mathcal{U}^h$ such that

$$\sum_{E \in \Gamma_h \setminus \Gamma_0} \int_E \left([\![\underline{\underline{k}} \cdot \underline{\nu}]\!] \cdot \underline{v} + [\![\underline{s} \cdot \underline{\nu}]\!] \, v_3 \right) ds \geq C_6 |\underline{\underline{k}}, \underline{s}|_J^2 \tag{4.35}$$

and

$$\sum_{K \in \mathcal{C}_h} h_K^{-2} \|\vec{v}\|_{0,K}^2 + \|\vec{v}\|_1^2 \leq |\underline{\underline{k}}, \underline{s}|_J^2. \tag{4.36}$$

The proof of this result is rather technical and hence we will postpone it to Lemma 4.4 below.

With this $\vec{v}$ we first write

$$B_h(\vec{w}, \underline{\tau}, \underline{\underline{k}}, \underline{s}; \vec{v}, \underline{0}, \underline{\underline{0}}, \underline{0}) = I + II, \tag{4.37}$$

with

$$\begin{aligned} I \ \stackrel{\mathrm{def}}{=} \ & A(\vec{w}, \underline{\tau}; \vec{v}, \underline{0}) + \alpha_4 \int_\Omega \underline{\underline{\varepsilon}}(\vec{w}) : \underline{\underline{\underline{E}}} : \underline{\underline{\varepsilon}}(\vec{v}) \, dS \\ & + \alpha_5 \int_\Omega \underline{\gamma}(\vec{w}, \underline{\tau}) \cdot \underline{\gamma}(\vec{v}, \underline{0}) \, dS - \alpha_1 \left\langle \underline{\mathrm{div}}\, \underline{\underline{m}}(\vec{w}, \underline{\tau}) - \underline{s}, \underline{\mathrm{div}}\, \underline{\underline{m}}(\vec{v}, \underline{0}) \right\rangle_h \\ & - \alpha_2 \left\langle \underline{\mathrm{div}}\, \underline{\underline{\tilde{n}}}(\vec{w}, \underline{\tau}, \underline{\underline{k}}) - \underline{\underline{b}} \cdot \underline{s}, \underline{\mathrm{div}}\, \underline{\underline{\tilde{n}}}(\vec{v}, \underline{0}, \underline{\underline{0}}) \right\rangle_h \\ & - \alpha_3 \left\langle \mathrm{div}\, \underline{s} + \underline{\underline{b}} : \underline{\underline{\tilde{n}}}(\vec{w}, \underline{\tau}, \underline{\underline{k}}), \underline{\underline{b}} : \underline{\underline{\tilde{n}}}(\vec{v}, \underline{0}, \underline{\underline{0}}) \right\rangle_h, \end{aligned} \tag{4.38}$$

and

$$II \stackrel{\text{def}}{=} (1-\alpha_4 t^2)\int_\Omega \underline{\underline{\varepsilon}}(\vec{v}) : \underline{\underline{k}}\, dS + (1-\alpha_5 t^2)\int_\Omega \underline{\gamma}(\vec{v},\underline{0})\cdot \underline{s}\, dS. \tag{4.39}$$

Note that estimates (4.21) and (4.22) immediately imply, for all $(\vec{x},\underline{\chi}) \in \mathcal{U}^h$:

$$\sum_{K\in\mathcal{C}_h} h_K^2 \|\underline{\text{div}}\, \underline{\underline{m}}(\vec{x},\underline{\chi})\|_{0,K}^2 \le C\|\vec{x},\underline{\chi}\|_1^2, \tag{4.40}$$

$$\sum_{K\in\mathcal{C}_h} h_K^2 \|\underline{\text{div}}\, \Big[\underline{\underline{b}}\cdot \underline{\underline{m}}(\vec{x},\underline{\chi})\Big]\|_{0,K}^2 \le C\|\vec{x},\underline{\chi}\|_1^2. \tag{4.41}$$

In order to bound I, we apply these relations to both $(\vec{w},\underline{\tau})$ and $(\vec{v},\underline{0})$, and we use the Schwarz inequality and (4.36). We obtain

$$|I| \le C_7\|\vec{v}\|_1\Big(\|\vec{w},\underline{\tau}\|_1^2 + |\underline{\underline{k}},\underline{s}|_I^2\Big)^{1/2} \le C_7|\underline{\underline{k}},\underline{s}|_J\Big(\|\vec{w},\underline{\tau}\|_1^2 + |\underline{\underline{k}},\underline{s}|_I^2\Big)^{1/2} \tag{4.42}$$

We transform the term II by using the integration by parts formulas (3.13) and (3.14) over each element

$$\begin{aligned} II &= III - (1-\alpha_4 t^2)\sum_{K\in\mathcal{C}_h}\int_K \Big(\underline{\text{div}}\,\underline{\underline{k}} - \underline{\underline{b}}\cdot\underline{s}\Big)\cdot\underline{v}\, dS \\ &\quad -(1-\alpha_5 t^2)\sum_{K\in\mathcal{C}_h}\int_K \Big(\text{div}\,\underline{s} + \underline{\underline{b}} : \underline{\underline{k}}\Big)\, v_3\, dS, \end{aligned} \tag{4.43}$$

with

$$III \stackrel{\text{def}}{=} (1-\alpha_4 t^2)\sum_{E\in\Gamma_h\setminus\Gamma_0}\int_E [\![\underline{\underline{k}}\cdot\underline{\nu}]\!]\cdot\underline{v}\, ds + (1-\alpha_5 t^2)\sum_{E\in\Gamma_h\setminus\Gamma_0}\int_E [\![\underline{s}\cdot\underline{\nu}]\!]\, v_3\, ds.$$

Hence, using (4.31), (4.35), (4.36) and the arithmetic-geometric mean inequality we get

$$\begin{aligned} II &\ge III - |\underline{\underline{k}},\underline{s}|_I\Big(\sum_{K\in\mathcal{C}_h} h_K^{-2}\|\vec{v}\|_{0,K}^2\Big)^{1/2} \\ &\ge III - |\underline{\underline{k}},\underline{s}|_I|\underline{\underline{k}},\underline{s}|_J \\ &\ge \frac{C_6}{2}|\underline{\underline{k}},\underline{s}|_J^2 - |\underline{\underline{k}},\underline{s}|_I|\underline{\underline{k}},\underline{s}|_J \\ &\ge \frac{C_6}{4}|\underline{\underline{k}},\underline{s}|_J^2 - \frac{1}{C_6}|\underline{\underline{k}},\underline{s}|_I^2. \end{aligned} \tag{4.44}$$

Combining (4.37)–(4.44) and using, once again, the arithmetic-geometric mean inequality, we obtain

$$\begin{aligned} &B_h(\vec{w},\underline{\tau},\underline{\underline{k}},\underline{s};\vec{v},\underline{0},\underline{\underline{0}},\underline{0}) \\ &\qquad \ge -C_7|\underline{\underline{k}},\underline{s}|_J\Big(\|\vec{w},\underline{\tau}\|_1^2 + |\underline{\underline{k}},\underline{s}|_I^2\Big)^{1/2} + \frac{C_6}{4}|\underline{\underline{k}},\underline{s}|_J^2 - \frac{1}{C_6}|\underline{\underline{k}},\underline{s}|_I^2 \\ &\qquad \ge \Big(\frac{C_6}{4} - \frac{C_7\sigma}{2}\Big)|\underline{\underline{k}},\underline{s}|_J^2 - \frac{C_7}{2\sigma}\Big(\|\vec{w},\underline{\tau}\|_1^2 + |\underline{\underline{k}},\underline{s}|_I^2\Big) - \frac{1}{C_6}|\underline{\underline{k}},\underline{s}|_I^2\ , \end{aligned}$$

for any strictly positive σ. By choosing σ small enough, we then have

$$B_h(\vec{w},\underline{\tau},\underline{\underline{k}},\underline{s};\vec{v},\underline{0},\underline{\underline{0}},\underline{0}) \geq C_8|\underline{\underline{k}},\underline{s}|_J^2 - C_9\Big(\|\vec{w},\underline{\tau}\|_1^2 + |\underline{\underline{k}},\underline{s}|_I^2\Big). \tag{4.45}$$

Step 3. Choose now $(\vec{z},\underline{\eta},\underline{\underline{p}},\underline{r}) = (\vec{w}+\delta\vec{v},\underline{\tau},-\underline{\underline{k}},-\underline{s})$ with $\delta > 0$. The relations (4.34) and (4.45) proved in the previous steps then give

$$\begin{aligned} B_h(\vec{w},\underline{\tau},\underline{\underline{k}},\underline{s};\vec{z},\underline{\eta},\underline{\underline{p}},\underline{r}) & \\ &\geq (C_5 - \delta C_9)\left(\|\vec{w},\underline{\tau}\|_1^2 + |\underline{\underline{k}},\underline{s}|_I^2\right) + C_5 t^2\|\underline{\underline{k}},\underline{s}\|_0^2 + \delta C_8|\underline{\underline{k}},\underline{s}|_J^2 \\ &\geq C_{10}\|\vec{w},\underline{\tau},\underline{\underline{k}},\underline{s}\|_{t,h}^2, \end{aligned}$$

when choosing $\delta < C_5/C_9$. Using (4.36) we get

$$\begin{aligned} \|\vec{z},\underline{\eta},\underline{\underline{p}},\underline{r}\|_{t,h} &\leq \|\vec{w},\underline{\tau},\underline{\underline{k}},\underline{s}\|_{t,h} + \delta\|\vec{v},\underline{0},\underline{\underline{0}},\underline{0}\|_{t,h} = \|\vec{w},\underline{\tau},\underline{\underline{k}},\underline{s}\|_{t,h} + \delta\|\vec{v}\|_1 \\ &\leq \|\vec{w},\underline{\tau},\underline{\underline{k}},\underline{s}\|_{t,h} + \delta|\underline{\underline{k}},\underline{s}|_J \leq C_{11}\|\vec{w},\underline{\tau},\underline{\underline{k}},\underline{s}\|_{t,h}, \end{aligned} \tag{4.46}$$

and the assertion is thus proved. ■

Next, we will prove the missing steps in the proof above. We start with the inverse estimates (4.21) and (4.22).

Lemma 4.3 *The following inequalities hold for any* $(\vec{v},\underline{\eta}) \in \mathcal{U}^h$*:*

$$\sum_{K\in\mathcal{C}_h} h_K^2\|\underline{\mathrm{div}}\ \underline{\underline{m}}(\vec{v},\underline{\eta})\|_{0,K}^2 \leq C_1(A(\vec{v},\underline{\eta};\vec{v},\underline{\eta}) + h^2\|\vec{v},\underline{\eta}\|_0^2), \tag{4.47}$$

$$\sum_{K\in\mathcal{C}_h} h_K^2\|\underline{\mathrm{div}}\ \left[\underline{\underline{b}}\cdot\underline{\underline{m}}(\vec{v},\underline{\eta})\right]\|_{0,K}^2 \leq C_2(A(\vec{v},\underline{\eta};\vec{v},\underline{\eta}) + h^2\|\vec{v},\underline{\eta}\|_0^2). \tag{4.48}$$

Proof: First we note that

$$\sum_{K\in\mathcal{C}_h} h_K^2\|\underline{\mathrm{div}}\ \underline{\underline{m}}(\vec{v},\underline{\eta})\|_{0,K}^2 \leq C\sum_{K\in\mathcal{C}_h} h_K^2\|\underline{\underline{m}}(\vec{v},\underline{\eta})\|_{1,K}^2, \tag{4.49}$$

while, by virtue of the smoothness of $\underline{\underline{b}}$ we also have

$$\sum_{K\in\mathcal{C}_h} h_K^2\|\underline{\mathrm{div}}\ \left[\underline{\underline{b}}\cdot\underline{\underline{m}}(\vec{v},\underline{\eta})\right]\|_{0,K}^2 \leq C\sum_{K\in\mathcal{C}_h} h_K^2\|\underline{\underline{m}}(\vec{v},\underline{\eta})\|_{1,K}^2. \tag{4.50}$$

Therefore the proof will be identical for both estimates. Further, the above definition of Sobolev norms for tensors implies

$$\|\underline{\underline{m}}(\vec{v},\underline{\eta})\|_{1,K}^2 \leq C\sum_{\alpha,\beta}\|m^{\alpha\beta}(\vec{v},\underline{\eta})\|_{1,K}^2, \tag{4.51}$$

where the norm in the right hand side now denotes the standard H^1-norm. Our argument will not require the detailed expression of $m^{\alpha\beta}(\vec{v},\underline{\eta})$, so we instead symbolically denote

$$m^{\alpha\beta}(\vec{v},\underline{\eta}) = \rho_1^{\alpha\beta i\mu}v_{i,\mu} + \rho_2^{\alpha\beta i}v_i + \rho_3^{\alpha\beta i\mu}\eta_{i,\mu} + \rho_4^{\alpha\beta i}\eta_i, \tag{4.52}$$

where all coefficients involved are smooth functions that incorporate geometric and material terms. For a smooth function ρ we now define $T^h\rho$ as the linear part of its Taylor expansion at the center of each element. Using standard expansion properties, we have for any (α,β) and any element

$$\begin{aligned}\|(\rho_1^{\alpha\beta i\mu} - T^h\rho_1^{\alpha\beta i\mu})v_{i,\mu} + (\rho_2^{\alpha\beta i} - T^h\rho_2^{\alpha\beta i})v_i \\ +(\rho_3^{\alpha\beta i\mu} - T^h\rho_3^{\alpha\beta i\mu})\eta_{i,\mu} + (\rho_4^{\alpha\beta i} - T^h\rho_4^{\alpha\beta i})\eta_i\|_{0,K}^2 \leq Ch_K^4\|\vec{v},\underline{\eta}\|_{1,K}^2,\end{aligned} \tag{4.53}$$

$$\begin{aligned}\|(\rho_1^{\alpha\beta i\mu} - T^h\rho_1^{\alpha\beta i\mu})v_{i,\mu} + (\rho_2^{\alpha\beta i} - T^h\rho_2^{\alpha\beta i})v_i + (\rho_3^{\alpha\beta i\mu} - T^h\rho_3^{\alpha\beta i\mu})\eta_{i,\mu} \\ +(\rho_4^{\alpha\beta i} - T^h\rho_4^{\alpha\beta i})\eta_i\|_{1,K}^2 \leq C(h_K^4\|\vec{v},\underline{\eta}\|_{2,K}^2 + h_K^2\|\vec{v},\underline{\eta}\|_{1,K}^2).\end{aligned} \tag{4.54}$$

Hence, we obtain

$$\begin{aligned}&h_K^2\|m^{\alpha\beta}(\vec{v},\underline{\eta})\|_{1,K}^2 \\ &\quad\leq\ h_K^2\|(T^h\rho_1^{\alpha\beta i\mu})v_{i,\mu} + (T^h\rho_2^{\alpha\beta i})v_i + (T^h\rho_3^{\alpha\beta i\mu})\eta_{i,\mu} + (T^h\rho_4^{\alpha\beta i})\eta_i\|_{1,K}^2 \\ &\qquad +Ch_K^6\|\vec{v},\underline{\eta}\|_{2,K}^2 + Ch_K^4\|\vec{v},\underline{\eta}\|_{1,K}^2.\end{aligned} \tag{4.55}$$

We can now invoke standard inverse estimates on polynomial functions to obtain

$$\begin{aligned}&h_K^2\|m^{\alpha\beta}(\vec{v},\underline{\eta})\|_{1,K}^2 \\ &\quad\leq\ \|(T^h\rho_1^{\alpha\beta i\mu})v_{i,\mu} + (T^h\rho_2^{\alpha\beta i})v_i + (T^h\rho_3^{\alpha\beta i\mu})\eta_{i,\mu} + (T^h\rho_4^{\alpha\beta i})\eta_i\|_{0,K}^2 \\ &\qquad +Ch_K^2\|\vec{v},\underline{\eta}\|_{0,K}^2,\end{aligned} \tag{4.56}$$

and, using (4.53) combined with an inverse estimate, we finally get

$$h_K^2\|m^{\alpha\beta}(\vec{v},\underline{\eta})\|_{1,K}^2 \leq \|m^{\alpha\beta}(\vec{v},\underline{\eta})\|_{0,K}^2 + Ch_K^2\|\vec{v},\underline{\eta}\|_{0,K}^2, \tag{4.57}$$

which, combined with (4.24), concludes the proof. ■

Remark 4.3 It is not clear whether estimates (4.47) and (4.48) would hold without the terms $h^2\|\vec{v},\underline{\eta}\|_0^2$. These estimates, however, are sufficient for our purposes in their present form. ■

Remark 4.4 In the above proof, it clearly appears that estimates (4.47) and (4.48) follow from similar local estimates that we can explicitly write as:

$$h_K^2\|\underline{\mathrm{div}}\,\underline{\underline{m}}(\vec{v},\underline{\eta})\|_{0,K}^2 \leq C_1(\frac{1}{12}\int_K \underline{\underline{\kappa}}(\vec{v},\underline{\eta}) : \underline{\underline{\underline{E}}} : \underline{\underline{\kappa}}(\vec{v},\underline{\eta})\,dS + h_K^2\|\vec{v},\underline{\eta}\|_{0,K}^2), \tag{4.58}$$

$$\begin{aligned}&h_K^2\|\underline{\mathrm{div}}\,\left[\underline{b}\cdot\underline{\underline{m}}(\vec{v},\underline{\eta})\right]\|_{0,K}^2 \\ &\qquad\leq C_2(\frac{1}{12}\int_K \underline{\underline{\kappa}}(\vec{v},\underline{\eta}) : \underline{\underline{\underline{E}}} : \underline{\underline{\kappa}}(\vec{v},\underline{\eta})\,dS + h_K^2\|\vec{v},\underline{\eta}\|_{0,K}^2).\end{aligned} \tag{4.59}$$

This implies that these inverse inequality constants can be computed for each element using a simple local eigenvalue problem, cf. [25] where this idea is introduced for several stabilized methods. Therefore the choice of α_1 and α_2 can be made automatically in the course of the assembling process. ■

We proceed to establish the result required at the second step of the proof of Lemma 4.2. For this, the technical Lemmas A.1 and A.2 are essential.

Lemma 4.4 *There is a positive constant C such that, for all $(\underline{\underline{k}}, \underline{s}) \in \mathcal{Q}^h$, there exists $(\vec{v}, \underline{0}) \in \mathcal{U}^h$ such that*

$$\sum_{E \in \Gamma_h \setminus \Gamma_0} \int_E \left([\![\underline{\underline{k}} \cdot \underline{\nu}]\!] \cdot \underline{v} \, ds + [\![\underline{s} \cdot \underline{\nu}]\!] \, v_3 \right) ds \geq C |\underline{\underline{k}}, \underline{s}|_J$$

and

$$\sum_{K \in \mathcal{C}_h} h_K^{-2} \|\vec{v}\|_{0,K}^2 + \|\vec{v}\|_1^2 \leq 1.$$

Proof: We use the normal Lagrange degrees of freedom for $\vec{v}$ and we choose these so that $\vec{v} = \vec{0}$ at the vertices and the internal nodes of all elements. To the remaining degrees of freedom of $\vec{v}$ along edges we assign values so that

$$v_\alpha|_E = h_E \, \bar{a}_{\alpha\lambda} [\![k^{\lambda\mu} \bar{\nu}_\mu]\!] \, b_E, \quad v_3|_E = h_E \, [\![s^\mu \bar{\nu}_\mu]\!] \, b_E, \tag{4.60}$$

where a bar over the symbol of a continuous function denotes its value at the midpoint of the edge, and b_E stands for the second-degree "bubble function" along the edge with value one at the midpoint (i.e. $\bar{b}_E = 1$). As it holds

$$[\![k^{\lambda\mu}]\!]|_E \in P_{k-1}(E), \quad [\![s^\mu]\!]|_E \in P_{k-1}(E) \ \text{ and } \ b_E \in P_2(E), \tag{4.61}$$

this construction yields a function $\vec{v}$ lying in the appropriate finite element subspace for the displacement.

Now, we first note that Lemma A.2 gives

$$\begin{aligned} \sum_{E \in \Gamma_h \setminus \Gamma_0} \int_E [\![\underline{\underline{k}} \cdot \underline{\nu}]\!] \cdot \underline{v} \, ds &= \sum_{E \in \Gamma_h \setminus \Gamma_0} h_E \int_E [\![k^{\alpha\lambda} \nu_\lambda]\!] \bar{a}_{\alpha\beta} [\![k^{\beta\mu} \bar{\nu}_\mu]\!] \, b_E \, ds \\ &= \sum_{E \in \Gamma_h \setminus \Gamma_0} h_E \int_E [\![k^{\alpha\lambda} \bar{\nu}_\lambda]\!] \bar{a}_{\alpha\beta} [\![k^{\alpha\mu} \bar{\nu}_\mu]\!] \, b_E \zeta \, ds, \end{aligned}$$

where ζ is a smooth positive function bounded away from zero. Then we can apply Lemma A.1 to obtain

$$\begin{aligned} \sum_{E \in \Gamma_h \setminus \Gamma_0} \int_E [\![\underline{\underline{k}} \cdot \underline{\nu}]\!] \cdot \underline{v} \, ds &\geq C \sum_{E \in \Gamma_h \setminus \Gamma_0} h_E \int_E [\![k^{\alpha\lambda} \bar{\nu}_\lambda]\!] \, [\![k^{\alpha\mu} \bar{\nu}_\mu]\!] \, b_E \zeta \, ds \\ &\geq C \sum_{E \in \Gamma_h \setminus \Gamma_0} h_E \int_E [\![k^{\alpha\lambda} \bar{\nu}_\lambda]\!] \, [\![k^{\alpha\mu} \bar{\nu}_\mu]\!] \, b_E \, ds, \end{aligned}$$

and since the quantities involved are now polynomial functions, a standard scaling argument yields

$$\sum_{E\in\Gamma_h\setminus\Gamma_0} h_E \int_E [\![k^{\alpha\lambda}\bar\nu_\lambda]\!]\,[\![k^{\alpha\mu}\bar\nu_\mu]\!]\,b_E\,ds \geq C \sum_{E\in\Gamma_h\setminus\Gamma_0} h_E \int_E [\![k^{\alpha\lambda}\bar\nu_\lambda]\!]\,[\![k^{\alpha\mu}\bar\nu_\mu]\!]\,ds$$

so that, using again Lemmas A.1 and A.2, we have

$$\begin{aligned}
\sum_{E\in\Gamma_h\setminus\Gamma_0} \int_E [\![\underline{\underline{k}}\cdot\underline{\nu}]\!]\cdot\underline{v}\,ds &\geq C \sum_{E\in\Gamma_h\setminus\Gamma_0} h_E \int_E [\![k^{\alpha\lambda}\nu_\lambda]\!]\,[\![k^{\alpha\mu}\nu_\mu]\!]\,ds \\
&\geq C \sum_{E\in\Gamma_h\setminus\Gamma_0} h_E \int_E [\![k^{\alpha\lambda}\nu_\lambda]\!]\,a_{\alpha\beta}[\![k^{\alpha\mu}\nu_\mu]\!]\,ds \\
&\geq C \sum_{E\in\Gamma_h\setminus\Gamma_0} h_E \|[\![\underline{\underline{k}}\cdot\underline{\nu}]\!]\|^2_{0,E}\,. \qquad (4.62)
\end{aligned}$$

To bound $\underline{v}$, we first note that inverse estimates imply

$$\|\underline{v}\|_1^2 \leq C \sum_{K\in\mathcal{C}_h} h_K^{-2}\|\underline{v}\|^2_{0,K}\,.$$

Then, using the fact that all degrees of freedom are equal to zero except along edges, we get by scaling arguments, and Lemmas A.1 and A.2

$$\|\underline{v}\|_1^2 + \sum_{K\in\mathcal{C}_h} h_K^{-2}\|\underline{v}\|^2_{0,K} \leq C \sum_{E\in\Gamma_h\setminus\Gamma_0} h_E \|[\![\underline{\underline{k}}\cdot\underline{\nu}]\!]\|^2_{0,E}\,. \qquad (4.63)$$

For the component v_3 similar arguments give

$$\sum_{E\in\Gamma_h\setminus\Gamma_0} \int_E [\![\underline{s}\cdot\underline{\nu}]\!]\,v_3\,ds \geq C \sum_{E\in\Gamma_h\setminus\Gamma_0} h_E \|[\![\underline{s}\cdot\underline{\nu}]\!]\|^2_{0,E} \qquad (4.64)$$

and

$$\|v_3\|_1^2 + \sum_{K\in\mathcal{C}_h} h_K^{-2}\|v_3\|^2_{0,K} \leq C \sum_{E\in\Gamma_h\setminus\Gamma_0} h_E \|[\![\underline{s}\cdot\underline{\nu}]\!]\|^2_{0,E}\,. \qquad (4.65)$$

The estimates (4.62)–(4.65) then prove the assertion. ■

To perform the error analysis, interpolation estimates in non-standard norms will be required.

Lemma 4.5 *For* $(\vec u,\underline\theta)\in\mathcal{U}\cap\left([H^{k+1}(\Omega)]^3\times[H^{k+1}(\Omega)]^2\right)$ *it holds*

$$\begin{aligned}
\inf_{(\vec v,\underline\eta)\in\mathcal{U}^h} \Big\{ &\|\vec u-\vec v,\underline\theta-\underline\eta\|_1^2 + \sum_{E\in\Gamma_h\setminus\Gamma_0} h_E^{-1}\|\vec u-\vec v\|^2_{0,E} \qquad (4.66)\\
&+ \sum_{K\in\mathcal{C}_h} (h_K^{-2}\|\vec u-\vec v,\underline\theta-\underline\eta\|^2_{0,K} + h_K^2\|\vec u-\vec v,\underline\theta-\underline\eta\|^2_{2,K})\Big\}^{1/2} \\
&\leq C h^k \|\vec u,\underline\theta\|_{k+1}\,.
\end{aligned}$$

Proof: The estimate follows from scaling arguments. ■

Lemma 4.6 *For* $(\underline{\underline{n}}, \underline{q}) \in \mathcal{Q} \cap \left([H^k(\Omega)]^{2\times 2} \times [H^k(\Omega)]^2\right)$ *it holds*

$$\inf_{(\underline{\underline{p}},\underline{r})\in\mathcal{Q}^h} \left\{ \|\underline{\underline{n}} - \underline{\underline{p}}, \underline{q} - \underline{r}\|_{t,h} + \|\underline{\underline{n}} - \underline{\underline{p}}, \underline{q} - \underline{r}\|_0 \right\} \leq Ch^k \|\underline{\underline{n}}, \underline{q}\|_k \ . \tag{4.67}$$

Proof: Let $(\underline{\underline{p}}, \underline{r})$ be the L^2-projection of $(\underline{\underline{n}}, \underline{q})$ in $\mathcal{Q}^h$. Since no continuity is imposed on the finite element functions, this projection is defined locally element by element. We now have

$$\begin{aligned}\|\underline{\underline{n}} - \underline{\underline{p}}, \underline{q} - \underline{r}\|_{t,h}^2 + \|\underline{\underline{n}} - \underline{\underline{p}}, \underline{q} - \underline{r}\|_0^2 \leq C\Big(&\|\underline{\underline{n}} - \underline{\underline{p}}, \underline{q} - \underline{r}\|_0^2 \\ &+ \sum_{K\in\mathcal{C}_h} h_K^2 \|\underline{\underline{n}} - \underline{\underline{p}}, \underline{q} - \underline{r}\|_{1,K}^2 + \sum_{K\in\mathcal{C}_h} \sum_{E\in\partial K} h_E (\|\underline{\underline{n}} - \underline{\underline{p}}\|_{0,E}^2 + \|\underline{q} - \underline{r}\|_{0,E}^2)\Big)\end{aligned}$$

The last term above is handled by a scaling argument

$$\begin{aligned}\sum_{K\in\mathcal{C}_h} \sum_{E\in\partial K} h_E (\|\underline{\underline{n}} - \underline{\underline{p}}\|_{0,E}^2 + \|\underline{q} - \underline{r}\|_{0,E}^2) \\ \leq C\Big(\|\underline{\underline{n}} - \underline{\underline{p}}, \underline{q} - \underline{r}\|_0^2 + \sum_{K\in\mathcal{C}_h} h_K^2 \|\underline{\underline{n}} - \underline{\underline{p}}, \underline{q} - \underline{r}\|_{1,K}^2 \Big)\end{aligned}$$

and the asserted estimate then follows from standard estimates. ■

Let us now prove the error estimate for our numerical scheme. In the proof we will repeatedly need the estimate

$$\Big(\sum_{K\in\mathcal{C}_h} h_K^2 \|\vec{v}, \underline{\eta}\|_{2,K}^2 \Big)^{1/2} \leq C \|\vec{v}, \underline{\eta}\|_{1,K} \quad \forall (\vec{v}, \underline{\eta}) \in \mathcal{U}^h, \tag{4.68}$$

which is a direct consequence of standard inverse inequalities.

Theorem 4.1 *Assume that conditions* (S1)–(S4) *of Lemma* 4.2 *hold and that the solution of* $\mathcal{M}_t$ *is such that* $\vec{u} \in [H^{k+1}(\Omega)]^3$, $\underline{\theta} \in [H^{k+1}(\Omega)]^2$, $\underline{\underline{n}} \in [H^k(\Omega)]^{2\times 2}$ *and* $\underline{q} \in [H^k(\Omega)]^2$. *Then the finite element formulation* $\mathcal{S}_t^h$ *has a unique solution satisfying*

$$\|\vec{u} - \vec{u}^h, \underline{\theta} - \underline{\theta}^h\|_1 + \|\underline{\underline{n}} - \underline{\underline{n}}^h, \underline{q} - \underline{q}^h\|_{t,h} \leq Ch^k (\|\vec{u}, \underline{\theta}\|_{k+1} + \|\underline{\underline{n}}, \underline{q}\|_k). \tag{4.69}$$

Proof: Let $(I^h\vec{u}, I^h\underline{\theta}, I^h\underline{\underline{n}}, I^h\underline{q}) \in \mathcal{U}^h \times \mathcal{Q}^h$ be the interpolant to $(\vec{u}, \underline{\theta}, \underline{\underline{n}}, \underline{q})$ satisfying the estimates of Lemmas 4.5 and 4.6. By the stability result of Lemma 4.2 there exists $(\vec{v}, \underline{\eta}, \underline{\underline{p}}, \underline{r}) \in \mathcal{U}^h \times \mathcal{Q}^h$ such that

$$\|\vec{v}, \underline{\eta}, \underline{\underline{p}}, \underline{r}\|_{t,h} \leq 1 \tag{4.70}$$

and

$$C\|\vec{u}^h - I^h\vec{u}, \underline{\theta}^h - I^h\underline{\theta}, \underline{\underline{n}}^h - I^h\underline{\underline{n}}, \underline{q}^h - I^h\underline{q}\|_{t,h} \leq B_h(\vec{u}^h - I^h\vec{u}, \underline{\theta}^h - I^h\underline{\theta}, \underline{\underline{n}}^h - I^h\underline{\underline{n}}, \underline{q}^h - I^h\underline{q}; \vec{v}, \underline{\eta}, \underline{\underline{p}}, \underline{r}). \tag{4.71}$$

Next, the consistency of Lemma 4.1 gives

$$\begin{aligned} B_h(\vec{u}^h - I^h\vec{u}, \underline{\theta}^h - I^h\underline{\theta}, \underline{\underline{n}}^h - I^h\underline{\underline{n}}, \underline{q}^h - I^h\underline{q}; \vec{v}, \underline{\eta}, \underline{\underline{p}}, \underline{r}) \\ = B_h(\vec{u} - I^h\vec{u}, \underline{\theta} - I^h\underline{\theta}, \underline{\underline{n}} - I^h\underline{\underline{n}}, \underline{q} - I^h\underline{q}; \vec{v}, \underline{\eta}, \underline{\underline{p}}, \underline{r}). \end{aligned} \tag{4.72}$$

To proceed, we write out the expression for the bilinear form

$$\begin{aligned} B_h(\vec{u} - I^h\vec{u}, \underline{\theta} - I^h\underline{\theta}, \underline{\underline{n}} - I^h\underline{\underline{n}}, \underline{q} - I^h\underline{q}; \vec{v}, \underline{\eta}, \underline{\underline{p}}, \underline{r}) & \\ = A(\vec{u} - I^h\vec{u}, \underline{\theta} - I^h\underline{\theta}; \vec{v}, \underline{\eta}) + M(\vec{u} - I^h\vec{u}, \underline{\theta} - I^h\underline{\theta}, ; \underline{\underline{p}}, \underline{r}) & \\ + M(\vec{v}, \underline{\eta}; \underline{\underline{n}} - I^h\underline{\underline{n}}, \underline{q} - I^h\underline{q}) - N(\underline{\underline{n}} - I^h\underline{\underline{n}}, \underline{q} - I^h\underline{q}; \underline{\underline{p}}, \underline{r}) & \\ + S_h^e(\vec{u} - I^h\vec{u}, \underline{\theta} - I^h\underline{\theta}, \underline{\underline{n}} - I^h\underline{\underline{n}}, \underline{q} - I^h\underline{q}; \vec{v}, \underline{\eta}, \underline{\underline{p}}, \underline{r}) & \\ + S_h^c(\vec{u} - I^h\vec{u}, \underline{\theta} - I^h\underline{\theta}, \underline{\underline{n}} - I^h\underline{\underline{n}}, \underline{q} - I^h\underline{q}; \vec{v}, \underline{\eta}, \underline{\underline{p}}, \underline{r}). & \end{aligned} \tag{4.73}$$

For the first term above the Schwarz inequality, Lemma 4.5 and (4.70) give

$$|A(\vec{u} - I^h\vec{u}, \underline{\theta} - I^h\underline{\theta}; \vec{v}, \underline{\eta})| \leq C\|\vec{u} - I^h\vec{u}, \underline{\theta} - I^h\underline{\theta}\|_1\|\vec{v}, \underline{\eta}\|_1 \leq Ch^k\|\vec{u}, \underline{\theta}\|_{k+1}. \tag{4.74}$$

By using the integration by parts formulas (3.13), (3.14) and Lemma 4.5 we get

$$\begin{aligned} & |M(\vec{u} - I^h\vec{u}, \underline{\theta} - I^h\underline{\theta}, ; \underline{\underline{p}}, \underline{r})| \\ & \leq \Big(\sum_{E \in \Gamma_h \setminus \Gamma_0} h_E^{-1}\|\vec{u} - I^h\vec{u}\|_{0,E}^2 + \sum_{K \in \mathcal{C}_h} h_K^{-2}\|\vec{u} - I^h\vec{u}, \underline{\theta} - I^h\underline{\theta}\|_{0,K}^2\Big)^{1/2} \\ & \qquad \times \|\underline{\underline{p}}, \underline{r}\|_{t,h} \\ & \leq Ch^k\|\vec{u}, \underline{\theta}\|_{k+1}. \end{aligned} \tag{4.75}$$

Next, we directly get

$$|M(\vec{v}, \underline{\eta}; \underline{\underline{n}} - I^h\underline{\underline{n}}, \underline{q} - I^h\underline{q})| \leq C\|\underline{\underline{n}} - I^h\underline{\underline{n}}, \underline{q} - I^h\underline{q}\|_0\|\vec{v}, \underline{\eta}\|_1 \leq Ch^k\|\underline{\underline{n}}, \underline{q}\|_k \tag{4.76}$$

and

$$|N(\underline{\underline{n}} - I^h\underline{\underline{n}}, \underline{q} - I^h\underline{q}; \underline{\underline{p}}, \underline{r})| \leq Ct\|\underline{\underline{n}} - I^h\underline{\underline{n}}, \underline{q} - I^h\underline{q}\|_0\|\underline{\underline{p}}, \underline{r}\|_{t,h} \leq Ch^k\|\underline{\underline{n}}, \underline{q}\|_k. \tag{4.77}$$

The first stabilizing term is estimated using Lemmas 4.5 and 4.6 as follows

$$\begin{aligned} & |S_h^e(\vec{u} - I^h\vec{u}, \underline{\theta} - I^h\underline{\theta}, \underline{\underline{n}} - I^h\underline{\underline{n}}, \underline{q} - I^h\underline{q}; \vec{v}, \underline{\eta}, \underline{\underline{p}}, \underline{r})| \\ & \leq \Big(\Big(\sum_{K \in \mathcal{C}_h} h_K^2\|\vec{u} - I^h\vec{u}, \underline{\theta} - I^h\underline{\theta}\|_{2,K}^2\Big)^{1/2} + \|\underline{\underline{n}} - I^h\underline{\underline{n}}, \underline{q} - I^h\underline{q}\|_{t,h}\Big) \\ & \qquad \times \Big(\Big(\sum_{K \in \mathcal{C}_h} h_K^2\|\vec{v}, \underline{\eta}\|_{2,K}^2\Big)^{1/2} + \|\underline{\underline{p}}, \underline{r}\|_{t,h}\Big) \\ & \leq Ch^k(\|\vec{u}, \underline{\theta}\|_{k+1} + \|\underline{\underline{n}}, \underline{q}\|_k), \end{aligned} \tag{4.78}$$

where we used the inverse inequality (4.68). For the second stabilizing term we obtain

$$\begin{aligned}
&|S_h^c(\vec{u}-I^h\vec{u},\underline{\theta}-I^h\underline{\theta},\underline{\underline{n}}-I^h\underline{\underline{n}},\underline{q}-I^h\underline{q};\vec{v},\underline{\eta},\underline{\underline{p}},\underline{r})| \\
&\quad\leq Ct^2\Big(|M(\vec{u}-I^h\vec{u},\underline{\theta}-I^h\underline{\theta},;\underline{\underline{p}},\underline{r})|+|M(\vec{v},\underline{\eta};\underline{\underline{n}}-I^h\underline{\underline{n}},\underline{q}-I^h\underline{q})| \\
&\qquad\quad +|N(\underline{\underline{n}}-I^h\underline{\underline{n}},\underline{q}-I^h\underline{q};\underline{\underline{p}},\underline{r})|\Big)+C\|\vec{u}-I^h\vec{u},\underline{\theta}-I^h\underline{\theta}\|_1\|\vec{v},\underline{\eta}\|_1 \\
&\quad\leq Ch^k(\|\vec{u},\underline{\theta}\|_{k+1}+\|\underline{\underline{n}},\underline{q}\|_k),
\end{aligned} \tag{4.79}$$

by virtue of the previous estimates (4.75)–(4.77), the Schwarz inequality and Lemma 4.5. Finally, we note that Lemmas 4.5 and 4.6 also give

$$\|\vec{u}-I^h\vec{u},\underline{\theta}-I^h\underline{\theta},\underline{\underline{n}}-I^h\underline{\underline{n}},\underline{q}-I^h\underline{q}\|_{t,h}\leq Ch^k(\|\vec{u},\underline{\theta}\|_{k+1}+\|\underline{\underline{n}},\underline{q}\|_k). \tag{4.80}$$

Hence, collecting (4.71) to (4.80) and using the triangle inequality we get the asserted estimate. ■

We will close this section by giving some remarks on the analysis of the method. In [2] the analysis is performed using an abstract semi-norm for the stresses, instead of the discrete norm we have used. This abstract semi-norm is defined as follows

$$|||\underline{\underline{p}},\underline{r}|||\stackrel{\text{def}}{=}\sup_{(\vec{v},\underline{\eta})\in\mathcal{U}}\frac{M(\vec{v},\underline{\eta};\underline{\underline{p}},\underline{r})}{\|\vec{v},\underline{\eta}\|_1}. \tag{4.81}$$

For our finite element method it is possible to prove the stability using the same global norm as in [2], i.e.

$$|||\vec{v},\underline{\eta},\underline{\underline{p}},\underline{r}|||_t\stackrel{\text{def}}{=}\left(\|\vec{v},\underline{\eta}\|_1^2+t^2\|\underline{\underline{p}},\underline{r}\|_0^2+|||\underline{\underline{p}},\underline{r}|||^2\right)^{1/2}. \tag{4.82}$$

This is achieved using the "Pitkäranta-Verfürth trick" introduced by Pitkäranta in connection with Babuška's method for approximating Dirichlet boundary conditions [38] and Verfürth for mixed methods for the Stokes equation [51]. The idea has later been extensively used for both classical and stabilized formulations of the Stokes problem, cf. [46, 21].

We are, however, not able to carry through the whole error analysis with this norm. The reason is that the stabilizing term (4.78) which we are not able to bound.

One possibility for an analysis would be to use the combined norm $|||\cdot|||+\|\cdot\|_{t,h}$ for the stresses. This is possible since the same stability construction can be shown to yield the stability with respect to both norms for the stresses. We choose a different approach and show that the estimate for the stresses in the norm (4.81) can be proven "a posteriori".

Theorem 4.2 *Under the assumptions of Theorem* 4.1 *it also holds*

$$|||\underline{\underline{n}}-\underline{\underline{n}}^h,\underline{q}-\underline{q}^h|||\leq Ch^k(\|\vec{u},\underline{\theta}\|_{k+1}+\|\underline{\underline{n}},\underline{q}\|_k). \tag{4.83}$$

Proof: From the definition of the semi-norm (4.81) there is $(\vec{v},\underline{\eta}) \in \mathcal{U} \times \mathcal{Q}$ such that

$$|||\underline{\underline{n}} - \underline{\underline{n}}^h, \underline{q} - \underline{q}^h||| = M(\vec{v},\underline{\eta};\underline{\underline{n}} - \underline{\underline{n}}^h, \underline{q} - \underline{q}^h) \tag{4.84}$$

and

$$\|\vec{v},\underline{\eta}\|_1 = 1. \tag{4.85}$$

Now we let $(\vec{v}^h,\underline{\eta}^h) \in \mathcal{U}^h$ be the Clément interpolant to $(\vec{v},\underline{\eta})$. From [17] we have

$$\begin{aligned}\sum_{K\in\mathcal{C}_h} h_K^{-2}\|\vec{v}-\vec{v}^h, \underline{\eta}-\underline{\eta}^h\|_{0,K}^2 + \sum_{E\in\Gamma_h\setminus\Gamma_0} h_E^{-1}\|\vec{v}-\vec{v}^h\|_{0,E}^2 \\ +\|\vec{v}-\vec{v}^h, \underline{\eta}-\underline{\eta}^h\|_1^2 \le C\|\vec{v},\underline{\eta}\|_1^2.\end{aligned} \tag{4.86}$$

Next, we write

$$\begin{aligned} M(\vec{v},\underline{\eta};\underline{\underline{n}} - \underline{\underline{n}}^h, \underline{q} - \underline{q}^h) &= M(\vec{v}-\vec{v}^h, \underline{\eta}-\underline{\eta}^h; \underline{\underline{n}} - \underline{\underline{n}}^h, \underline{q} - \underline{q}^h) \\ &\quad + M(\vec{v}^h, \underline{\eta}^h; \underline{\underline{n}} - \underline{\underline{n}}^h, \underline{q} - \underline{q}^h). \end{aligned} \tag{4.87}$$

For the first term above an integration by part using (3.13) and (3.14) gives

$$\begin{aligned} &M(\vec{v}-\vec{v}^h, \underline{\eta}-\underline{\eta}^h; \underline{\underline{n}} - \underline{\underline{n}}^h, \underline{q} - \underline{q}^h) \\ &\le C\Big(\sum_{K\in\mathcal{C}_h} h_K^{-2}\|\vec{v}-\vec{v}^h, \underline{\eta}-\underline{\eta}^h\|_{0,K}^2 + \sum_{E\in\Gamma_h\setminus\Gamma_0} h_E^{-1}\|\vec{v}-\vec{v}^h\|_{0,E}^2\Big)^{1/2} \\ &\qquad \times \|\underline{\underline{n}} - \underline{\underline{n}}^h, \underline{q} - \underline{q}^h\|_{t,h} \\ &\le C\|\vec{v},\underline{\eta}\|_1 \|\underline{\underline{n}} - \underline{\underline{n}}^h, \underline{q} - \underline{q}^h\|_{t,h} \\ &\le C\|\underline{\underline{n}} - \underline{\underline{n}}^h, \underline{q} - \underline{q}^h\|_{t,h}. \end{aligned} \tag{4.88}$$

From the consistency we have

$$B_h(\vec{u}-\vec{u}^h, \underline{\theta}-\underline{\theta}^h, \underline{\underline{n}} - \underline{\underline{n}}^h, \underline{q} - \underline{q}^h; \vec{v}^h, \underline{\eta}^h, \underline{\underline{0}}, \underline{0}) = 0. \tag{4.89}$$

Hence, it holds

$$\begin{aligned} &M(\vec{v}^h, \underline{\eta}^h; \underline{\underline{n}} - \underline{\underline{n}}^h, \underline{q} - \underline{q}^h) \\ &= -A(\vec{u}-\vec{u}^h, \underline{\theta}-\underline{\theta}^h; \vec{v}^h, \underline{\eta}^h) - S_h^e(\vec{u}-\vec{u}^h, \underline{\theta}-\underline{\theta}^h, \underline{\underline{n}} - \underline{\underline{n}}^h, \underline{q} - \underline{q}^h; \vec{v}^h, \underline{\eta}^h, \underline{\underline{0}}, \underline{0}) \\ &\quad -S_h^c(\vec{u}-\vec{u}^h, \underline{\theta}-\underline{\theta}^h, \underline{\underline{n}} - \underline{\underline{n}}^h, \underline{q} - \underline{q}^h; \vec{v}^h, \underline{\eta}^h, \underline{\underline{0}}, \underline{0}). \end{aligned} \tag{4.90}$$

For the first term above we get

$$|A(\vec{u}-\vec{u}^h, \underline{\theta}-\underline{\theta}^h; \vec{v}^h, \underline{\eta}^h)| \le C\|\vec{u}-\vec{u}^h, \underline{\theta}-\underline{\theta}^h\|_1 \|\vec{v}^h, \underline{\eta}^h\|_1. \tag{4.91}$$

The second term is estimated as follows

$$|S_h^e(\vec{u}-\vec{u}^h, \underline{\theta}-\underline{\theta}^h, \underline{\underline{n}} - \underline{\underline{n}}^h, \underline{q} - \underline{q}^h; \vec{v}^h, \underline{\eta}^h, \underline{\underline{0}}, \underline{0})|$$

$$
\begin{aligned}
&\leq C\Big(\big(\sum_{K\in\mathcal{C}_h} h_K^2\|\vec{u}-\vec{u}^h,\underline{\theta}-\underline{\theta}^h\|_{2,K}^2\big)^{1/2}+\|\underline{\underline{n}}-\underline{\underline{n}}^h,\underline{q}-\underline{q}^h\|_{t,h}\Big)\\
&\qquad\qquad\times\Big(\sum_{K\in\mathcal{C}_h} h_K^2\|\vec{v}^h,\underline{\eta}^h\|_{2,K}^2\Big)^{1/2}\\
&\leq C\Big(\big(\sum_{K\in\mathcal{C}_h} h_K^2\|\vec{u}-\vec{u}^h,\underline{\theta}-\underline{\theta}^h\|_{2,K}^2\big)^{1/2}+\|\underline{\underline{n}}-\underline{\underline{n}}^h,\underline{q}-\underline{q}^h\|_{t,h}\Big)\\
&\qquad\qquad\times\|\vec{v}^h,\underline{\eta}^h\|_1 \qquad (4.92)
\end{aligned}
$$

where we used the inverse inequality (4.68). For the third term we directly get

$$
\begin{aligned}
&|S_h^c(\vec{u}-\vec{u}^h,\underline{\theta}-\underline{\theta}^h,\underline{\underline{n}}-\underline{\underline{n}}^h,\underline{q}-\underline{q}^h;\vec{v}^h,\underline{\eta}^h,\underline{0},\underline{0})| \qquad (4.93)\\
&\qquad\leq C(\|\vec{u}-\vec{u}^h,\underline{\theta}-\underline{\theta}^h\|_1+t^2\|\underline{\underline{n}}-\underline{\underline{n}}^h,\underline{q}-\underline{q}^h\|_0)\|\vec{v}^h,\underline{\eta}^h\|_1,
\end{aligned}
$$

Combining the above estimates gives

$$
\begin{aligned}
|||\underline{\underline{n}}-\underline{\underline{n}}^h,\underline{q}-\underline{q}^h||| \leq C\Big(&\|\vec{u}-\vec{u}^h,\underline{\theta}-\underline{\theta}^h,\underline{\underline{n}}-\underline{\underline{n}}^h,\underline{q}-\underline{q}^h\|_{t,h} \qquad (4.94)\\
&+\big(\sum_{K\in\mathcal{C}_h} h_K^2\|\vec{u}-\vec{u}^h,\underline{\theta}-\underline{\theta}^h\|_{2,K}^2\big)^{1/2}\Big).
\end{aligned}
$$

From Lemma 4.5, the inverse estimate (4.68) and Theorem 4.1 it follows that

$$
\big(\sum_{K\in\mathcal{C}_h} h_K^2\|\vec{u}-\vec{u}^h,\underline{\theta}-\underline{\theta}^h\|_{2,K}^2\big)^{1/2} \leq Ch^k(\|\vec{u},\underline{\theta}\|_{k+1}+\|\underline{\underline{n}},\underline{q}\|_k). \qquad (4.95)
$$

Hence, the assertion follows from Theorem 4.1. ■

5 A second stabilized method

In this section, we propose an alternative stabilized scheme, obtained from the previous formulation by changing the weights in front of the stabilizing terms that originate from the constitutive equations, in the spirit of what is considered for plates in [39, 27, 47, 34, 33, 32]. This second stabilized method reads:

$\tilde{\mathcal{S}}_t^h$: *Find* $(\vec{u}^h,\underline{\theta}^h,\underline{\underline{n}}^h,\underline{q}^h)\in\mathcal{U}^h\times\mathcal{Q}^h$ *such that*

$$
\tilde{B}_h(\vec{u}^h,\underline{\theta}^h,\underline{\underline{n}}^h,\underline{q}^h;\vec{v},\underline{\eta},\underline{\underline{p}},\underline{r}) = F_h(\vec{v},\underline{\eta},\underline{\underline{p}},\underline{r}) \quad \forall(\vec{v},\underline{\eta},\underline{\underline{p}},\underline{r})\in\mathcal{U}^h\times\mathcal{Q}^h, \qquad (5.1)
$$

where

$$
\begin{aligned}
\tilde{B}_h(\vec{w},\underline{\tau},\underline{\underline{k}},\underline{s};\vec{v},\underline{\eta},\underline{\underline{p}},\underline{r}) \stackrel{\text{def}}{=}\ & B(\vec{w},\underline{\tau},\underline{\underline{k}},\underline{s};\vec{v},\underline{\eta},\underline{\underline{p}},\underline{r})+S_h^e(\vec{w},\underline{\tau},\underline{\underline{k}},\underline{s};\vec{v},\underline{\eta},\underline{\underline{p}},\underline{r})\\
&+\tilde{S}_h^c(\vec{w},\underline{\tau},\underline{\underline{k}},\underline{s};\vec{v},\underline{\eta},\underline{\underline{p}},\underline{r}), \qquad (5.2)
\end{aligned}
$$

with the stabilizing term S_h^e originating from the equilibrium equations defined as before in (4.6), and the term from the constitutive equations redefined by

$$\begin{aligned}
&\tilde{S}_h^c(\vec{w},\underline{\tau},\underline{\underline{k}},\underline{s};\vec{v},\underline{\eta},\underline{\underline{p}},\underline{r}) \qquad (5.3)\\
&\quad\stackrel{\text{def}}{=}\ \frac{1}{2}\sum_{K\in\mathcal{C}_h}\frac{1}{t^2+\alpha_4 h_K^2}\int_K\left[\underline{\underline{\varepsilon}}(\vec{w})-t^2\breve{\underline{\underline{E}}}:\underline{\underline{k}}\right]:\underline{\underline{E}}:\left[\underline{\underline{\varepsilon}}(\vec{v})-t^2\breve{\underline{\underline{E}}}:\underline{\underline{p}}\right]\,dS\\
&\qquad+\frac{1}{2}\sum_{K\in\mathcal{C}_h}\frac{1}{t^2+\alpha_5 h_K^2}\int_K\left[\underline{\gamma}(\vec{w},\underline{\tau})-t^2\breve{\underline{\underline{G}}}\cdot\underline{s}\right]\cdot\underline{\underline{G}}\cdot\left[\underline{\gamma}(\vec{v},\underline{\eta})-t^2\breve{\underline{\underline{G}}}\cdot\underline{r}\right]\,dS.
\end{aligned}$$

The right hand side F_h remains unchanged as defined by (4.8).

For the displacement variables we will use the following norm

$$|\!|\!|\vec{v},\underline{\eta}|\!|\!|_{t,h}\stackrel{\text{def}}{=}\left(\|\vec{v},\underline{\eta}\|_1^2+\sum_{K\in\mathcal{C}_h}\frac{1}{t^2+h_K^2}(\|\underline{\gamma}(\vec{v},\underline{\eta})\|_{0,K}^2+\|\underline{\underline{\varepsilon}}(\vec{v})\|_{0,K}^2)\right)^{1/2},$$

which is obviously stronger than the H^1-norm used for our first method. The norm for the stresses we now have to weaken by first redefining the "jump" term

$$\|\underline{\underline{p}},\underline{r}\|_J\stackrel{\text{def}}{=}\Big(\sum_{E\in\Gamma_h\setminus\Gamma_0}h_E^3\int_E\left[[\![\underline{\underline{p}}\cdot\underline{\nu}]\!]^2+[\![\underline{r}\cdot\underline{\nu}]\!]^2\right]\,ds\Big)^{1/2}$$

and then defining

$$|\!|\!|\underline{\underline{p}},\underline{r}|\!|\!|_{t,h}\stackrel{\text{def}}{=}\left(t^2\|\underline{\underline{p}},\underline{r}\|_0^2+|\underline{\underline{p}},\underline{r}|_I^2+\|\underline{\underline{p}},\underline{r}\|_J^2\right)^{1/2},$$

by keeping the interior part as defined in (4.11). We will also use the combined notation

$$|\!|\!|\vec{v},\underline{\eta},\underline{\underline{p}},\underline{r}|\!|\!|_{t,h}\stackrel{\text{def}}{=}\left(|\!|\!|\vec{v},\underline{\eta}|\!|\!|_{t,h}^2+|\!|\!|\underline{\underline{p}},\underline{r}|\!|\!|_{t,h}^2\right)^{1/2}.$$

The analysis of this new scheme will now essentially follow the same lines as in Section 4, so there is no need to give the full proofs for all results. We will instead highlight the differences between the two formulations, whenever these differences are significant.

First, we again note that the formulation is consistent.

Lemma 5.1 *Suppose that $\vec{f}\in[L^2(\Omega)]^3$. Then the solution $(\vec{u},\underline{\theta},\underline{\underline{n}},\underline{q})\in\mathcal{U}\times\mathcal{Q}$ of $\mathcal{M}_t$ satisfies*

$$\tilde{B}_h(\vec{u},\underline{\theta},\underline{\underline{n}},\underline{q};\vec{v},\underline{\eta},\underline{\underline{p}},\underline{r})=F_h(\vec{v},\underline{\eta},\underline{\underline{p}},\underline{r})\quad\forall(\vec{v},\underline{\eta},\underline{\underline{p}},\underline{r})\in\mathcal{U}^h\times\mathcal{Q}^h.\quad\blacksquare \qquad (5.4)$$

The finite element subspace $\mathcal{U}^h\times\mathcal{Q}^h$ we keep as defined by (4.16) and (4.17) in Section 4. With this choice we can state the new stability result.

Lemma 5.2 *Assume that:*

($\tilde{\text{S}}1$) $0 < \alpha_i < C_i^I$, *for* $i = 1, 2$, *where the positive constants* C_i^I *are fixed, derived from inverse inequalities.*

($\tilde{\text{S}}2$) $\alpha_i > 0$, *for* $i = 3, 4, 5$.

($\tilde{\text{S}}3$) h *is sufficiently small.*

Then the stabilized formulation $\tilde{\mathcal{S}}_t^h$ *is stable, i.e. there is a positive constant* C *such that, for all* $(\vec{w}, \underline{\tau}, \underline{\underline{k}}, \underline{s}) \in \mathcal{U}^h \times \mathcal{Q}^h$, *there exist* $(\vec{z}, \underline{\eta}, \underline{\underline{p}}, \underline{r}) \in \mathcal{U}^h \times \mathcal{Q}^h$ *with*

$$\tilde{B}_h(\vec{w}, \underline{\tau}, \underline{\underline{k}}, \underline{s}; \vec{z}, \underline{\eta}, \underline{\underline{p}}, \underline{r}) \geq C [\![\vec{w}, \underline{\tau}, \underline{\underline{k}}, \underline{s}]\!]_{t,h} \tag{5.5}$$

and

$$[\![\vec{z}, \underline{\eta}, \underline{\underline{p}}, \underline{r}]\!]_{t,h} \leq 1. \tag{5.6}$$

Proof: We again go through 3 steps.
Step 1. First, a direct substitution yields

$$\begin{aligned}
&\tilde{B}_h(\vec{w}, \underline{\tau}, \underline{\underline{k}}, \underline{s}; \vec{w}, \underline{\tau}, -\underline{\underline{k}}, -\underline{s}) \\
&\quad = A(\vec{w}, \underline{\tau}; \vec{w}, \underline{\tau}) + \frac{1}{2} \sum_{K \in \mathcal{C}_h} \frac{1}{t^2 + \alpha_4 h_K^2} \int_K \underline{\underline{\varepsilon}}(\vec{w}) : \underline{\underline{\underline{\underline{E}}}} : \underline{\underline{\varepsilon}}(\vec{w})\, dS \\
&\qquad + \frac{1}{2} \sum_{K \in \mathcal{C}_h} \frac{1}{t^2 + \alpha_5 h_K^2} \int_K \underline{\gamma}(\vec{w}, \underline{\tau}) \cdot \underline{\underline{G}} \cdot \underline{\gamma}(\vec{w}, \underline{\tau})\, dS \\
&\qquad + \frac{t^2}{2} \sum_{K \in \mathcal{C}_h} \left(1 + \frac{\alpha_4 h_K^2}{t^2 + \alpha_4 h_K^2}\right) \int_K \underline{\underline{k}} : \breve{\underline{\underline{\underline{\underline{E}}}}} : \underline{\underline{k}}\, dS \\
&\qquad + \frac{t^2}{2} \sum_{K \in \mathcal{C}_h} \left(1 + \frac{\alpha_5 h_K^2}{t^2 + \alpha_5 h_K^2}\right) \int_K \underline{s} \cdot \breve{\underline{\underline{G}}} \cdot \underline{s}\, dS \\
&\qquad + \alpha_1 \sum_{K \in \mathcal{C}_h} h_K^2 \left(\|\underline{s}\|_{0,K}^2 - \|\underline{\mathrm{div}}\, \underline{\underline{m}}(\vec{w}, \underline{\tau})\|_{0,K}^2 \right) \\
&\qquad + \alpha_2 \sum_{K \in \mathcal{C}_h} h_K^2 \left(\|\underline{\mathrm{div}}\, \underline{\underline{k}} - \underline{b} \cdot \underline{s}\|_{0,K}^2 - \|\underline{\mathrm{div}}\, \left[\underline{b} \cdot \underline{\underline{m}}(\vec{w}, \underline{\tau})\right] \|_{0,K}^2 \right) \\
&\qquad + \alpha_3 \sum_{K \in \mathcal{C}_h} h_K^2 \left(\|\mathrm{div}\, \underline{s} + \underline{\underline{b}} : \underline{\underline{k}}\|_{0,K}^2 - \|\underline{\underline{c}} : \underline{\underline{m}}(\vec{w}, \underline{\tau})\|_{0,K}^2 \right).
\end{aligned}$$

Note that, for any K,

$$\frac{1}{\max\{1, \alpha_i\}} \frac{1}{t^2 + h_K^2} \leq \frac{1}{t^2 + \alpha_i h_K^2} \leq \frac{1}{\min\{1, \alpha_i\}} \frac{1}{t^2 + h_K^2}, \qquad i = 4, 5. \tag{5.7}$$

Using as before estimates (4.21), (4.22) and (4.25), we thus obtain

$$\begin{aligned}
&\tilde{B}_h(\vec{w}, \underline{\tau}, \underline{\underline{k}}, \underline{s}; \vec{w}, \underline{\tau}, -\underline{\underline{k}}, -\underline{s}) \\
&\quad \geq (1 - \alpha_1 C_1 - \alpha_2 C_2 - C h^2) A(\vec{w}, \underline{\tau}; \vec{w}, \underline{\tau}) - C h^2 \|\vec{w}, \underline{\tau}\|_0^2
\end{aligned} \tag{5.8}$$

$$+C_3\Big\{t^2\|\underline{\underline{k}},\underline{s}\|_0^2+\sum_{K\in\mathcal{C}_h}\frac{1}{t^2+h_K^2}(\|\underline{\gamma}(\vec{w},\underline{\tau})\|_{0,K}^2+\|\underline{\underline{\varepsilon}}(\vec{w})\|_{0,K}^2)\Big\}$$
$$+\alpha_1\sum_{K\in\mathcal{C}_h}h_K^2\|\underline{s}\|_{0,K}^2+\alpha_2\sum_{K\in\mathcal{C}_h}h_K^2\|\underline{\mathrm{div}}\,\underline{\underline{k}}-\underline{b}\cdot\underline{s}\|_{0,K}^2$$
$$+\alpha_3\sum_{K\in\mathcal{C}_h}h_K^2\|\mathrm{div}\,\underline{s}+\underline{\underline{b}}:\underline{\underline{k}}\|_{0,K}^2.$$

We again set

$$C_i^I\stackrel{\mathrm{def}}{=}(2C_i)^{-1},\ i=1,2.$$

Then, proceeding like in the proof of Lemma 4.2, we can conclude that

$$\tilde{B}_h(\vec{w},\underline{\tau},\underline{\underline{k}},\underline{s};\vec{w},\underline{\tau},-\underline{\underline{k}},-\underline{s})\geq C_4\left(|\!|\!|\vec{w},\underline{\tau}|\!|\!|_{h,t}^2+t^2\|\underline{\underline{k}},\underline{s}\|_0^2+|\underline{\underline{k}},\underline{s}|_I^2\right).\tag{5.9}$$

Step 2. By changing the weights from h_E to h_E^3 in (4.60), Lemma 4.4 changes as follows: there exists $(\vec{v},\underline{0})\in\mathcal{U}^h$ such that

$$\sum_{E\in\Gamma_h\setminus\Gamma_0}\int_E(\llbracket\underline{\underline{k}}\cdot\underline{\nu}\rrbracket\cdot\underline{v}+\llbracket\underline{s}\cdot\underline{\nu}\rrbracket\,v_3)\,ds\geq C_5\|\underline{\underline{k}},\underline{s}\|_J^2\tag{5.10}$$

and

$$\sum_{K\in\mathcal{C}_h}(h_K^{-4}\|\vec{v}\|_{0,K}^2+h_K^{-2}\|\vec{v}\|_{1,K}^2)\leq\|\underline{\underline{k}},\underline{s}\|_J^2.\tag{5.11}$$

We first note that

$$|\!|\!|\vec{v},\underline{0}|\!|\!|_{t,h}\leq C_6\Big(\sum_{K\in\mathcal{C}_h}h_K^{-2}\|\vec{v}\|_{1,K}^2\Big)^{1/2}\leq C_6\|\underline{\underline{k}},\underline{s}\|_J.\tag{5.12}$$

By substituting this $\vec{v}$ we get

$$\tilde{B}_h(\vec{w},\underline{\tau},\underline{\underline{k}},\underline{s};\vec{v},\underline{0},\underline{0},\underline{0})=I+II+III\tag{5.13}$$

with

$$\begin{aligned}I\ \stackrel{\mathrm{def}}{=}\ & A(\vec{w},\underline{\tau};\vec{v},\underline{0})+\frac{1}{2}\sum_{K\in\mathcal{C}_h}\frac{1}{t^2+\alpha_4h_K^2}\int_K\underline{\underline{\varepsilon}}(\vec{w}):\underline{\underline{\underline{\underline{E}}}}:\underline{\underline{\varepsilon}}(\vec{v})\,dS\\
&+\frac{1}{2}\sum_{K\in\mathcal{C}_h}\frac{1}{t^2+\alpha_5h_K^2}\int_K\underline{\gamma}(\vec{w},\underline{\tau})\cdot\underline{G}\cdot\underline{\gamma}(\vec{v},\underline{0})\,dS\\
&-\alpha_1\left\langle\underline{\mathrm{div}}\,\underline{\underline{m}}(\vec{w},\underline{\tau})-\underline{s},\underline{\mathrm{div}}\,\underline{\underline{m}}(\vec{v},\underline{0})\right\rangle_h\\
&-\alpha_2\left\langle\underline{\mathrm{div}}\,\underline{\underline{\tilde{n}}}(\vec{w},\underline{\tau},\underline{\underline{k}})-\underline{\underline{b}}\cdot\underline{s},\underline{\mathrm{div}}\,\underline{\underline{\tilde{n}}}(\vec{v},\underline{0},\underline{0})\right\rangle_h\\
&-\alpha_3\left\langle\mathrm{div}\,\underline{s}+\underline{\underline{b}}:\underline{\underline{\tilde{n}}}(\vec{w},\underline{\tau},\underline{\underline{k}}),\underline{\underline{b}}:\underline{\underline{\tilde{n}}}(\vec{v},\underline{0},\underline{0})\right\rangle_h,\end{aligned}$$

$$II\stackrel{\mathrm{def}}{=}\int_\Omega\underline{\underline{\varepsilon}}(\vec{v}):\underline{\underline{k}}\,dS+\int_\Omega\underline{\gamma}(\vec{v},\underline{0})\cdot\underline{s}\,dS,\tag{5.14}$$

and

$$III \stackrel{\text{def}}{=} -\frac{1}{2}\sum_{K\in\mathcal{C}_h}\frac{t^2}{t^2+\alpha_4 h_K^2}\int_K \underline{\underline{\varepsilon}}(\vec{v}):\underline{\underline{k}}\,dS - \frac{1}{2}\sum_{K\in\mathcal{C}_h}\frac{t^2}{t^2+\alpha_5 h_K^2}\int_K \underline{\gamma}(\vec{v},\underline{0})\cdot\underline{s}\,dS. \quad (5.15)$$

Using the Schwarz inequality and estimates (4.40)-(4.41), then (5.12), we obtain

$$|I| \le C_7 [\![\vec{v},\underline{0}]\!]_{t,h}\Big([\![\vec{w},\underline{\tau}]\!]_{h,t}^2 + |\underline{\underline{k}},\underline{s}|_I^2\Big)^{1/2} \le C_8\|\underline{\underline{k}},\underline{s}\|_J\Big([\![\vec{w},\underline{\tau}]\!]_{h,t}^2 + |\underline{\underline{k}},\underline{s}|_I^2\Big)^{1/2}. \quad (5.16)$$

Further, recalling (5.7), we have

$$\begin{aligned}
|III| &\le C_9\Big(\sum_{K\in\mathcal{C}_h}\frac{t^2}{t^2+h_K^2}\Big|\int_K \underline{\underline{\varepsilon}}(\vec{v}):\underline{\underline{k}}\,dS\Big| + \sum_{K\in\mathcal{C}_h}\frac{t^2}{t^2+h_K^2}\Big|\int_K \underline{\gamma}(\vec{v},\underline{0})\cdot\underline{s}\,dS\Big|\Big)\\
&\le C_9\Big(\sum_{K\in\mathcal{C}_h}\sqrt{\frac{t^2}{t^2+h_K^2}}\Big|\int_K \underline{\underline{\varepsilon}}(\vec{v}):\underline{\underline{k}}\,dS\Big| + \sum_{K\in\mathcal{C}_h}\sqrt{\frac{t^2}{t^2+h_K^2}}\Big|\int_K \underline{\gamma}(\vec{v},\underline{0})\cdot\underline{s}\,dS\Big|\Big)\\
&\le C_9[\![\vec{v},\underline{0}]\!]_{t,h}\times t\|\underline{\underline{k}},\underline{s}\|_0 \le C_{10}\|\underline{\underline{k}},\underline{s}\|_J\times t\|\underline{\underline{k}},\underline{s}\|_0, \qquad (5.17)
\end{aligned}$$

using the Schwarz inequality and (5.12). In the second term we again integrate by parts using (3.13) and (3.14) , which gives

$$II = IV + V, \quad (5.18)$$

with

$$IV \stackrel{\text{def}}{=} \sum_{E\in\Gamma_h\setminus\Gamma_0}\int_E \left([\![\underline{\underline{k}}\cdot\underline{\nu}]\!]\cdot\underline{v} + [\![\underline{s}\cdot\underline{\nu}]\!]\,v_3\right)ds, \quad (5.19)$$

and

$$V \stackrel{\text{def}}{=} -\sum_{K\in\mathcal{C}_h}\int_K\Big(\underline{\operatorname{div}}\,\underline{\underline{k}} - \underline{\underline{b}}\cdot\underline{s}\Big)\cdot\underline{v}\,dS - \sum_{K\in\mathcal{C}_h}\int_K\Big(\operatorname{div}\underline{s} + \underline{\underline{b}}:\underline{\underline{k}}\Big)\,v_3\,dS. \quad (5.20)$$

Equation (5.11) implies

$$|V| \le |\underline{\underline{k}},\underline{s}|_I\Big(\sum_{K\in\mathcal{C}_h}h_K^{-2}\|\vec{v}\|_{0,K}^2\Big)^{1/2} \le |\underline{\underline{k}},\underline{s}|_I\|\underline{\underline{k}},\underline{s}\|_J. \quad (5.21)$$

Hence, recalling (5.10), the arithmetic-geometric mean inequality gives

$$II \ge \frac{C_5}{2}\|\underline{\underline{k}},\underline{s}\|_J^2 - \frac{1}{2C_5}|\underline{\underline{k}},\underline{s}|_I^2. \quad (5.22)$$

As in the proof of Lemma 4.2, we now combine (5.13)–(5.22) and use the arithmetic-geometric mean inequality to obtain

$$\tilde{B}_h(\vec{w},\underline{\tau},\underline{\underline{k}},\underline{s};\vec{v},\underline{0},\underline{\underline{0}},\underline{0}) \ge C_{11}\|\underline{\underline{k}},\underline{s}\|_J^2 - C_{12}\Big([\![\vec{w},\underline{\tau}]\!]_{t,h}^2 + t^2\|\underline{\underline{k}},\underline{s}\|_0^2 + |\underline{\underline{k}},\underline{s}|_I^2\Big). \quad (5.23)$$

Step 3. The assertion again follows from (5.9) and (5.23) by choosing $(\vec{z},\underline{\eta},\underline{\underline{p}},\underline{r}) = (\vec{w}+\delta\vec{v},\underline{\tau},-\underline{\underline{k}},-\underline{s})$ with $\delta>0$ small enough. ■

We now examine the modifications in the interpolation estimates. In view of the stronger norm used for the displacements, a stronger assumption on the regularity of $\vec{u}$ is required.

Lemma 5.3 *For* $(\vec{u},\underline{\theta}) \in \mathcal{U} \cap \left([H^{k+2}(\Omega)]^3 \times [H^{k+1}(\Omega)]^2\right)$ *it holds*

$$\inf_{(\vec{v},\underline{\eta})\in\mathcal{U}^h} \Big\{ [\![\vec{u}-\vec{v}, \underline{\theta}-\underline{\eta}]\!]_{t,h}^2 + \sum_{E\in\Gamma_h\setminus\Gamma_0} h_E^{-3} \|\vec{u}-\vec{v}\|_{0,E}^2 \tag{5.24}$$
$$+ \sum_{K\in\mathcal{C}_h} (h_K^{-2}\|\vec{u}-\vec{v},\underline{\theta}-\underline{\eta}\|_{0,K}^2 + h_K^2 \|\vec{u}-\vec{v},\underline{\theta}-\underline{\eta}\|_{2,K}^2) \Big\}^{1/2}$$
$$\leq C h^k (\|\vec{u}\|_{k+2} + \|\underline{\theta}\|_{k+1}) \ .$$

Proof: Using the definition of the norm $[\![\cdot]\!]_{t,h}$, we have

$$[\![\vec{u}-\vec{v}, \underline{\theta}-\underline{\eta}]\!]_{t,h}^2 \leq C\Big(\|\vec{u}-\vec{v},\underline{\theta}-\underline{\eta}\|_1^2 + \sum_{K\in\mathcal{C}_h} h_K^{-2} (\|\vec{u}-\vec{v}\|_1^2 + \|\underline{\theta}-\underline{\eta}\|_0^2) \Big).$$

Therefore, recalling that the interpolation space is one degree higher for $\vec{u}$ than for $\underline{\theta}$, scaling arguments imply

$$[\![\vec{u}-\vec{v}, \underline{\theta}-\underline{\eta}]\!]_{t,h} \leq C h^k (\|\vec{u}\|_{k+2} + \|\underline{\theta}\|_{k+1}).$$

The next term in (5.24) is also treated by standard scaling arguments, and the remaining terms are unchanged from Lemma 4.5. ■

By contrast, the new norm for the stresses is weaker than the previous one, so the interpolation estimates remain valid.

Lemma 5.4 *For* $(\underline{\underline{n}},\underline{q}) \in \mathcal{Q} \cap \left([H^k(\Omega)]^{2\times 2} \times [H^k(\Omega)]^2\right)$ *it holds*

$$\inf_{(\underline{\underline{p}},\underline{r})\in\mathcal{Q}^h} \Big\{ [\![\underline{\underline{n}}-\underline{\underline{p}}, \underline{q}-\underline{r}]\!]_{t,h} + \|\underline{\underline{n}}-\underline{\underline{p}}, \underline{q}-\underline{r}\|_0 \Big\} \leq C h^k \|\underline{\underline{n}}, \underline{q}\|_k \ . \quad \blacksquare \tag{5.25}$$

We can now state and prove our final approximation result.

Theorem 5.1 *Assume that conditions* ($\tilde{\text{S}}1$)–($\tilde{\text{S}}3$) *of Lemma* 5.2 *hold and that the solution of* $\mathcal{M}_t$ *is such that* $\vec{u} \in [H^{k+2}(\Omega)]^3$, $\underline{\theta} \in [H^{k+1}(\Omega)]^2$, $\underline{\underline{n}} \in [H^k(\Omega)]^{2\times 2}$ *and* $\underline{q} \in [H^k(\Omega)]^2$. *Then the finite element formulation* $\tilde{\mathcal{S}}_t^h$ *has a unique solution satisfying*

$$[\![\vec{u}-\vec{u}^h, \underline{\theta}-\underline{\theta}^h, \underline{\underline{n}}-\underline{\underline{n}}^h, \underline{q}-\underline{q}^h]\!]_{t,h} \leq C h^k (\|\vec{u}\|_{k+2} + \|\underline{\theta}\|_{k+1} + \|\underline{\underline{n}}, \underline{q}\|_k). \tag{5.26}$$

Proof: Let $(I^h\vec{u}, I^h\underline{\theta}, I^h\underline{\underline{n}}, I^h\underline{q}) \in \mathcal{U}^h \times \mathcal{Q}^h$ be the interpolant of $(\vec{u},\underline{\theta},\underline{\underline{n}},\underline{q})$ satisfying the estimates of Lemmas 5.3 and 5.4. By the stability result of Lemma 5.2 there exists $(\vec{v},\underline{\eta},\underline{\underline{p}},\underline{r}) \in \mathcal{U}^h \times \mathcal{Q}^h$ such that

$$[\![\vec{v}, \underline{\eta}, \underline{\underline{p}}, \underline{r}]\!]_{t,h} \leq 1 \tag{5.27}$$

and

$$C[\![\vec{u}^h - I^h\vec{u}, \underline{\theta}^h - I^h\underline{\theta}, \underline{\underline{n}}^h - I^h\underline{\underline{n}}, \underline{q}^h - I^h\underline{q}]\!]_{t,h} \leq \tilde{B}_h(\vec{u}^h - I^h\vec{u}, \underline{\theta}^h - I^h\underline{\theta}, \underline{\underline{n}}^h - I^h\underline{\underline{n}}, \underline{q}^h - I^h\underline{q}; \vec{v}, \underline{\eta}, \underline{\underline{p}}, \underline{r}). \tag{5.28}$$

Next, the consistency of Lemma 5.1 gives

$$\tilde{B}_h(\vec{u}^h - I^h\vec{u}, \underline{\theta}^h - I^h\underline{\theta}, \underline{\underline{n}}^h - I^h\underline{\underline{n}}, \underline{q}^h - I^h\underline{q}; \vec{v}, \underline{\eta}, \underline{\underline{p}}, \underline{r}) = \tilde{B}_h(\vec{u} - I^h\vec{u}, \underline{\theta} - I^h\underline{\theta}, \underline{\underline{n}} - I^h\underline{\underline{n}}, \underline{q} - I^h\underline{q}; \vec{v}, \underline{\eta}, \underline{\underline{p}}, \underline{r}). \tag{5.29}$$

To proceed, we write out the expression for the bilinear form

$$\begin{aligned}
&\tilde{B}_h(\vec{u} - I^h\vec{u}, \underline{\theta} - I^h\underline{\theta}, \underline{\underline{n}} - I^h\underline{\underline{n}}, \underline{q} - I^h\underline{q}; \vec{v}, \underline{\eta}, \underline{\underline{p}}, \underline{r}) \\
&= A(\vec{u} - I^h\vec{u}, \underline{\theta} - I^h\underline{\theta}; \vec{v}, \underline{\eta}) + M(\vec{u} - I^h\vec{u}, \underline{\theta} - I^h\underline{\theta}, ; \underline{\underline{p}}, \underline{r}) \\
&\quad + M(\vec{v}, \underline{\eta}; \underline{\underline{n}} - I^h\underline{\underline{n}}, \underline{q} - I^h\underline{q}) - N(\underline{\underline{n}} - I^h\underline{\underline{n}}, \underline{q} - I^h\underline{q}; \underline{\underline{p}}, \underline{r}) \\
&\quad + S_h^e(\vec{u} - I^h\vec{u}, \underline{\theta} - I^h\underline{\theta}, \underline{\underline{n}} - I^h\underline{\underline{n}}, \underline{q} - I^h\underline{q}; \vec{v}, \underline{\eta}, \underline{\underline{p}}, \underline{r}) \\
&\quad + \tilde{S}_h^c(\vec{u} - I^h\vec{u}, \underline{\theta} - I^h\underline{\theta}, \underline{\underline{n}} - I^h\underline{\underline{n}}, \underline{q} - I^h\underline{q}; \vec{v}, \underline{\eta}, \underline{\underline{p}}, \underline{r}).
\end{aligned} \tag{5.30}$$

For the first term above, the Schwarz inequality, Lemma 5.3 and (5.27) give

$$\begin{aligned}
|A(\vec{u} - I^h\vec{u}, \underline{\theta} - I^h\underline{\theta}; \vec{v}, \underline{\eta})| &\leq C\|\vec{u} - I^h\vec{u}, \underline{\theta} - I^h\underline{\theta}\|_1\|\vec{v}, \underline{\eta}\|_1 \\
&\leq Ch^k(\|\vec{u}\|_{k+2} + \|\underline{\theta}\|_{k+1}).
\end{aligned} \tag{5.31}$$

By using the integration by parts formulas (3.13), (3.14) and Lemma 5.3 we get

$$\begin{aligned}
&|M(\vec{u} - I^h\vec{u}, \underline{\theta} - I^h\underline{\theta}, ; \underline{\underline{p}}, \underline{r})| \\
&\leq \Big(\sum_{E\in\Gamma_h\setminus\Gamma_0} h_E^{-3}\|\vec{u} - I^h\vec{u}\|_{0,E}^2 + \sum_{K\in\mathcal{C}_h} h_K^{-2}\|\vec{u} - I^h\vec{u}, \underline{\theta} - I^h\underline{\theta}\|_{0,K}^2\Big)^{1/2} [\![\underline{\underline{p}}, \underline{r}]\!]_{t,h} \\
&\leq Ch^k(\|\vec{u}\|_{k+2} + \|\underline{\theta}\|_{k+1}).
\end{aligned} \tag{5.32}$$

Next, we directly get

$$|M(\vec{v}, \underline{\eta}; \underline{\underline{n}} - I^h\underline{\underline{n}}, \underline{q} - I^h\underline{q})| \leq C\|\underline{\underline{n}} - I^h\underline{\underline{n}}, \underline{q} - I^h\underline{q}\|_0\|\vec{v}, \underline{\eta}\|_1 \leq Ch^k\|\underline{\underline{n}}, \underline{q}\|_k \tag{5.33}$$

and

$$|N(\underline{\underline{n}} - I^h\underline{\underline{n}}, \underline{q} - I^h\underline{q}; \underline{\underline{p}}, \underline{r})| \leq Ct\|\underline{\underline{n}} - I^h\underline{\underline{n}}, \underline{q} - I^h\underline{q}\|_0\|\underline{\underline{p}}, \underline{r}\|_{t,h} \leq Ch^k\|\underline{\underline{n}}, \underline{q}\|_k. \tag{5.34}$$

The first stabilizing term is estimated using Lemmas 5.3 and 5.4 as follows

$$\begin{aligned}
&|S_h^e(\vec{u} - I^h\vec{u}, \underline{\theta} - I^h\underline{\theta}, \underline{\underline{n}} - I^h\underline{\underline{n}}, \underline{q} - I^h\underline{q}; \vec{v}, \underline{\eta}, \underline{\underline{p}}, \underline{r})| \\
&\leq \Big((\sum_{K\in\mathcal{C}_h} h_K^2\|\vec{u} - I^h\vec{u}, \underline{\theta} - I^h\underline{\theta}\|_{2,K}^2)^{1/2} + [\![\underline{\underline{n}} - I^h\underline{\underline{n}}, \underline{q} - I^h\underline{q}]\!]_{t,h}\Big) \\
&\qquad \times \Big((\sum_{K\in\mathcal{C}_h} h_K^2\|\vec{v}, \underline{\eta}\|_{2,K}^2)^{1/2} + [\![\underline{\underline{p}}, \underline{r}]\!]_{t,h}\Big) \\
&\leq Ch^k(\|\vec{u}\|_{k+2} + \|\underline{\theta}\|_{k+1} + \|\underline{\underline{n}}, \underline{q}\|_k),
\end{aligned} \tag{5.35}$$

where we used the inverse inequality (4.68). For the second stabilizing term we obtain

$$\begin{aligned}
&|\tilde{S}_h^c(\vec{u}-I^h\vec{u},\underline{\theta}-I^h\underline{\theta},\underline{\underline{n}}-I^h\underline{\underline{n}},\underline{q}-I^h\underline{q};\vec{v},\underline{\eta},\underline{\underline{p}},\underline{r})| \\
&\quad\leq\ C\Big(|M(\vec{u}-I^h\vec{u},\underline{\theta}-I^h\underline{\theta},;\underline{\underline{p}},\underline{r})|+|M(\vec{v},\underline{\eta};\underline{\underline{n}}-I^h\underline{\underline{n}},\underline{q}-I^h\underline{q})| \\
&\qquad\qquad +|N(\underline{\underline{n}}-I^h\underline{\underline{n}},\underline{q}-I^h\underline{q};\underline{\underline{p}},\underline{r})|\Big) \\
&\qquad +C[\![\vec{u}-I^h\vec{u},\underline{\theta}-I^h\underline{\theta}]\!]_{t,h}[\![\vec{v},\underline{\eta}]\!]_{t,h} \\
&\quad\leq\ Ch^k(\|\vec{u}\|_{k+2}+\|\underline{\theta}\|_{k+1}+\|\underline{\underline{n}},\underline{q}\|_k), \qquad (5.36)
\end{aligned}$$

by virtue of the previous estimates (5.32)–(5.34), the Schwarz inequality and Lemma 5.3. Finally, we note that Lemmas 5.3 and 5.4 also give

$$[\![\vec{u}-I^h\vec{u},\underline{\theta}-I^h\underline{\theta},\underline{\underline{n}}-I^h\underline{\underline{n}},\underline{q}-I^h\underline{q}]\!]_{t,h}\leq Ch^k(\|\vec{u}\|_{k+2}+\|\underline{\theta}\|_{k+1}+\|\underline{\underline{n}},\underline{q}\|_k). \quad (5.37)$$

Hence, collecting (5.28) to (5.37) and using the triangle inequality we get the asserted estimate. ■

6 Numerical results

6.1 Numerical procedure

We have implemented the two stabilized methods analysed in Sections 4 and 5, for the lowest-order finite element spaces, i.e. letting $k=1$ in (4.16) and (4.17), in the simplified case of circular cylinders, using the MODULEF library. In addition to the obvious implementational simplification, circular cylinders indeed feature two major advantages:

- They provide one of the very few instances of shell geometries for which some reference solutions can be derived, either in closed form or with arbitrary numerical precision, so that reliable benchmarks are available.
- This geometry allows a wide variety of possible limit behaviours when the thickness is very small, according to how boundary conditions are imposed. In particular, if essential boundary conditions are not imposed outside of rulings (straight lines parallel to the axis), it is easily seen that the limit problem is bending-dominated.

We considered two different benchmarks. In the sequel, we refer to the natural coordinate system, shown in Figure 1, in which $\vec{a}_1$ and $\vec{a}_2$ are unit vectors, respectively tangent and normal to generators. The benchmarks are defined as follows:

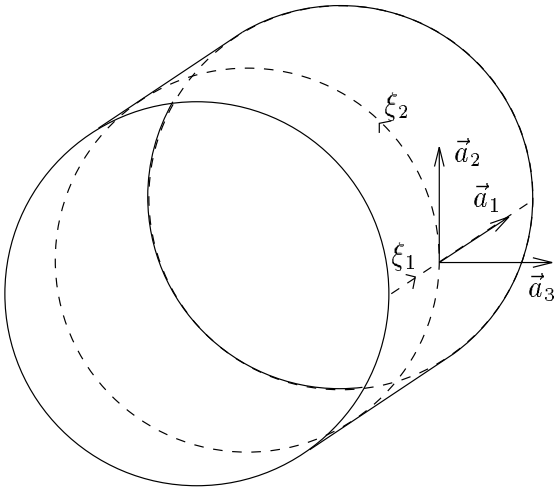

Figure 1: The natural coordinate system

a) The boundary conditions and the loading are defined in such a way that the shell behaves like an arch. Considering a cylindrical "slice" at the ends of which one imposes $u_1 = \theta_1 = 0$, and assuming that the loading $\vec{f}$ is a function of ξ_2 such that $f_1 = 0$, it is indeed easily seen that the resulting behaviour is that of an arch, i.e.:

$$\begin{cases} \vec{u} = \vec{u}(\xi_2),\ \underline{\theta} = \underline{\theta}(\xi_2) \\ u_1 = \theta_1 = 0 \end{cases} . \tag{6.1}$$

We used an example, borrowed from [12], for which a closed-form arch solution can be derived. This example, described in Figure 2, corresponds to a semi-circular arch clamped at both ends and loaded by a uniformly-distributed constant force. We considered a slice of length equal to 1.5 times the radius R and, for symmetry reasons, one half of this slice was computed, so that the actual computational domain was a $1.5 \times \frac{\pi}{2}$ rectangle.

b) A fully-circular cylinder is loaded by a periodic pressure, so that the shell reduces to a one-dimensional model along the axis. This problem is analysed in detail in [41], and we applied the procedure described therein to obtain numerical solutions with arbitrary precision, using symbolic calculus software. The specific example considered is a cylinder with free ends and of length $2R$, loaded by a pressure $p = p_0 \cos(2\xi_2/R)$, see Figure 3. Due to symmetry, only one sixteenth of the structure was effectively computed, using a $1 \times \frac{\pi}{4}$ rectangle.

In each case, numerical solutions were computed using triangular meshes automatically generated by a Voronoï method available in MODULEF [23], for gradually

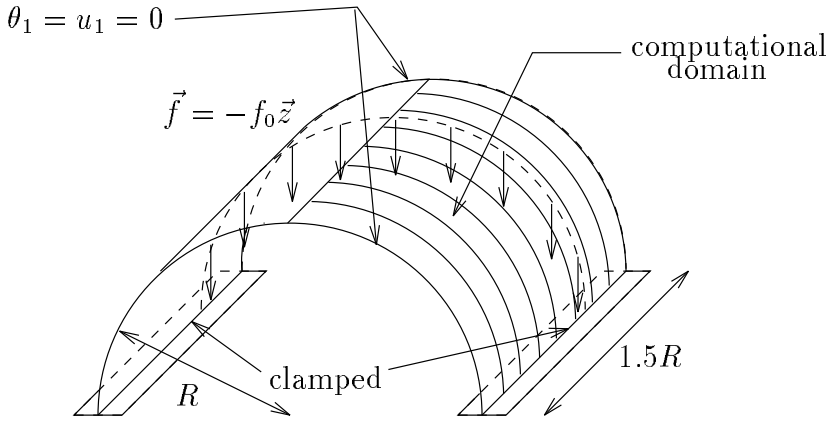

Figure 2: Semi-circular arch

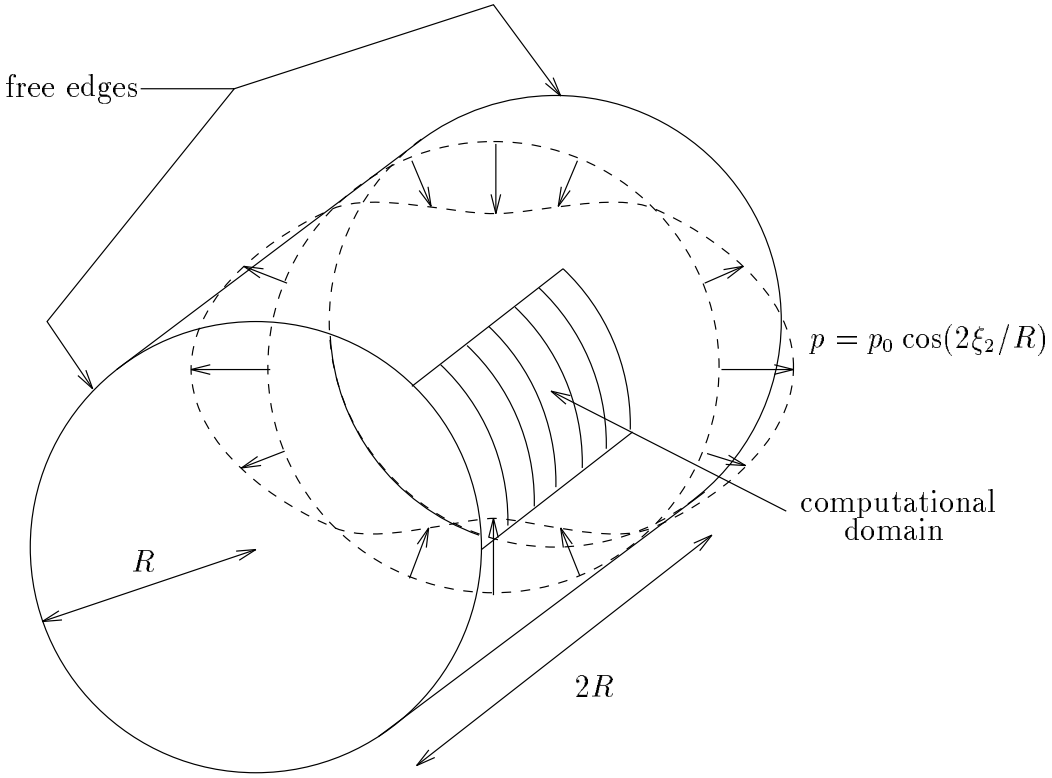

Figure 3: Cylinder loaded by periodic pressure

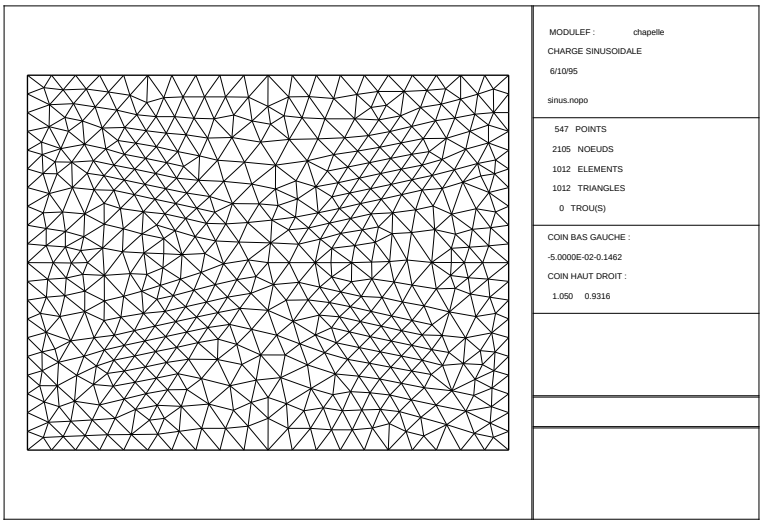

Figure 4: Mesh generated for the $1 \times \frac{\pi}{4}$ domain (N=20)

refined discretizations of the rectangular boundary, dividing each side into N equal parts, with $N = 10, 20, 40$. One of these meshes is shown in Figure 4. We deliberately avoided using meshes aligned with the sides of the domain, or with the axis of the cylinder, that seem to alleviate locking phenomena in some instances [40].

The Poisson ratio ν was set to 0.3. The stabilization constants were determined by performing some preliminary calculations on the first benchmark. For the first stabilized method, we then used $\alpha_1 = 0.3\ 10^{-3}/E$, $\alpha_2 = 0$, $\alpha_3 = 0.1R^2/E$, $\alpha_4 = \alpha_5 = 50/R^2$. For the second method we chose $\alpha_1 = 0.3\ 10^{-3}/E$, $\alpha_2 = 0$, $\alpha_3 = 0.1R^2/E$, $\alpha_4 = \alpha_5 = 0.01$. In both cases the choice $\alpha_2 = 0$ is allowed because, with $\underline{\mathrm{div}}\ \underline{\underline{n}} = \underline{0}$ inside each element for any discrete membrane stress field $\underline{\underline{n}}$ (since all $n^{\alpha\beta}$ are constant), the corresponding least-square term provides no further stabilization than the first one.

In order to compute the stabilized numerical solutions, we used standard solvers for symmetric positive matrices. To that purpose we eliminated the stress degrees of freedom element by element, as indicated in Remark 4.1 above. Denoting by U the column vector relative to the displacements and rotations, and by P that relative to the stresses, the stabilized mixed methods indeed lead to the following typical matrix equation:

$$\begin{pmatrix} M_{UU} & M_{PU}^T \\ M_{PU} & -M_{PP} \end{pmatrix} \begin{pmatrix} U \\ P \end{pmatrix} = \begin{pmatrix} F_U \\ F_P \end{pmatrix}. \tag{6.2}$$

From the proof of the stability results, it is clear that both M_{UU} and M_{PP} are symmetric positive definite matrices so that we can eliminate P from this equation. Note that, since the stress finite element functions are discontinuous, this elimination can be carried out at the element level as the assembling is performed. Thus,

instead of (6.2), the problem solved in practice is

$$(M_{UU} + M_{PU}^T M_{PP}^{-1} M_{PU})U = F_U - M_{PU}^T M_{PP}^{-1} F_P, \tag{6.3}$$

where the matrix to be inverted, $(M_{UU} + M_{PU}^T M_{PP}^{-1} M_{PU})$, is now symmetric positive definite. Hence we solve a matrix equation similar to what would be obtained from a standard displacement-based finite element method.

6.2 Analysis of the results

Figures 5 and 7 show relative errors in the H^1-semi-norm for the displacements, i.e.

$$e^h = \frac{|\vec{u} - \vec{u}^h|_1}{|\vec{u}|_1},$$

computed from solutions obtained by the first stabilized method, and also by a purely displacement-based scheme corresponding to the standard Galerkin approximation of problem $\mathcal{P}_t$ using $\mathcal{U}^h$. Thickness values of 0.1, 0.01, 0.001 (scaled by R) were considered. Figures 6 and 8 compare similar errors for the rotations with interpolation errors, since optimal convergence is expected from the theoretical analysis. These interpolation errors turned out to vary very little with the thickness (typically within 10% ranges), so for the sake of legibility only their values for $t = 0.1$ are displayed. Note that, from the theory, optimal error estimates are not expected for the displacements since convergence is governed by the approximability properties of the lowest degree finite element space, i.e. that of the linear rotations. This is confirmed by the numerical results, as can be seen in Figures 5 and 7 where the slope of the best convergence curves barely exceeds unity. This explains why we do not plot interpolation estimates for the displacements.

In Figures 5 and 6, i.e. for the arch benchmark, we see no significant influence of the thickness on approximation errors of the stabilized method. Moreover, the errors for the rotations are very close to interpolation errors. By contrast the displacement-based method exhibits strong locking, with completely erroneous (in fact vanishing) solutions for $t = 0.001$ for all meshes except the finest one.

In the second example, a marked deterioration of convergence appears in Figures 7 and 8 for the first stabilized method when the thickness decreases. Approximation errors are visibly affected in absolute value for each mesh, as well as in rates of convergence. However, even errors for $t = 0.001$ remain acceptable, especially when compared with the displacement-based method which dramatically fails here again.

We next compare the two stabilized methods. Figures 9 and 11 display the relative errors for the displacements approximated by both methods for thickness

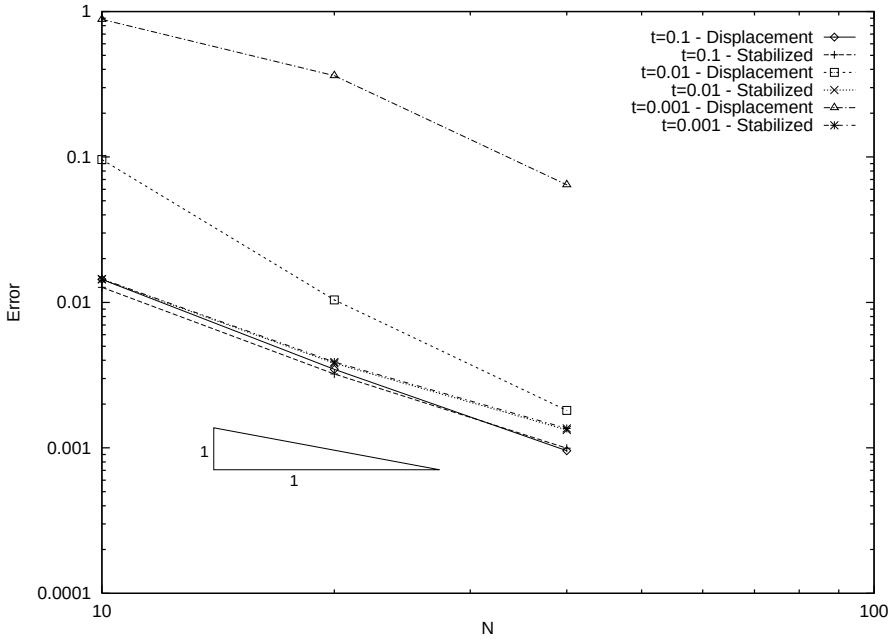

Figure 5: Benchmark a - Relative errors for $\vec{u}$

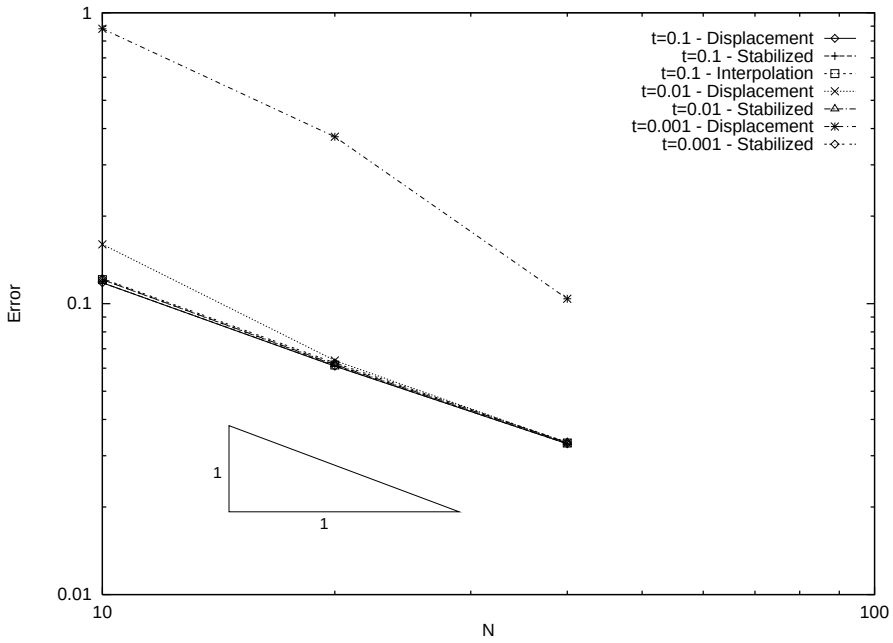

Figure 6: Benchmark a - Relative errors for $\underline{\theta}$

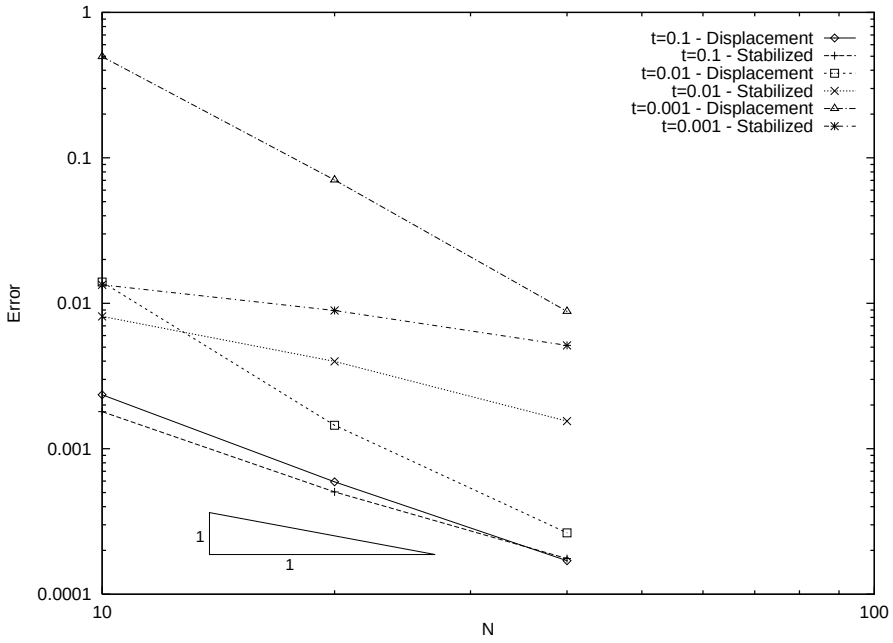

Figure 7: Benchmark b - Relative errors for $\vec{u}$

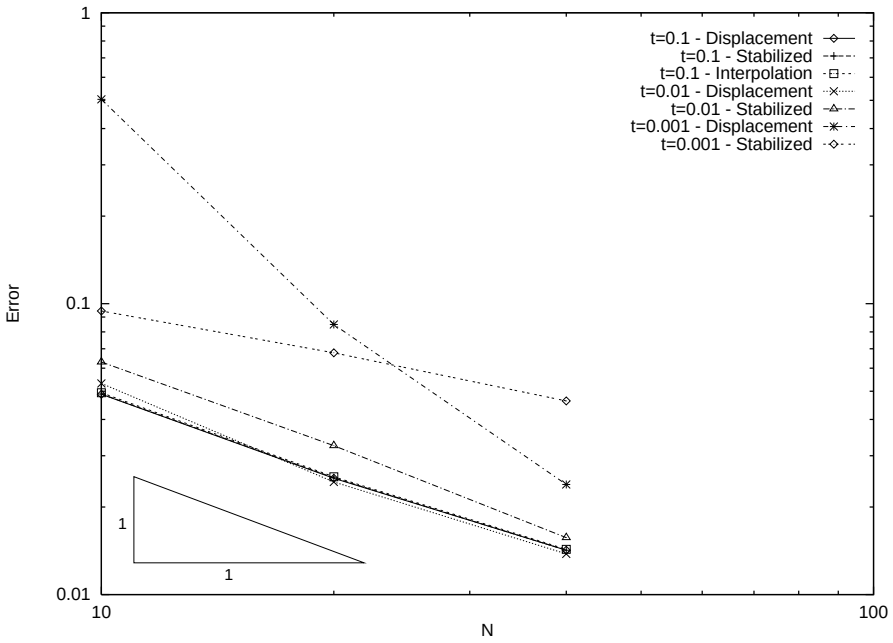

Figure 8: Benchmark b - Relative errors for $\underline{\theta}$

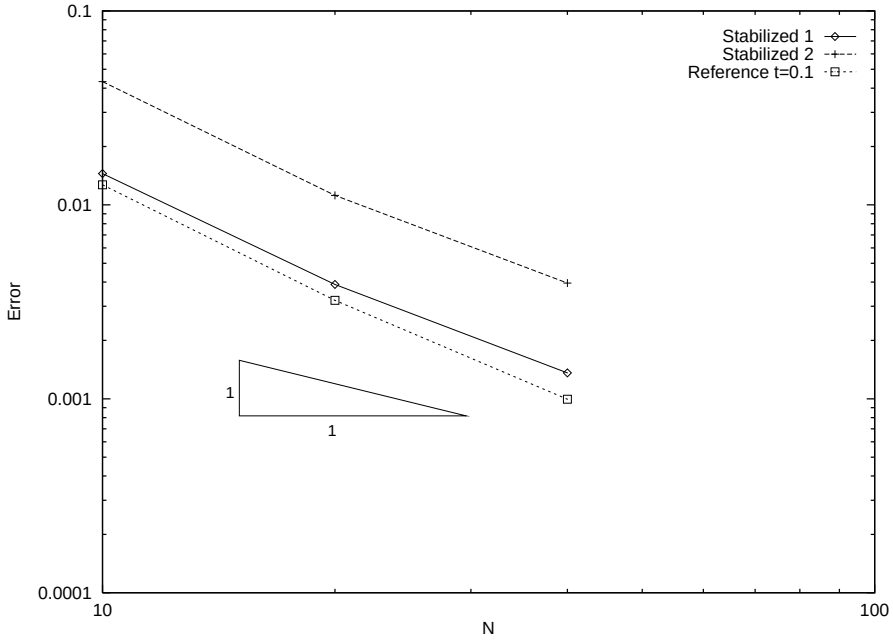

Figure 9: Benchmark a - Compared relative errors for $\vec{u}$

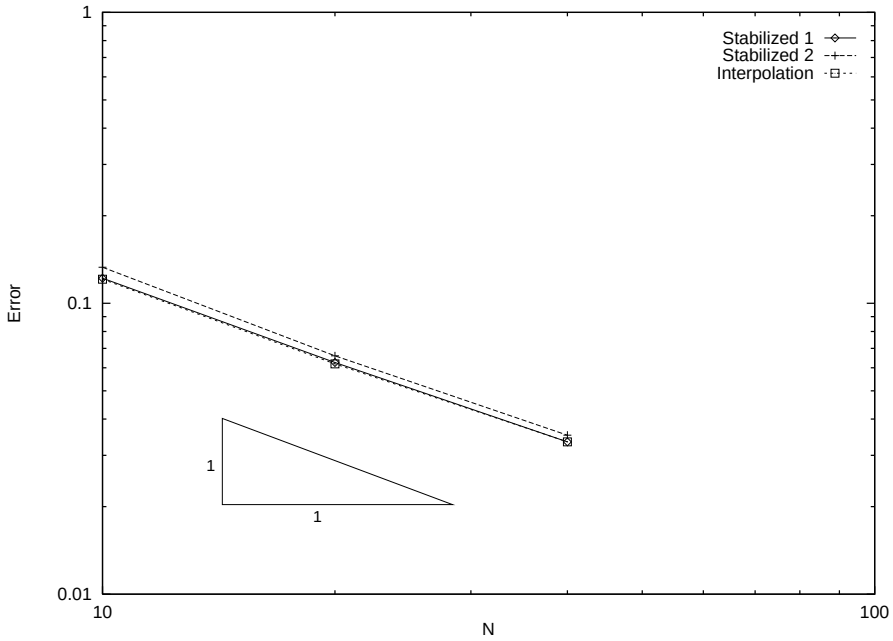

Figure 10: Benchmark a - Compared relative errors for $\underline{\theta}$

0.001, compared with the best numerical solutions obtained for thickness 0.1. Similar errors for the rotations are plotted in Figures 10 and 12, with the corresponding interpolation errors.

In the case of the arch benchmark, rotation errors remain essentially unchanged in Figure 10, but displacement errors now feature a more sensitive behaviour with

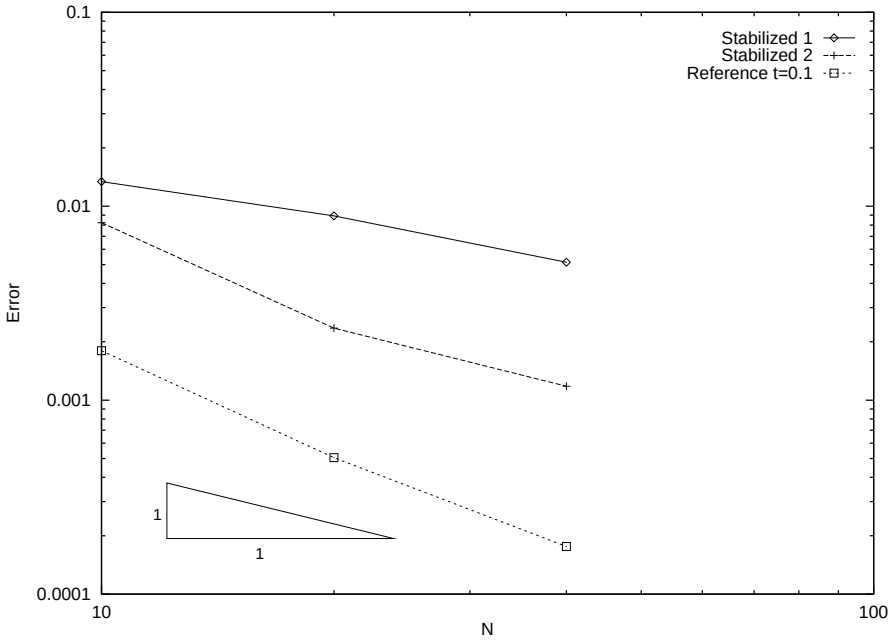

Figure 11: Benchmark b - Compared relative errors for $\vec{u}$

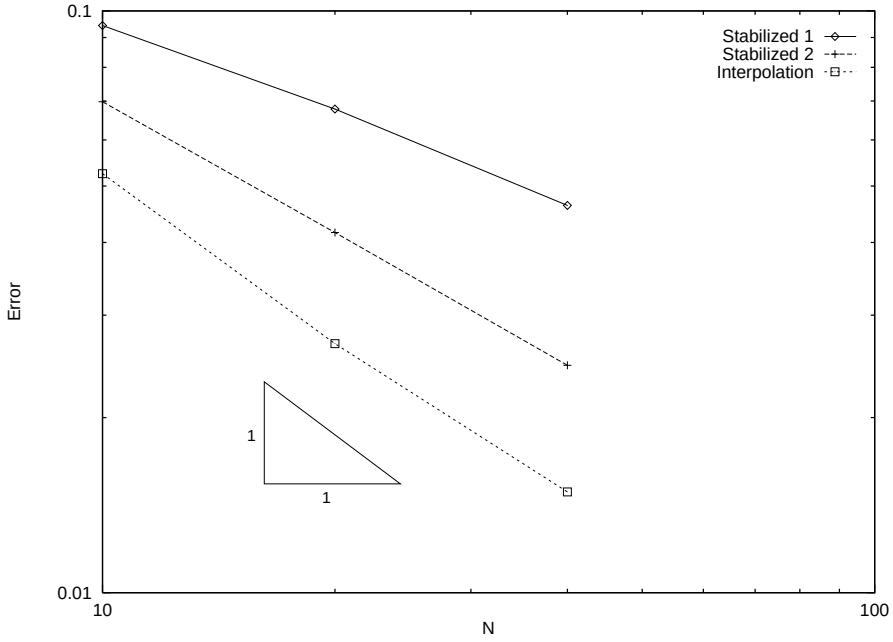

Figure 12: Benchmark b - Compared relative errors for $\underline{\theta}$

respect to the thickness parameter (Figure 9). However, this phenomenon is much more limited than the deterioration observed with the first stabilized method in the second benchmark. Furthermore, the numerical results obtained with this second stabilized method here still strikingly differ from the locking behaviour exhibited by the displacement-based method.

Regarding the results obtained for the second example and displayed in Figures 11 and 12, we see that the second stabilized method provides significantly improved approximations for displacements as well as rotations.

In partial conclusion to our numerical tests, it appears that, even though both stabilized methods give reasonably good results, the second one seems more promising in that it limits the sensitivity of convergence behaviours with respect to the thickness parameter in the two examples that we considered. This sensitivity is not completely ruled out, but it does not compare with a real locking behaviour such as the one displayed by the displacement-based method.

A Appendix

This section contains results of essentially geometric character that are used in the numerical analysis of the shell formulation. We start with a classical property of the first fundamental form (see e.g. [43]), the proof of which we give for completeness.

Lemma A.1 *There exist two strictly positive constants c and C such that, at any point of $\bar{\Omega}$ and for any surface tensor $\underline{\eta}$:*

$$c[(\eta^1)^2 + (\eta^2)^2] \leq a_{\alpha\beta}\eta^\alpha\eta^\beta \leq C[(\eta^1)^2 + (\eta^2)^2], \tag{A.1}$$

$$\frac{1}{C}[(\eta_1)^2 + (\eta_2)^2] \leq a^{\alpha\beta}\eta_\alpha\eta_\beta \leq \frac{1}{c}[(\eta_1)^2 + (\eta_2)^2]. \tag{A.2}$$

Proof: Consider the following function defined on $\mathcal{R}^2 \times \bar{\Omega}$:

$$\Big((\eta^1, \eta^2), M\Big) \mapsto a_{\alpha\beta}(M)\eta^\alpha\eta^\beta.$$

Since $\underline{\underline{a}}$ is the metric tensor of the surface, this function is strictly positive everywhere except when $(\eta^1, \eta^2) = (0, 0)$. Moreover it is continuous since $\vec{r}$ is smooth. Define c and C as its minimum and maximum values on the compact set $\{(\eta^1, \eta^2), (\eta^1)^2 + (\eta^2)^2 = 1\} \times \bar{\Omega}$. Hence $c > 0$ and (A.1) holds. Then (A.2) immediately follows from the fact that the matrix $(a^{\alpha\beta})$ is the inverse of $(a_{\alpha\beta})$, so that they have inverse eigenvalues. ■

We next establish a property regarding vectors orthogonal to curves which are images in $\mathcal{S}$ of straight lines in Ω. This result enters as a crucial element in the stability of the proposed approximation schemes.

Lemma A.2 *Let C be an oriented straight segment in Ω, $\underline{\tau}$ the unit tangent vector of $\vec{r}(C)$ in $\mathcal{S}$, $\underline{\nu}$ the unit normal vector ($\nu_\alpha = \epsilon_{\alpha\beta}\tau^\beta$, see [24]). Then, for any two points M_1, M_2 on C, there exists a positive number ζ such that:*

$$\begin{cases} \nu_1(M_2) = \zeta\, \nu_1(M_1) \\ \nu_2(M_2) = \zeta\, \nu_2(M_1) \end{cases} \tag{A.3}$$

Moreover, ζ lies in an interval $[\zeta_I, \zeta_S]$, with $\zeta_I > 0$, the bounds of which are independent of the specific segment considered.

Proof: Let $(\xi_1(x) = \lambda_1 x + \mu_1, \xi_2(x) = \lambda_2 x + \mu_2)$ be a parametrization of C. Define:

$$\vec{t} \stackrel{\text{def}}{=} \frac{d}{dx}\left[\vec{r}(\xi_1(x), \xi_2(x))\right].$$

By the chain rule:

$$\vec{t} = \lambda_1 \vec{a}_1 + \lambda_2 \vec{a}_2,$$

and of course $\underline{\tau}$ equals $\vec{t}$ up to a normalizing factor. Now let:

$$\vec{n} \stackrel{\text{def}}{=} -\lambda_2 \vec{a}^1 + \lambda_1 \vec{a}^2,$$

so that $\vec{t} \cdot \vec{n} = 0$ and $(\vec{t}, \vec{n}, \vec{a}_3)$ is positively oriented. Therefore $\underline{\nu}$ equals $\vec{n}$ up to a normalizing factor also, i.e.:

$$\begin{cases} \nu_1 = -\lambda_2[(\lambda_2)^2 a^{11} + (\lambda_1)^2 a^{22} - 2\lambda_1\lambda_2 a^{12}]^{-\frac{1}{2}} \\ \nu_2 = \lambda_1[(\lambda_2)^2 a^{11} + (\lambda_1)^2 a^{22} - 2\lambda_1\lambda_2 a^{12}]^{-\frac{1}{2}} \end{cases}$$

Let now M_1 and M_2 be any two points on C. We infer:

$$\frac{\nu_1(M_2)}{\nu_1(M_1)} = \frac{\nu_2(M_2)}{\nu_2(M_1)} = \sqrt{\frac{(\lambda_2)^2 a^{11}(M_1) + (\lambda_1)^2 a^{22}(M_1) - 2\lambda_1\lambda_2 a^{12}(M_1)}{(\lambda_2)^2 a^{11}(M_2) + (\lambda_1)^2 a^{22}(M_2) - 2\lambda_1\lambda_2 a^{12}(M_2)}}, \tag{A.4}$$

so that (A.3) holds. Moreover, from Lemma A.1 we have for any point M in Ω:

$$\frac{1}{C}\left[(\lambda_1)^2 + (\lambda_2)^2\right] \leq (\lambda_2)^2 a^{11}(M) + (\lambda_1)^2 a^{22}(M) - 2\lambda_1\lambda_2 a^{12}(M) \leq \frac{1}{c}\left[(\lambda_1)^2 + (\lambda_2)^2\right].$$

Therefore:

$$\sqrt{\frac{c}{C}} \leq \sqrt{\frac{(\lambda_2)^2 a^{11}(M_1) + (\lambda_1)^2 a^{22}(M_1) - 2\lambda_1\lambda_2 a^{12}(M_1)}{(\lambda_2)^2 a^{11}(M_2) + (\lambda_1)^2 a^{22}(M_2) - 2\lambda_1\lambda_2 a^{12}(M_2)}} \leq \sqrt{\frac{C}{c}},$$

and setting $\zeta_I = \sqrt{\frac{c}{C}}$ and $\zeta_S = \sqrt{\frac{C}{c}}$, this completes the proof. ■

We now establish an ellipticity property for tensor $\underline{\underline{\underline{\underline{\breve{E}}}}}$.

Lemma A.3 *For any symmetric tensor* $\underline{\underline{X}}$ *at any point* M *in* Ω*, we have:*

$$\underline{\underline{X}} : \underline{\underline{\underline{\underline{\breve{E}}}}} : \underline{\underline{X}} \geq \frac{1}{2E} \underline{\underline{X}} : \underline{\underline{X}} \,. \tag{A.5}$$

Proof: From the definition of $\underline{\underline{\underline{\underline{\breve{E}}}}}$ in section 4, we have:

$$\underline{\underline{X}} : \underline{\underline{\underline{\underline{\breve{E}}}}} : \underline{\underline{X}} = \frac{1+\nu}{E} \left[X^{\lambda\mu} X_{\lambda\mu} - \frac{\nu}{1+\nu} \left(X^{\lambda}_{\lambda} \right)^2 \right] . \tag{A.6}$$

We can always construct, at least in a neighbourhood of M, a new coordinate system such that, at M, the covariant base vectors are orthogonal and of unit-length. We denote by $(\tilde{X}_{\lambda\mu})$ the components of $\underline{\underline{X}}$ in this new coordinate system where no difference subsists between covariant and contravariant forms. Since tensor invariants are independent of the coordinate system considered, we have:

$$\left(X^{\lambda}_{\lambda} \right)^2 = \left(\tilde{X}_{\lambda\lambda} \right)^2 \leq 2 \tilde{X}_{\lambda\mu} \tilde{X}_{\lambda\mu} = 2 X^{\lambda\mu} X_{\lambda\mu}.$$

Thus, from (A.6):

$$\underline{\underline{X}} : \underline{\underline{\underline{\underline{\breve{E}}}}} : \underline{\underline{X}} \geq \frac{1-\nu}{E} \underline{\underline{X}} : \underline{\underline{X}} \geq \frac{1}{2E} \underline{\underline{X}} : \underline{\underline{X}},$$

since $\nu \leq \frac{1}{2}$. ■

Acknowledgement: Most of this work was performed while R. Stenberg was invited at INRIA by Professor M. Bernadou, under whose supervision D. Chapelle was pursuing his doctorate. Both authors are grateful to Professor Bernadou for giving them this opportunity, for several helpful discussions, and for his continuous support during this project.

References

[1] D.N. Arnold. Discretization by finite elements of a model parameter dependent problem. *Numer. Math.*, 37:405–421, 1981.

[2] D.N. Arnold and F. Brezzi. Locking free finite elements for shells. Research Report 898, Istituto di Analisi Numerica, Pavia, 1993.

[3] D.N. Arnold and R.S. Falk. A uniformly accurate finite element method for the Reissner-Mindlin plate model. *SIAM J. Numer. Anal.*, 26(6):1276–1290, 1989.

[4] K. Arunakirinathar and B. D. Reddy. Mixed finite element methods for elastic rods of arbitrary geometry. *Numer. Math.*, 64:13–43, 1993.

[5] M. Bernadou. *Méthodes d'Eléments Finis pour les Problèmes de Coques Minces*. Masson, 1994.

[6] M. Bernadou and P.G. Ciarlet. Sur l'ellipticité du modèle linéaire de coques de W.T. Koiter. In R. Glowinski and J.L. Lions, editors, *Computing Methods in Applied Sciences and Engineering*, 1975.

[7] D. Braess. *Finite Elemente*. Springer-Verlag, 1992.

[8] F. Brezzi, K.J. Bathe, and M. Fortin. Mixed-interpolated elements for Reissner-Mindlin plates. *Internat. J. Numer. Methods Engrg.*, 28:1787–1801, 1989.

[9] F. Brezzi and M. Fortin. *Mixed and Hybrid Finite Element Methods*. Springer-Verlag, 1991.

[10] F. Brezzi, M. Fortin, and R. Stenberg. Error analysis of mixed-interpolated elements for Reissner-Mindlin plates. *Math. Models Methods Appl. Sci.*, 1(2):125–151, 1991.

[11] M.L. Bucalem and K.-J. Bathe. Finite element analysis of shell structures. *Arch. Comput. Methods Engrg.* To appear.

[12] D. Chapelle. A locking-free approximation of curved rods by straight beam elements. Research Report 2733, INRIA, 1995.

[13] D. Chenais and J.-C. Paumier. On the locking phenomenon for a class of elliptic problems. *Numer. Math.*, 67:427–440, 1994.

[14] P.G. Ciarlet. *The Finite Element Method for Elliptic Problems*. North-Holland, 1978.

[15] P.G. Ciarlet and V. Lods. Analyse asymptotique des coques linéairement élastiques. I. Coques "membranaires". *C. R. Acad. Sci. Paris*, t.318:863–868, 1994. Série I.

[16] P.G. Ciarlet, V. Lods, and B. Miara. Analyse asymptotique des coques linéairement élastiques. II. Coques "en flexion". *C. R. Acad. Sci. Paris*, t.319:95–100, 1994. Série I.

[17] P. Clément. Approximation by finite element functions using local regularization. *R.A.I.R.O.*, R-2:77–84, 1975.

[18] N. Coutris. Théorème d'existence et d'unicité pour un problème de coque élastique dans le cas d'un modèle linéaire de P.M. Naghdi. *RAIRO Analyse Numérique*, 12:51–57, 1978.

[19] P. Destuynder. *Modélisation des Coques Minces Elastiques.* Masson, 1990.

[20] P. Destuynder and T. Nevers. A new finite element scheme for bending plates. *Comput. Methods Appl. Mech. Engrg.*, 68:127–139, 1988.

[21] L. Franca and R. Stenberg. Error analysis of some Galerkin-least-squares methods for the elasticity equations. *SIAM J. Numer. Anal.*, 28:1680–1697, 1991.

[22] L.P. Franca, T.J.R. Hughes, A.F.D. Loula, and I. Miranda. A new family of stable elements for nearly incompressible elasticity based on a mixed Petrov-Galerkin finite element formulation. *Numer. Math.*, 53:123–141, 1988.

[23] P.L. George. *MODULEF, Guide n° 3 : Création et Modification de Maillages.*

[24] A.E. Green and W. Zerna. *Theoretical Elasticity.* Oxford University Press, 2nd edition, 1968.

[25] I. Harari and T.J.R. Hughes. What is C and h: Inequalities for the analysis and design of finite element methods. *Comput. Methods Appl. Mech. Engrg.*, 97:157–192, 1992.

[26] T.J.R. Hughes and L.P. Franca. A new finite element formulation for computational fluid dynamics: VII. The Stokes problem with various well-posed boundary conditions: Symmetric formulations that converge for all velocity/pressure spaces. *Comput. Methods Appl. Mech. Engrg.*, 65:85–96, 1987.

[27] T.J.R. Hughes and L.P. Franca. A mixed finite element formulation for Reissner-Mindlin plate theory: uniform convergence of all higher-order spaces. *Comput. Methods Appl. Mech. Engrg.*, 67:223–240, 1988.

[28] A. Kirmse. Bending-dominated deformation of thin spherical shells: analysis and finite element approximation. *SIAM J. Numer. Anal.*, 30(4):1015–1040, 1993.

[29] W.T. Koiter. A consistent first approximation in the general theory of thin elastic shells. Part 1, Foundations and linear theory. Research Report 713, Delft Technological University, 1959.

[30] A.F.D. Loula, L.P. Franca, T.J.R. Hughes, and I. Miranda. Stability, convergence and accuracy of a new finite element method for the circular arch problem. *Comput. Methods Appl. Mech. Engrg.*, 63:281–303, 1987.

[31] A.F.D. Loula, T.J.R. Hughes, L.P. Franca, and I. Miranda. Mixed Petrov-Galerkin methods for the Timoshenko beam. *Comput. Methods Appl. Mech. Engrg.*, 63:133–154, 1987.

[32] M. Lyly and R. Stenberg. In preparation.

[33] M. Lyly and R. Stenberg. Stabilized MITC plate bending elements. In M. Papadrakakis and B.H.V. Topping, editors, *Advances in Finite Element Techniques*, pages 11–16. CIVIL-COMP Ltd, 1994.

[34] M. Lyly, R. Stenberg, and T. Vihinen. A stable bilinear element for Reissner-Mindlin plates. *Comput. Methods Appl. Mech. Engrg.*, 110:343–357, 1993.

[35] P.M. Naghdi. Foundations of elastic shell theory. In *Progress in Solid Mechanics*, volume 4, pages 1–90. North-Holland, 1963.

[36] J. Piila and J. Pitkäranta. Energy estimates relating different linear elastic models of a thin cylindrical shell. I. The membrane-dominated case. *SIAM J. Math. Anal.*, 24(1):1–22, 1993.

[37] J. Piila and J. Pitkäranta. Energy estimates relating different linear elastic models of a thin cylindrical shell. II: The case of free boundary. *SIAM J. Math. Anal.*, 26(4):820–849, 1995.

[38] J. Pitkäranta. Boundary subspaces for the finite element method with Lagrange multipliers. *Numer. Math.*, 33:273–289, 1979.

[39] J. Pitkäranta. Analysis of some low-order finite element schemes for Mindlin-Reissner and Kirchhoff plates. *Numer. Math.*, 53:237–254, 1988.

[40] J. Pitkäranta. The problem of membrane locking in finite element analysis of cylindrical shells. *Numer. Math.*, 61:523–542, 1992.

[41] J. Pitkäranta, Y. Leino, O. Ovaskainen, and J. Piila. Shell deformation states and the finite element method: a benchmark study of cylindrical shells. *Comput. Methods Appl. Mech. Engrg.*, 128:81–121, 1995.

[42] P.A. Raviart and J.-M. Thomas. *Introduction à l'Analyse Numérique des Equations aux Dérivées Partielles.* Masson, 1988.

[43] P. Rougée. *Equilibre des Coques Elastiques Minces Inhomogènes en Théorie Non Linéaire.* Thèse d'Etat, Université de Paris, 1969.

[44] E. Sanchez-Palencia. Statique et dynamique des coques minces. I. Cas de flexion pure non inhibée. *C. R. Acad. Sci. Paris*, t.309:411–417, 1989. Série I.

[45] E. Sanchez-Palencia. Statique et dynamique des coques minces. II. Cas de flexion pure inhibée - Approximation membranaire. *C. R. Acad. Sci. Paris*, t.309:531–537, 1989. Série I.

[46] R. Stenberg. Error analysis of some finite element methods for the Stokes problem. *Math. Comp.*, 54:495–508, 1990.

[47] R. Stenberg. A new finite element formulation for the plate bending problem. In Ciarlet, Trabucho, and Viano, editors, *Asymptotic Methods for Elastic Structures*, pages 209–221. Walter de Gruyter & Co., 1995.

[48] R. Stenberg and M. Suri. An *hp* error analysis of MITC plate elements. *SIAM J. Numer. Anal.* To appear in 1996.

[49] M.. Suri. A reduced constraint *hp* finite element method for shell problems. Research report, University of Maryland, Baltimore County, 1994.

[50] R. Valid. *The Nonlinear Theory of Shells through Variational Principles.* John Wiley & Sons, 1995.

[51] R. Verfürth. Error estimates for a mixed finite element approximation of the Stokes problem. *RAIRO Anal. Num.*, 18:175–182, 1984.

[52] G.R. Wempner. *Mechanics of Solids with Applications to Thin Bodies.* Sijthoff & Noordhoff, 1981.

Unit´e de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unit´e de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unit´e de recherche INRIA Rhône-Alpes, 655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN
Unit´e de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unit´e de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
ISSN 0249-6399