Helsinki University of Technology Institute of Mathematics Research Reports
Teknillisen korkeakoulun matematiikan laitoksen tutkimusraporttisarja
Espoo 2003 A449

MATHEMATICAL AND HISTORICAL REFLECTIONS ON THE LOWEST-ORDER FINITE ELEMENT MODELS FOR THIN STRUCTURES

Juhani Pitkäranta

TEKNILLINEN KORKEAKOULU
TEKNISKA HÖGSKOLAN
HELSINKI UNIVERSITY OF TECHNOLOGY
TECHNISCHE UNIVERSITÄT HELSINKI
UNIVERSITE DE TECHNOLOGIE D'HELSINKI

Helsinki University of Technology Institute of Mathematics Research Reports
Teknillisen korkeakoulun matematiikan laitoksen tutkimusraporttisarja
Espoo 2003 A449

MATHEMATICAL AND HISTORICAL REFLECTIONS ON THE LOWEST-ORDER FINITE ELEMENT MODELS FOR THIN STRUCTURES

Juhani Pitkäranta

Helsinki University of Technology
Department of Engineering Physics and Mathematics
Institute of Mathematics

Juhani Pitkäranta: *Mathematical and historical reflections on the lowest-order finite element models for thin structures*; Helsinki University of Technology Institute of Mathematics Research Reports A449 (2003).

Abstract: *We discuss the mathematical theory and history of the lowest-order linear and bilinear finite element models for beams, arches, plates and shells. The finite element formulations considered are based on the non-asymptotic Timoshenko beam and Reissner-Mindlin plate models and the analogies of these models for arches and shells. We follow some of the historical roots of the successful linear and bilinear elements, to find various physical justifications for formulations that now may be understood as purely numerical modifications within the usual energy principle. The simplified mathematical theory of such formulations is outlined, first in cases of the beam, arch and plate. We finally focus on the still challenging and largely open problems arising in the modelling of shell deformations. We consider here a simplified shallow shell model and an interpretation of the MITC4 shell element within that model, called MITC4-S. We sum up the results of the recent finite element theory for MITC4-S, concerning the approximation of bending- and membrane-dominated deformations of a shallow shell.*

AMS subject classifications: 73V05, 65N30

Keywords: finite elements, thin bodies, history

juhani.pitkaranta@hut.fi

ISBN 951-22-6001-8
ISSN 0784-3143
Espoo, 2002

Helsinki University of Technology
Department of Engineering Physics and Mathematics
Institute of Mathematics
P.O. Box 1100, 02015 HUT, Finland
email:math@hut.fi http://www.math.hut.fi/

1 Introduction

One of the dominant trends in computational structural mechanics over the last fifty years is the intensive search of simple low-order finite element formulations that avoid parametric locking when modelling thin structures like beams, arches, plates, and shells. The thin structures have in common that when modelled with standard finite elements of lowest order, convergence typically fails completely unless the mesh spacing is set below the thickness of the body. Our focus here is on the most challenging of these problems, the shell problem. For shells, the construction of efficient low-order elements has been attempted practically over the entire history of the finite element method. A number of special finite element constructions known as "shell elements" has resulted. We consider here one of the scientifically open ones of these formulations, the bilinear shell element of the code ADINA, known as MITC4. The formulation is due to Bathe and Dvorkin [1, 2]. Our aim is to put this element in a mathematical and historical frame in view of the theory developed recently [3, 4, 5, 6].

Our mathematical approach to MITC4 is based on a simplified "twin" formulation of the element in the context of shallow shell models [3, 4]. The simplified formulation, called MITC4-S, preserves the essential numerical ideas involved in the original (three-dimensional) formulation but makes the ideas more transparent for mathematical error analysis. In essence, the MITC4-S interprets the physical and geometric assumptions of the original 3D formulation as purely numerical modifications of the otherwise standard bilinear scheme for a classical 2D shell model.

The numerical modifications in MITC4, when uncovered in the mentioned way, turn out to have a long history. We find that some of the historical roots actually lead back to the early 1950's (or perhaps to the 1940's), to the prime stock of finite element methodology, then known as the "matrix methods" of structural analysis [7]. The two main ideas that we find already here are the modifications required to make the C^0 linear finite element locking-free (as we now say) a) in the Timoshenko beam problem and b) in the corresponding parametric (non-asymptotic) model for an arch. As often in finite element engineering, the original justification of the numerical schemes was intuitive, or physical. However, as in case of the MITC4 shell element, we can now understand these formulations in retrospect as purely numerical modifications within the standard linear element framework. In this way we see more clearly the historical connections and, what is more important, we can understand the actual mathematical reasoning behind — which typically is numerical rather than physical.

Due to the historical connections involved, and also to make the mathematical theory more transparent, we follow the historical order in our presentation. We start from the one-dimensional beam and arch problems to explain first the oldest ideas of dealing with the numerical (shear and membrane) locking in these parametric problems. We can explain these ideas quite easily in retrospect, when using the latest (version 2000) update of finite element theory as presented in [8, Section 6]. Once the one-dimensional locking effects and their classical remedies are understood, we take a step from the Timoshenko beam to its two-dimensional analogue, the Reissner-Mindlin plate model. Here we follow the historical evolution of the simplest bilinear elements. We observe first an unsuccessful first attempt, then a surprising success in a simple formulation by MacNeal [9], Hughes and Tezduyar [10], and (a bit later) by Bathe and Dvorkin [1, 11]. This formulation is still one of the great "bilinear miracles" in the FEM models of structures. Another (somewhat related) bilinear phenomenon is found in plane elasticity [8]. In plate element technology, the QUAD4 of NASTRAN and the MITC4 of ADINA are among the trademarks built upon this successful formulation. The mathematical justification was first given by Bathe and Brezzi [12]. We outline the reasoning briefly using the later evolution of the theory as presented in [8].

We finally take the step from plate bending to the most challenging problem in classical structural mechanics: the shell problem. Many attempts have been made to achieve a finite element control over the various locking effects that disturb the FEM modelling of shells. After a series of successful formulations in the mentioned simpler parametric problems, one would naturally expect a final success also here. However, the step from structures like beams, arches and plates to a shell is very long mathematically. The mentioned simpler structures all have relatively simple asymptotic behavior at zero thickness, whereas the shell problem breaks into a multitude of subproblems, each with its own asymptotics and characteristic locking phenomena, c.f. [13, 14]. For such reasons, it appears less likely that some magic bilinear (or other low-oder) finite element formulation could handle all these parametric effects simultaneously. — Note, for example, that a fully locking-free four-node plane elastic element of arbitrary quadrilateral shape is still a dream element only [8]. — Anyway, the MITC4 shell element, and its suspected relatives like the QUAD4 of NASTRAN, are brave attempts to beat a very strong enemy using traditional weapons. One should analyze these attempts mathematically under realistic assumptions, in order to be able (at least) to set the actual limits of the possible. We sum up briefly what the theory of MITC4-S can say so far.

2 The beam and the "shear trick"

According to the Timoshenko beam model, the total energy of a beam loaded by a distributed normal load f is

$$\mathcal{F}(w,\theta) = \frac{D}{2}\int_0^L \left[\left(\frac{d\theta}{dx}\right)^2 + \frac{k}{t^2}\left(\theta - \frac{dw}{dx}\right)^2\right] dx \;-\; \int_0^L f w\, dx, \tag{2.1}$$

where L is the length and t the depth of the beam, w is the transverse deflection and θ the rotation of the cross-section. Assuming rectangular cross-section we have $D = Et^3/12$ and $k = 12\gamma G/E$, where E is the Young modulus, G the shear modulus, and γ the shear correction factor. We write the energy shortly as

$$\mathcal{F}(\mathbf{u}) = \frac{1}{2}|||\mathbf{u}|||^2 - \mathcal{L}(\mathbf{u}), \tag{2.2}$$

where $\mathbf{u} = (w,\theta)$ is the (generalized) displacement field, $|||\cdot|||$ is the energy norm (= square root of the strain energy), and $\mathcal{L}$ is the load functional (= potential energy of the load). In a standard finite element scheme one minimizes the energy as given, over a chosen finite element space U_h and under the kinematic constraints of the problem. This defines the finite element solution $\mathbf{u}_h = (w_h, \theta_h)$ as the best approximation of the exact solution $\mathbf{u}$ in the energy norm, so a rather natural error indicator is the relative error in that norm:

$$e(\mathbf{u}_h) = \frac{|||\mathbf{u} - \mathbf{u}_h|||}{|||\mathbf{u}|||}. \tag{2.3}$$

This indicator is obviously dimensionless and scaling-invariant.

To obtain a bound for the error in the sense of Eq. (2.3), one needs basically only the mentioned best approximation property

$$|||\mathbf{u} - \mathbf{u}_h||| \;=\; \min_{\mathbf{v} \in U_h} |||\mathbf{u} - \mathbf{v}|||. \tag{2.4}$$

In a parametric situation like the one considered, however, one should also know how the denominator in Eq. (2.3) scales with the parameters. In a complex problem like a shell problem this can be far from obvious, but here we can resolve the parametric dependence of $|||\mathbf{u}|||$ quite easily. The parameters in the energy norm are D, k and t, but D is inactive due to the scaling invariance of the error indicator, and the dimensionless parameter k may be considered harmless as well. Thus the only truly active parameter is t — or rather the ratio t/L. From the classical asymptotics of the problem we know that bending effects dominate when t/L is small, so we may assume that the denominator in Eq. (2.3) is essentially independent of t/L and scales like

$$|||\mathbf{u}|||^2 \sim \int_0^L D \left(\frac{d\theta}{dx}\right)^2 dx, \quad \mathbf{u} = (w, \theta). \tag{2.5}$$

(For a different type of loading, the scaling of $|||\mathbf{u}|||$ could be different.)

Consider now the simplest finite element approximation of the above problem, as based on the two-node linear element with w and θ approximated independently. To estimate the error $e(\mathbf{u}_h)$ of such an approximation, use first Eq. (2.4) to conclude that

$$|||\mathbf{u} - \mathbf{u}_h||| \;\le\; |||\mathbf{u} - \tilde{\mathbf{u}}|||, \tag{2.6}$$

where $\tilde{\mathbf{u}} = (\tilde{w}, \tilde{\theta})$ is the interpolant of $\mathbf{u}$ in U_h. Then use standard interpolation error estimates, assuming that the solution is sufficiently smooth (c.f. [8]), and Eq. (2.5), to conclude that the error is at most of order

$$e(\mathbf{u}_h) = \mathcal{O}\left(\frac{h}{t}\right), \tag{2.7}$$

where h is the (maximal) mesh spacing. Here the factor $1/t$ appears because of the parametric dependence of the energy norm, and because the denominator in Eq. (2.3) is essentially independent of t as assumed in Eq. (2.5). — Note that for the lowest-oder FEM, the optimal error bound should be $e(\mathbf{u}_h) = \mathcal{O}(h/L)$ when the deformation is smooth in the length scale L, as may be assumed here. Thus the error bound (2.7) predicts error magnification by factor L/t from the optimal rate when the beam aspect ratio t/L is small. Practice confirms that this prediction is not pessimistic — in fact, a separate *lower* bound for the error shows that also, see [8, Section 4]. We are facing a typical *parametric error amplification* or *locking* effect in a low-order finite element model.

A remedy for the above problem was found quite early, in fact, by the early FEM pioneers. We may consult here one of the classics, the paper by Turner et al. of 1956 [15] (the idea of the remedy is probably older). In [15], several low-order finite element formulations (as now called) are presented, one of which is a modified linear element for the Timoshenko beam problem. (A related, modified bilinear plane elastic formulation is also found in [15]. The theory and later successors of this were discussed in [8].) Consider a reference beam element with nodes at $x = -h/2$ and $x = h/2$. In the standard linear element one proceeds from the local expansion

$$(w, \theta) \;=\; A_1\,(1,0) + A_2\,(x,0) + B_1\,(0,1) + B_2\,(0,x). \tag{2.8}$$

Turner et al. [15] propose intstead the expansion

$$(w, \theta) \;=\; A_1\,(1,0) + A_2\,(x,0) + B_1\,(0,1) + B_2\,(\tfrac{1}{2}x^2 - \tfrac{1}{8}h^2, x). \tag{2.9}$$

Note that as compared with Eq. (2.8), this expansion may be considered more natural physically, as the last term here is the simplest asymptotic bending mode of the element when considered a beam of length h. So, with good understanding of physics (and with perhaps less understanding of finite elements), one might consider expansion (2.9) as the most natural first attempt. Anyway, let us consider this a modification of the (to us more natural) linear expansion (2.8). Note that we can still use the ordinary nodal degrees of freedom after the modification, since the added quadratic term in Eq. (2.9) vanishes at the nodal points.

When embedded in the energy formulation, the above modification apparently leaves the first (bending) term $(d\theta/dx)^2$ unchanged in Eq. (2.1). In the second (shear) term, the added quadratic term has the effect of cancelling the linear part of $\theta - dw/dx$ in each element, so the effect is the same as if we used the standard linear element together with the modification

$$\theta - \frac{dw}{dx} \;\hookrightarrow\; \Pi_h \left(\theta - \frac{dw}{dx} \right), \tag{2.10}$$

where Π_h stands for the operator of averaging over each element. The quadratic term in Eq. (2.9) finally affects also the third (load energy) term in Eq. (2.1), but this effect is easily shown to be small, so we ignore that below. We thus end up in a linear finite element scheme where the formulation is standard, except for the "shear trick" (2.10). By experiment, this modified scheme (which is as simple as the standard linear scheme) works very well: The error amplification on thin beams is no more observed.

The above successful formulation can be achieved in many ways, as shown by the later evolution of finite element methodology. Following roughly the historical order, the main alternatives of the above derivation are:

(1) *Mixed method*: Instead of the energy principle, use a mixed variational formulation where the shear stress $q = (k/t^2)\,(\theta - dw/dx)$ acts as an independent unknown. Approximate q by a piecewise constant function in the FEM model.

(2) *Reduced integration* (Underintegration): Evaluate the strain energy numerically using the elementwise midpoint rule.

(3) *Mixed interpolation*: Interpret Π_h in Eq. (2.10) as the interpolation operator at the midpoints of the elements.

Let us now consider the above formulation in view of the finite element theory. We want an error bound, hopefully of the optimal order $\mathcal{O}(h/L)$, and an (a posteriori) explation that presumably comes with the error analysis. In the theory, we have many options as well. For example, we could follow the original formulation above, or we could use the mixed finite element theory, as was done in the first error analysis by Arnold [16]. Here we outline what is perhaps the most straightforward theoretical reasoning, following the guidelines of [8, Section 6].

We start by defining a modified error indicator

$$e(\mathbf{u}_h) = \frac{|\!|\!|\mathbf{u} - \mathbf{u}_h|\!|\!|_h}{|\!|\!|\mathbf{u}|\!|\!|}, \tag{2.11}$$

where $|\!|\!| \cdot |\!|\!|_h$ is the modified energy norm. (We interpret Π_h in Eq. (2.10) as the averaging operator.) Following [8] we split the error (2.11) in two parts, called the *approximation error* and the *consistency error*. (The terminology comes from the tradition of finite element theory. The consistency error is sometimes referred to as the *equilibrium error* in the literature. See the further references in [8].) The approximation error $e_a(\mathbf{u}_h)$ is defined simply as the error of the best approximation according to indicator (2.11), i.e.,

$$e_a(\mathbf{u}_h) = \min_{\mathbf{v} \in U_h} \frac{|\!|\!|\mathbf{u} - \mathbf{v}|\!|\!|_h}{|\!|\!|\mathbf{u}|\!|\!|}. \tag{2.12}$$

(The kinematic constraints are obeyed here.) The remaining part of the error, the consistency error $e_c(\mathbf{u}_h)$, is then

$$e_c(\mathbf{u}_h) = \frac{|\!|\!|\mathbf{z}_h|\!|\!|_h}{|\!|\!|\mathbf{u}|\!|\!|}, \tag{2.13}$$

where $\mathbf{z}_h$ is the finite element solution (according to the modified scheme) when a) the kinematic constraints of the problem are replaced by the corresponding homogeneous constraints, and b) the load functional $\mathcal{L}$ is replaced by the generalized load

$$\ell_h(\mathbf{v}) = \mathcal{A}(\mathbf{u}, \mathbf{v}) - \mathcal{A}_h(\mathbf{u}, \mathbf{v}), \tag{2.14}$$

where $\mathbf{u}$ is the exact solution and $\mathcal{A}(\cdot, \cdot)$ is the energy inner product and $\mathcal{A}_h(\cdot, \cdot)$ its counterpart after modification (2.10).

To bound the approximation error, we choose $\mathbf{v} = (\tilde{w}, \tilde{\theta})$ in Eq. (2.12), where $\tilde{w} \approx w$ and $\tilde{\theta} \approx \theta$ are properly chosen piecewise linear approximations. The idea is to choose these as generalized interpolants under the constraint

$$\Pi_h \left[\theta - \tilde{\theta} - \frac{d}{dx}(w - \tilde{w}) \right] = 0. \tag{2.15}$$

Since here $\Pi_h(d\tilde{w}/dx) = d\tilde{w}/dx$, we can solve Eq. (2.15) for $\tilde{w}$, and thus we achieve our goal by first choosing $\tilde{\theta}$ (so far freely) and then solving $\tilde{w}$ from Eq. (2.15). Setting $\tilde{w}(0) = w(0)$ (as possibly forced by a kinematic coinstraint) we conclude that $\tilde{w}$ is then defined uniquely at each nodal point x_j as

$$\begin{aligned}\tilde{w}(x_j) &= w(0) + \int_0^{x_j} \Pi_h \frac{dw}{dx}\, dx - \int_0^{x_j} \Pi_h(\theta - \tilde{\theta})\, dx \\ &= w(0) + \int_0^{x_j} \frac{dw}{dx}\, dx - \int_0^{x_j} (\theta - \tilde{\theta})\, dx \\ &= w(x_j) - \int_0^{x_j} (\theta - \tilde{\theta})\, dx. \end{aligned} \tag{2.16}$$

Since the choice of $\tilde{\theta}$ is so far free, we may enforce here the possible kinematic constraint $\tilde{w}(L) = w(L)$ by choosing $\tilde{\theta}$ so that

$$\int_0^L (\theta - \tilde{\theta})\, dx = 0. \tag{2.17}$$

This is the only constraint that limits the choice of $\tilde{\theta}$.

With $\tilde{\theta}$ chosen so that Eq. (2.17) holds (otherwise so far freely) and $\tilde{w}$ defined by Eq. (2.16), set $\mathbf{v} = (\tilde{w}, \tilde{\theta})$. Then by Eq. (2.15)

$$\|\mathbf{u} - \mathbf{v}\|_h \;=\; \left\{ D \int_0^L \left[\frac{d}{dx}(\theta - \tilde{\theta}) \right]^2 dx \right\}^{1/2}. \tag{2.18}$$

Denoting henceforth by $\|\cdot\|$ the L_2 - norm on the interval $[0, L]$, i.e.,

$$\|\phi\| \;=\; \left[\int_0^L \phi^2\, dx \right]^{1/2}, \tag{2.19}$$

we thus conclude from Eqs. (2.18) and (2.12) that

$$e_a(\mathbf{u}_h) \;\le\; \frac{\|\mathbf{u} - \mathbf{v}\|_h}{\|\mathbf{u}\|} \;=\; \frac{D^{1/2}\, \|\theta' - \tilde{\theta}'\|}{\|\mathbf{u}\|}. \tag{2.20}$$

Define now finally $\tilde{\theta}$ by minimizing the right side of Eq. (2.20) under constraint (2.17). This defines $\tilde{\theta}$ as a slightly deflected interpolant that satisfies the sharp error bound [8, Section 10]

$$\|\theta' - \tilde{\theta}'\| \;\le\; Ch\, \|\theta''\|, \quad C \;=\; \frac{1}{\pi} + \mathcal{O}\left(\frac{h}{L}\right). \tag{2.21}$$

Combining Eqs. (2.20) and (2.21) we conclude that the approximation error can be bounded as

$$e_a(\mathbf{u}_h) \;\le\; CKQ\, \frac{h}{L}, \tag{2.22}$$

where C comes from Eq. (2.21) and K and Q are likewise dimensionless constants that depend on the exact solution as

$$K = \frac{D^{1/2}\|\theta'\|}{|\!|\!|\mathbf{u}|\!|\!|}, \quad Q = \frac{L\,\|\theta''\|}{\|\theta'\|}. \tag{2.23}$$

Note that here K^2 defines the ratio of the bending energy to the total deformation energy in the exact deformation state, whereas Q relates to the regularity of the exact solution. At the asymptotic limit $t/L = 0$ one has $K = 1$. In that case the constraint (2.15) must be forced when bounding the approximation error, so our analysis is sharp when $t/L = 0$. We note that the bound (2.21) is not improvable under further regularity assumptions on θ, only the value of C can be reduced slightly. The smallest possible value of C, obtained when $\theta'''(x)$ is square integrable over the interval $[0, L]$, is $C = 1/\sqrt{12} + \mathcal{O}(h/L)$. (The value of C given in Eq. (2.21) is for the worst case where the mesh is uniform and $\theta = \theta(x, h) = \sin \pi x/h$. The improved value is obtained when θ is a quadratic polynomial.)

To bound the consistency error, we note that by the general theory [8], we have the bound $e_c(\mathbf{u}_h) \leq \varepsilon_h$ whenever the generalized load functional (2.14) obeys the bound

$$|\ell_h(\mathbf{v})| \;\leq\; \varepsilon_h |\!|\!|\mathbf{u}|\!|\!| \, |\!|\!|\mathbf{v}|\!|\!|_h \quad \forall \mathbf{v} \in U_h. \tag{2.24}$$

We can write $\ell_h(\mathbf{v})$ in terms of the shear stress

$$q = \frac{k}{t^2}\left(\theta - \frac{dw}{dx}\right) \tag{2.25}$$

as

$$\ell_h(\mathbf{v}) \;=\; \int_0^L \left[q\left(\eta - \frac{d\xi}{dx}\right) - \Pi_h q\, \Pi_h\left(\eta - \frac{d\xi}{dx}\right)\right] dx, \quad \mathbf{v} = (\xi, \eta) \in U_h. \tag{2.26}$$

Since Π_h is here an averaging operator and $d\xi/dx$ is piecewise constant, we can simplify Eq. (2.26) as

$$\ell_h(\mathbf{v}) \;=\; \int_0^L (q - \Pi_h q)(\eta - \Pi_h \eta)\, dx, \quad \mathbf{v} = (\xi, \eta) \in U_h, \tag{2.27}$$

so that by the Cauchy-Schwarz inequality

$$|\ell_h(\mathbf{v})| \;\leq\; \|q - \Pi_h q\| \|\eta - \Pi_h \eta\|. \tag{2.28}$$

Bounding here the first term on the right side as

$$\|q - \Pi_h q\| \leq \|q\|, \tag{2.29}$$

the second term by standard approximation theory as

$$\|\eta - \Pi_h \eta\| \;\leq\; \frac{1}{\pi}\, h\, \|\eta'\|, \tag{2.30}$$

noting that by the Euler equations of the beam model,

$$q = \frac{d^2\theta}{dx^2}, \tag{2.31}$$

and finally noting that

$$\|\eta'\| \leq D^{-1/2}|||\mathbf{v}|||_h, \quad \mathbf{v} = (\xi, \eta) \in U_h, \tag{2.32}$$

we conclude, combining Eqs. (2.28)–(2.32), that ε_h in (2.24) and hence the consistency error can be bounded as

$$e_c(\mathbf{u}_h) \leq \varepsilon_h \; \leq \; CKQ\,\frac{h}{L}, \tag{2.33}$$

where $C = 1/\pi$ and costants K and Q are defined by Eq. (2.23).

From the analysis so far we conclude that the approximation and consistency errors are both of order $\mathcal{O}(h/L)$. This result was obtained assuming that θ'' is square integrable, in which case the upper bounds in Eqs. (2.22) and (2.33) are nearly the same. As noted, the approximation error bound (2.22) is essentially optimal even under stronger regularity assumptions. Instead, the consistency error bound can be improved considerably by assuming that θ''' is square integrable. Namely, recalling Eq. (2.31), we can then replace Eq. (2.29) by the bound

$$\|q - \Pi_h q\| \;\leq\; \frac{1}{\pi}\,h\,\|q'\| \;=\; \frac{1}{\pi}\,h\,\|\theta'''\|. \tag{2.34}$$

Following the above reasinoning we then conclude that

$$e_c(\mathbf{u}_h) \leq C_1 K Q_1 \left(\frac{h}{L}\right)^2, \quad C_1 = \frac{1}{\pi^2}, \quad Q_1 = \frac{L^2\,\|\theta'''\|}{\|\theta'\|}. \tag{2.35}$$

Thus the consistency error is of order $\mathcal{O}(h/L)$ or $\mathcal{O}(h^2/L^2)$, depending on how smooth the solution is assumed to be.

The total error is bounded by the sum of the approximation and consistency error terms, ore more precisely [8]

$$e(\mathbf{u}_h) = [\, e_a^2(\mathbf{u}_h) + e_c^2(\mathbf{u}_h)\,]^{1/2}. \tag{2.36}$$

We conclude that the modified linear finite element scheme is free of parametric error amplification when the error is measured by indicator (2.11). As the error analysis shows, the success is mostly based on the existence of an accurate (in the sense of indicator (2.11)) generalized interpolant of the exact solution. The "interpolant" constructed above is actually close to the finite element solution itself in the case where t/L is and h/L are both small and the consitency error is of order $\mathcal{O}(h^2/L^2)$ so that this error term is negligible. In that case the above theory thus not only proves convergence but also explains how the finite element algorithm actually works when minimizing the modified energy.

3 The arch and the "beam trick"

Consider a circular arch of radius R and length L, of rectangular cross-section, and subject to a smoothly varying (in the length scale L) traction load distribution $\mathbf{f} = (f_1, f_2)$, where f_1 is the tangential and f_2 the normal load component. When the depth to length ratio t/L is mall, the total energy of the arch is given approximately by the expression

$$\mathcal{F}(u,w,\theta) = \frac{D}{2}\int_0^L \left[\left(\frac{d\theta}{dx}\right)^2 + \frac{k}{t^2}\left(\theta - \frac{dw}{dx} + \frac{u}{R}\right)^2 + \frac{m}{t^2}\left(\frac{du}{dx} + \frac{w}{R}\right)^2\right] dx$$
$$- \int_0^L (f_1 u + f_2 w)\, dx, \qquad (3.1)$$

where the three terms of the strain energy correspond to bending, transverse shear, and stretching/compression. Here x is the arc length variable, w is the transverse deflection (positive when the deflection is away from the center of curvature), θ is again the rotation of the cross-section, and u is the tangential displacement. The coefficients D and k are the same as before, and $m = 12$.

We will again assume the exact solution $\mathbf{u} = (u, w, \theta)$ to be such that Eq. (2.5) holds, i.e., the first bending term in Eq. (3.1) is dominant. For the majority of the problems of the assumed type this holds, but there are also cases where this assumption does *not* hold. In such exceptional cases the load term in Eq. (3.1) takes the specific form

$$\int_0^L (f_1 u + f_2 w)\, dx \;=\; \int_0^L R f_2 \left(\frac{du}{dx} + \frac{1}{R} w\right) dx. \qquad (3.2)$$

A necessary condition is that

$$f_1 = -R\frac{df_2}{dx}, \qquad (3.3)$$

which condition is also sufficient when the kinematic constraints $u(0) = u(L) = 0$ are imposed. When Eq. (3.2) holds, the deformation energy is concentrated on the third (stretching) term in Eq. (3.1) at small values of t/L. Such stretching-dominated deformation states violate assumption (2.5) and are thus excluded from our consideration. We come back to this problem when considering different deformation states of a shell in Section 5.

Obviously, the simplest finite element approximation of the above problem is again based on the two-node linear element where now the three components of the displacement field $\mathbf{u} = (u, w, \theta)$ are approximated independently. From the analogous beam model we may deduce that to avoid parametric error amplification due to the shear energy term in Eq. (3.1), we should perform the "shear trick"

$$\theta - \frac{dw}{dx} + \frac{u}{R} \;\hookrightarrow\; \Pi_h\left(\theta - \frac{dw}{dx} + \frac{u}{R}\right), \qquad (3.4)$$

where Π_h is the averaging (or midpoint interpolation) operator. Indeed, this can again be justified physically by assuming first that w is quadratic on each element and then eliminating the quadratic part using the known asymptotics of the model. Moreover, the same reasoning could be applied to the stretching energy term in Eq. (3.1) as well: Assuming first that u is likewise quadratic on each element and eliminating the quadratic term by imposing partly the asymptotic assumption $du/dx + w/R = 0$, one would be lead to the "stretching trick"

$$\frac{du}{dx} + \frac{w}{R} \hookrightarrow \Pi_h\left(\frac{du}{dx} + \frac{w}{R}\right). \tag{3.5}$$

This indeed is the right thing to in the linear-element model, but the original physical justification was different. In the old matrix methods for arches, it was considered more natural to approximate the arch just as an assembly of beams when evaluating the stretching energy [7]. Although such a "beam trick" appears rather violent, it proved efficient experimentally when assumed in the context of the lowest-order finite element approximation. The first full mathematical justification was given by Kikuchi [17]. Following the reasoning in [17], consider an element with nodes at $x_{j-1} = -h/2$ and $x_j = h/2$. Then under the beam approximation of the element, the relative stretching of the beam in terms of the nodal degrees of freedom of u and w is given by

$$\beta = \frac{1}{\tilde{h}}\left[\cos\left(\frac{h}{2R}\right)(u_j - u_{j-1}) + \sin\left(\frac{h}{2R}\right)(w_{j-1} + w_j)\right], \tag{3.6}$$

where $\tilde{h}$ is the length of the beam. Using here the approximations

$$\tilde{h} \approx h, \quad \cos\left(\frac{h}{2R}\right) \approx 1, \quad \sin\left(\frac{h}{2R}\right) \approx \frac{h}{2R}, \tag{3.7}$$

we get

$$\beta \;\approx\; \frac{1}{h}(u_j - u_{j-1}) + \frac{1}{2R}(w_{j-1} + w_j) \;=\; \Pi_h\left(\frac{du}{dx} + \frac{w}{R}\right). \tag{3.8}$$

The beam assumption thus has practically the same effect as the modification (3.5). — We can add one more item to the list of different derivations of the same numerical trick!

The error analysis of the linear finite element scheme with modifications (3.4) and (3.5) can be carried out along the same lines as for the beam. (The error analysis shows very little difference between the beam assumption and modification (3.5), so we simply assume the latter.) We define again the error indicator by Eq. (2.11), where $\|\cdot\|_h$ is the modified energy norm (= square root of the modified strain energy), and where we assume that the denominator is scaled according to Eq. (2.5). When bounding the approximation error, we construct again a generalized interpolant $\tilde{\mathbf{u}} = (\tilde{u}, \tilde{w}, \tilde{\theta}) \in U_h$ of the exact solution $\mathbf{u} = (u, w, \theta)$ in such a way that the interpolation error is insensitive to the main parameter t. This is achieved by constructing $(\tilde{u}, \tilde{w}, \tilde{\theta})$ in such a way that

$$\begin{aligned} \Pi_h\left[\theta - \tilde{\theta} - \frac{d}{dx}(w - \tilde{w}) + \frac{1}{R}(u - \tilde{u})\right] &= 0, \\ \Pi_h\left[\frac{d}{dx}(u - \tilde{u}) + \frac{1}{R}(w - \tilde{w})\right] &= 0. \end{aligned} \tag{3.9}$$

The construction is analogous to that given above for the beam. Given first $\tilde{\theta}$, we define $\tilde{w}, \tilde{u}$ at each nodal point x_j so that

$$\begin{aligned} \tilde{w}(x_j) &= w(x_j) - \int_0^{x_j} (\theta - \tilde{\theta})\, dx - \frac{1}{R}\int_0^{x_j} (u - \tilde{u})\, dx, \\ \tilde{u}(x_j) &= u(x_j) + \frac{1}{R}\int_0^{x_j} (w - \tilde{w})\, dx. \end{aligned} \tag{3.10}$$

That this system is solvable for $\tilde{w}(x_j), \tilde{u}(x_j)$ is shown in [18]. To impose the possible kinematic constraints $\tilde{u}(L) = u(L)$, $\tilde{w}(L) = w(L)$, one needs to restrict $\tilde{\theta}$ by two integral constraints, otherwise the choice of $\tilde{\theta}$ remains free [18]. With $\mathbf{v} = (\tilde{u}, \tilde{w}, \tilde{\theta})$ defined in this way, we may again bound the approximation error by the minimum of $\|\mathbf{u} - \mathbf{v}\|_h$ with respect to the degrees of freedom of $\tilde{\theta}$ that are left free by the mentioned constraints. We conclude then that Eqs. (2.20)–(2.23) remain valid also in the arch problem. Thus the approximation error is again of the uniformly optimal order $\mathcal{O}(h/L)$ when θ'' is square integrable.

The consistency error analysis is likewise a straightforward extension of that for the beam above: We can expand the generalized load functional (2.14) this time as

$$\begin{aligned} \ell_h(\mathbf{v}) \;=\; & \int_0^L (q - \Pi_h q)(\eta - \Pi_h \eta)\, dx \;+\; \int_0^L R^{-1}(q - \Pi_h q)(v - \Pi_h v)\, dx \\ & + \int_0^L R^{-1}(\sigma - \Pi_h \sigma)(\xi - \Pi_h \xi)\, dx, \quad \mathbf{v} = (v, \xi, \eta) \in U_h, \end{aligned} \tag{3.11}$$

where

$$q = \frac{k}{t^2}\left(\theta - \frac{dw}{dx} + \frac{u}{R}\right), \quad \sigma = \frac{m}{t^2}\left(\frac{du}{dx} + \frac{w}{R}\right). \tag{3.12}$$

Using Eq. (3.11) together with the Euler equations

$$-\frac{d\sigma}{dx} + \frac{1}{R}q = f_1, \quad \frac{1}{R}\sigma + \frac{dq}{dx} = f_2, \quad q = \frac{d^2\theta}{dx^2}, \tag{3.13}$$

we conclude, by the same reasoning as above, that the consistency error is again of order $\mathcal{O}(h/L)$ or $\mathcal{O}(h^2/L^2)$, depending on the smoothness of θ, f_1 and f_2. We skip the further details here.

The final conclusion of the error analysis so far is that the linear finite element scheme with modification (2.10) in case of a beam or modifications (3.4)–(3.5) in case of an arch is the lowest-order "dream scheme" where the parametric locking of the standard linear element is completely removed without any additional cost. We also conclude that although this scheme has been supported by many kinds of arguments over the (pre)history of FEM, there remains really just one basic *numerical* idea in the end: the idea of averaging elementwise. We note that this idea, or the nearly equivalent idea of selective reduced integration by the midpoint rule, works also in the context of more general 2D arches or 3D rods with varying curvature, c.f. [19] and the further references therein.

4 The plate and the great bilinear element

According to the Reissner-Mindlin model of plate bending, the deformation of the plate is expressed in terms of the vector field $\mathbf{u} = (w, \boldsymbol{\theta})$, where w is the transverse deflection and $\boldsymbol{\theta} = (\theta_1, \theta_2)$ the rotation of the normal at the midsurface. Consider, as a model problem, a square plate with side length L and thickness t. Assume that the plate consists of homogeneous isotropic material with Young modulus E and Poisson ratio ν and that the plate is loaded by a normal pressure distribution f. Then the energy of the plate according to the Reissner-Mindlin model is given by

$$\begin{aligned}\mathcal{F}(\mathbf{u}) = & \frac{D}{2}\int_0^L\int_0^L [\,\nu(\kappa_{11}+\kappa_{22})^2 + (1-\nu)(\kappa_{11}^2 + 2\kappa_{12}^2 + \kappa_{22}^2)\,]\,dxdy \\ & + \frac{kD}{2t^2}\int_0^L\int_0^L \left[\left(\theta_1 - \frac{\partial w}{\partial x}\right)^2 + \left(\theta_2 - \frac{\partial w}{\partial y}\right)^2\right] dxdy \\ & - \int_0^L\int_0^L fw\,dxdy, \end{aligned} \tag{4.1}$$

where the coefficients are defined in terms of E, ν and the shear correction factor γ as

$$D = \frac{Et^3}{12(1-\nu^2)}, \quad k = 6\gamma(1-\nu), \tag{4.2}$$

and $\kappa_{ij} = \kappa_{ij}(\boldsymbol{\theta})$ are the bending strains as defined by

$$\kappa_{11} = \frac{\partial\theta_1}{\partial x}, \quad \kappa_{22} = \frac{\partial\theta_2}{\partial y}, \quad \kappa_{12} = \frac{1}{2}\left(\frac{\partial\theta_1}{\partial y} + \frac{\partial\theta_2}{\partial x}\right). \tag{4.3}$$

One of the simplest finite element approaches to te above problem is to choose a four-node bilinear element where the three components of the displacement field are approximated independently. The finite element space is then defined as $U_h = V_h \times V_h \times V_h$, where V_h is the scalar bilinear space associated to a given rectangular mesh. We focus on this basic approach below. (The story of the simplest triangular plate element is interesting as well — but that is another story.) When the energy (4.1) is minimized as given over the bilinear finite element space U_h, the relative nergy-norm error indicator again turns red, showing error magnification by factor L/t as compared with the optimal rate $\mathcal{O}(h/L)$. Indeed, we know that this happens even if the problem reduces to a beam problem ($f = f(x)$, periodic boundary conditions at $y = 0, L$). The question then arises, whether the error magnification could again be avoided by modifying the strain energy numerically. The modifications should obviously be focused on the second shear energy term of Eq. (4.1), as the large coefficient in front of this term is the source of the problem.

Let us pause here for a few historical comments. As we have seen, the simple linear element formulations for beams and arches have deep historical roots, leading back some 50 years in time. The same is true for the simplest locking-free bilinear element for plane elasticity, see [8]. In this perspective, the simplest bilinear Reissner-Mindlin plate elements seem to represent a historical anomaly, as these constructions are some 20 years old only. One possible explanation could be that the mentioned early formulations were, perhaps, somewhat lucky constructions based on occasional physical ideas, whereas in case of plate bending, the physical intuition alone was (perhaps) insufficient. Note that, when formulating the question as above, we have taken a purely numerical approach where no physical justification is asked for. In the engineering applications of FEM, a notable transition towards this kind of thinking occurred with the advention of selective reduced integration techniques in the early 1970's. This purely numerical approach was "elevated from the realm of tricks to a legitimate methodology" when it was found that there were many connections to the earlier mixed finite element methodology developed since the 1960's [20]. At the time of this new synthesis, the simple locking-free formulation of the bilinear plate-bending element was found. In retrospect it appears that the formulation was rather close once the right (numerical rather than physical) approach was taken.

The first systematic attempt to improve the performance of the bilinear plate-bending element by numerical modification of the shear strains appears to be that by Hughes et al. in 1977 [21]. Here, as inspired by the successful linear element formulation for the beam, one proposes in Eq. (4.1) the anologous modifications

$$\theta_1 - \frac{\partial w}{\partial x} \hookrightarrow \Pi_h^{xy}\left(\theta_1 - \frac{\partial w}{\partial x}\right), \quad \theta_2 - \frac{\partial w}{\partial y} \hookrightarrow \Pi_h^{xy}\left(\theta_2 - \frac{\partial w}{\partial y}\right), \tag{4.4}$$

where Π_h^{xy} is the elementwise averaging operator. Alternatively, Π_h^{xy} may be understood as an interpolation operator at the midpoints of the elements, or one may consider (as in [21]) the modification arising from selective reduced integration in the shear energy term using the midpoint rule. Empirically, this modification does improve the performance of the bilinear element (in many cases at least), but the theory developed some years later [22] raises questions. The error analysis gives the desired optimal convergence rate, but it appears that the analysis can go through only if the mesh is uniform, the exact solution is extremely smooth, and the kinematic constraint at the boundary are strong enough (a clamped boundary was assumed in [22]). The extreme assumptions are needed basically because the modified scheme (interpreted as a mixed finite element method in [22]) is not sufficiently stable. Note that in case of no kinematic constraints, modification (4.4) creates an unphysical zero-energy mode: If $\mathbf{u} = (w, \boldsymbol{\theta}) \in U_h$ is such that $\boldsymbol{\theta} = \mathbf{0}$ and w oscillates at the nodal points as $w(x_j, y_j) = (-1)^{i+j}$, then $\|\mathbf{u}\|_h = 0$. Although kinematic constraints can easily remove this mode, the problem of weak stability remains. – We note that in light of the current error analysis philosophy (as outlined above, see [8]), weak stability causes in general the magnification of the consistency error. Apparently the consistency error due to modification (4.4) is so large that it can be controlled only under extreme assumptions, like those made in [22].

The conclusion from the theory is thus that the bilinear plate element with modification (4.4) is an unsuccessful formulation. Meanwhile the practice had drawn the same conclusion and abandoned the approach, in favor of a much better formulation found in [9, 10] (see also [11]). In this alternative formulation — which still is the final word — one replaces Eq. (4.4) by the modifications

$$\theta_1 - \frac{\partial w}{\partial x} \;\hookrightarrow\; \Pi_h^x\left(\theta_1 - \frac{\partial w}{\partial x}\right), \quad \theta_2 - \frac{\partial w}{\partial y} \;\hookrightarrow\; \Pi_h^y\left(\theta_2 - \frac{\partial w}{\partial y}\right), \tag{4.5}$$

where Π_h^x and Π_h^y may still be considered elementwise averaging operators, but this time the averaging is done only in one of the coordinate directions, as indicated by the superscript. There are again other interpretations: One may think of the modification as arising alternatively from selective reduced integration [9] or from mixed interpolation [10, 11] (see below). Note that since $\partial w/\partial x$ is constant in x and $\partial w/\partial y$ is constant in y in Eq. (4.5), the modification does not change these terms and hence causes no unphysical zero-energy modes. In practice the formulation works extremely well, even when extended (properly, see the referneces cited) to more general quadrilateral element shapes.

The first error analysis of the above scheme was given by Bathe and Brezzi [12]. Let us try to translate this analysis into our language where we bound the error using the modified energy-norm error indicator (2.11) and split the error in two parts. The first task then is to decide, how the modifications (4.5) should be understood when acting on more general than piecewise bilinear functions. Note that we need this interpretation in Eq. (2.11), since the modifications in $\|\mathbf{u}-\mathbf{u}_h\|_h$ act on $\mathbf{u}-\mathbf{u}_h$. We consult here the finite element designers, the majority of whom seem to recommend *mixed interpolation* as the most natural approach, especially when extending the formulation to more general quadrilateral element shapes [10, 11]. Below we choose to follow this advice when extending the definition of operators Π_h^x and Π_h^y beyond the finite element space.

Consider a field $\mathbf{q} = (q_1, q_2)$ defined over the assumed rectangular domain and satisfying the (minimal) regularity assumption that q_1 is integrable with respect to x for all y and q_2 is integrable with respect to y for all x. We want to define $\Pi_h\mathbf{q} = (\Pi_h^x q_1, \Pi_h^y q_2)$ under such assumptions. To this end, let K be any given rectangle in the finite element mesh. The mixed-interpolation idea is then to define $\Pi_h\mathbf{q}$ on K as

$$\Pi_h\mathbf{q}\,(x,y) \;=\; (\,c_1 + c_2 y\,,\, c_3 + c_4 x\,), \quad (x,y) \in K, \tag{4.6}$$

where the constants c_i are defined by requiring that on each edge E of the rectangle

$$\int_E \mathbf{t}\cdot(\mathbf{q} - \Pi_h\mathbf{q})\, ds = 0, \tag{4.7}$$

where $\mathbf{t}$ is the tangent vector on E. Obviously this defines $\Pi_h\mathbf{q} = (\Pi_h^x q_1, \Pi_h^y q_2)$ uniquely on each element and hence on the entire domain. We also observe that in the case where (a) q_1 and q_2 are both bilinear on each K and (b) q_1 is continuous in y and q_2 is continuous in x, we have not changed the definition of Π_h^x and Π_h^y from that assumed above. Hence we can rewrite the modifications (4.5) now as

$$\boldsymbol{\rho} \;\hookrightarrow\; \Pi_h\boldsymbol{\rho}, \tag{4.8}$$

where $\boldsymbol{\rho}$ is the vector of shear strains:

$$\boldsymbol{\rho} = (\rho_1, \rho_2) = \left(\theta_1 - \frac{\partial w}{\partial x}\,,\, \theta_2 - \frac{\partial w}{\partial y}\right). \tag{4.9}$$

As noted, the derivative terms in Eq. (4.5) are left unchanged when $w \in V_h$, that is

$$w \in V_h \quad \Rightarrow \quad \Pi_h \nabla w = \nabla w. \tag{4.10}$$

This is an important property for stability.

Let us now bound the approximation error under the above interpretation of modifications (4.5). To obtain a bound that is uniform with respect to parameter t/L, we need to construct a generalized interpolant $\tilde{\mathbf{u}} = (\tilde{w}, \tilde{\boldsymbol{\theta}}) \in U_h$ of the exact solution $\mathbf{u} = (w, \boldsymbol{\theta})$ such that

$$\Pi_h [\, \boldsymbol{\theta} - \tilde{\boldsymbol{\theta}} - \nabla(w - \tilde{w}) \,] = \mathbf{0}. \tag{4.11}$$

Under this constraint one has

$$\| \mathbf{u} - \tilde{\mathbf{u}} \|_h = \| \boldsymbol{\theta} - \tilde{\boldsymbol{\theta}} \|_b, \tag{4.12}$$

where $\|\boldsymbol{\theta}\|_b$ is defined as the square root of the bending energy (the first term in Eq. (4.1)). When bounding the approximation error by choosing $\mathbf{v} = \tilde{\mathbf{u}} = (\tilde{w}, \tilde{\boldsymbol{\theta}})$ in Eq. (2.12), the question then is: How small can the right side of Eq. (4.12) be made under constraint (4.11)?

To solve the above problem, we follow the footsteps of Bathe and Brezzi [12]. First note that Eq. (4.11) is equivalent to stating that for each side E of the finite element mesh

$$\int_E \mathbf{t} \cdot [\, \boldsymbol{\theta} - \tilde{\boldsymbol{\theta}} - \nabla(w - \tilde{w}) \,] \, ds = 0. \tag{4.13}$$

For a side $E = \mathbf{a}_1 \mathbf{a}_2$ this is further equivalent to

$$\int_E \mathbf{t} \cdot (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}) \, ds \;=\; (w - \tilde{w})(\mathbf{a}_2) - (w - \tilde{w})(\mathbf{a}_1). \tag{4.14}$$

When summing Eq. (4.14) over the edges of a rectangle K we get

$$\int_{\partial K} \mathbf{t} \cdot (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}) \, ds \;=\; 0, \tag{4.15}$$

or equivalently

$$\int_K rot\, (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}) \; dxdy \;=\; 0. \tag{4.16}$$

where $rot\, \boldsymbol{\theta} = -\partial \theta_1 / \partial y + \partial \theta_2 / \partial x$. That this holds for every rectangle K of the finite element mesh is thus a necessary condition for the constraint (4.11) to be fulfilled. On the other hand, when $\tilde{\boldsymbol{\theta}}$ is given so that Eq. (4.16) holds for every rectangle of the mesh, we can also find $\tilde{w} \in V_h$ so that Eq. (4.11) holds. Namely, we may start from an interpolation condition at any given node and then proceed to the other nodes using Eq. (4.14) as the definition of $\tilde{w}$. No conflict arises in this construction under the stated condition on $\tilde{\boldsymbol{\theta}}$. A more explicit way of defining $\tilde{w}$ is simply to solve Eq. (4.11) for $\nabla \tilde{w}$. To this end, let $\breve{w}$ be the usual interpolant of w in V_h. Then by the definition of Π_h one has $\Pi_h(\nabla w - \nabla \breve{w}) = \mathbf{0}$, so that

$$\Pi_h \nabla w = \Pi_h \nabla \breve{w}. \tag{4.17}$$

But by Eq. (4.10) one has also $\Pi_h \nabla \breve{w} = \nabla \breve{w}$ and $\Pi_h \nabla \tilde{w} = \nabla \tilde{w}$, so Eq. (4.11) can be written equivalently as

$$\nabla \tilde{w} = \nabla \breve{w} - \Pi_h(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}). \tag{4.18}$$

We have now come to a similar situation as in the beam problem above: We observe that minimizing $\|\mathbf{u} - \tilde{\mathbf{u}}\|_h$ with respect to $\tilde{\mathbf{u}} = (\tilde{w}, \tilde{\boldsymbol{\theta}}) \in U_h$ under constraint (4.11) is the same as first minimizing $\|\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}\|_b$ with respect to $\tilde{\boldsymbol{\theta}} \in V_h \times V_h$ under the elementwise constraints (4.16) and then solving $\tilde{w}$ from Eq. (4.18). The question that remains is then: How small can the right side of Eq. (4.12) be made under the elementwise constraints (4.16)?

The above problem takes a more familiar form when writing Eq. (4.16) as

$$\int_K div\,(\boldsymbol{\phi} - \tilde{\boldsymbol{\phi}})\; dxdy \;=\; 0, \tag{4.19}$$

where $\boldsymbol{\phi} = (\phi_1, \phi_2) = (\theta_2, -\theta_1)$. Continuous piecewise bilinear approximation under constraint (4.19) is a well-known problem that arises in the bilinear/constant velocity-pressure finite element model for the Stokes flow. The problem appears also in plane elasticity when modelling coplanar deformations of a solid body consisting of nearly incompressible material [8]. That accurate approximation under the elementwise constraints (4.19) is indeed possible, is a rather unique "bilinear miracle" that seems to be hidden in quite many four-node element constructions developed over the history of FEM. In the present context the connection was found in [12].

The approximation theory of the bilinear element under constraint (4.19) was developed in [22, 23]. To apply the theory here, we need to split the exact solution in two parts as

$$\mathbf{u} = \mathbf{u}_0 + \mathbf{u}_1 = (w_0, \boldsymbol{\theta}_0) + (w_1, \boldsymbol{\theta}_1), \tag{4.20}$$

where $(w_0, \boldsymbol{\theta}_0)$ is the asymptotic Kirchhoff solution that satisifies the constraint $\boldsymbol{\theta}_0 = \nabla w_0$. (From this on we depart somewhat from the reasoning in [12]). If the boundary layer part of $\mathbf{u}$ is neglected, or if the layer is sufficiently weak, it is realistic to assume that the second term in the expansion (4.20) is by factor $\mathcal{O}(t/L)$ smaller in the energy norm than the first term, c.f. [24]. (In [24], the effect of the layer is studied as well. In general, the approximation of the boundary layer has to be studied as a separate problem, see the remarks ahead.) Under this assumption the small amplitude of $\mathbf{u}_1$ cancels the parametric amplification effect, so that the accuracy of standard interpolation is sufficient when bounding the approximation error due to this term. By this reasoning we have then isolated the main difficulty in the case where the field $\mathbf{u} = (w, \boldsymbol{\theta})$ to be approximated satisfies the (Kirchhoff) constraint $\boldsymbol{\theta} = \nabla w$. In that case one has $rot\,\boldsymbol{\theta} = 0$ in Eq. (4.16), or equivalently, $div\,\boldsymbol{\phi} = 0$ in Eq. (4.19). The theory in [23] then applies and states that in the model problem considered, the accuracy of standard interpolation is maintained under constraints (4.16). (In fact, the assumptions in [23] do not quite cover a general rectangular mesh, but the analysis there can be extended easily.) The conclusion is then that the approximation error is of the uniformly optimal order $\mathcal{O}(h/L)$, assuming that $\boldsymbol{\theta}$ is sufficiently smooth in the length scale L. One requires here the standard regularity assumption that the second partial derivatives of θ_1 and θ_2 are square integrable over the domain.

The consistency error analysis is more straightforward. We may expand the generalized load functional ℓ_h in Eq. (2.14) in this case as

$$\ell_h(\mathbf{v}) = \int_0^L \int_0^L (\mathbf{q} - \Pi_h \mathbf{q}) \cdot \Pi_h(\boldsymbol{\eta} - \nabla \xi)\, dxdy \\ + \int_0^L \int_0^L \mathbf{q} \cdot (\boldsymbol{\eta} - \Pi_h \boldsymbol{\eta})\, dxdy, \quad \mathbf{v} = (\xi, \boldsymbol{\eta}) \in U_h, \quad (4.21)$$

where $\mathbf{q} = (k/t^2)(\boldsymbol{\theta} - \nabla w)$ is the shear stress. Under reasonable regularity assumptions on $\mathbf{q}$ one may conclude from this expansion that the consistency error is likewise of order $\mathcal{O}(h/L)$ uniformly with respect to parameter t/L. More details of this reasoning are found in [25], where the analysis is carried out for a family of mixed-interpolated elements.

Our final conclusion is that, insofar as the simplest four-node plate element based on the Reissner-Mindlin model is concerned, the bilinear element with mixed interpolation of the shear strains according to Eq. (4.8) is the ultimate dream element. Variations of the element can be obtained, e.g. by adding one degree of freedom for the tangential rotation on each edge of the element and then eliminating the added d.o.f. by discrete Kirchhoff constraints [25]. We note that the basic idea of setting the degrees of freedom of mixed interpolation on the *edges* of the element seems the right approach also when the boundary layer is taken into account. Indeed, there are examples of other kinds of numerical modifications that result in good performance when approximating smooth solutions but cause unwanted error growth at the layer. Such error growth is seen most clearly at a free boundary where the layer is relatively strong, see [25, 26]. We note finally that when the above element is extended to more general quadrilateral element shapes as in [10, 11], the theory can largely follow. Some weak restrictions on the mesh do arise, due to the still incomplete theory of the constrained approximation problem above. Otherwise the theory confirms what is seen in practice: The element preserves its efficiency on quadrilateral meshes.

5 The shell and MITC4

We consider as a model problem a shallow shell such that the midsurface of the shell is a small deviation from a plane. We assume that the midsurface occupies a rectangular domain $[0, L] \times [0, L]$ in the coordinates x, y on the plane and that the thickness t of the shell is small compared with L. The curvature tensor $\{b_{ij}\}$ of the midsurface of the shell is assumed constant. Below we write $a = b_{11}$, $b = b_{22}$, and $c = b_{12} = b_{21}$. The shell is then classified geometrically to be elliptic when $ab - c^2 > 0$, parabolic when $ab - c^2 = 0$, and hyperbolic when $ab - c^2 < 0$.

To model the deformation of the shell when loaded, we assume the Naghdi (or Reissner-Naghdi) shell model where the deformation is expressed in terms of the membrane strains β_{ij}, transverse shear strains ρ_i, and bending strains κ_{ij}, each defined along the midsurface of the shell. The strains are associated to the five-component displacement field $\mathbf{u} = (u, v, w, \theta_1, \theta_2)$, where u, v and w are, respectively, the tangential displacements and the transverse deflection of the shell midsurface, and $\boldsymbol{\theta} = (\theta_1, \theta_2)$ is the rotation vector of the normal. We consider a simplified shell model where the membrane strains are defined as

$$\beta_{11} = \frac{\partial u}{\partial x} + aw, \quad \beta_{22} = \frac{\partial v}{\partial y} + bw, \quad \beta_{12} = \frac{1}{2}\left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right) + cw, \tag{5.1}$$

and the bending and the tranverse shear strains are defined by Eqs. (4.3) and (4.9), i.e., in the same way as in the plate-bending model. This model may be considered as an approximation to the geometrically accurate 2D models of classical shell theory, see [14] for the mathematical reasoning. Here we may consider the model as the simplest model that contains all the essential features of the (linear) shell problem from the finite element modelling point of view.

To express the energy of the shell in terms of the strains, we consider separately two cases, a *bending-dominated* deformation state and a *membrane-dominated* deformation state. — We note that the concept "shell problem" hides mathematical diversity not encountered in the simpler parametric problems considered above. In particular, shell problems can be classidfied (roughly, see [13]) depending on which type of deformation becomes dominant in the strain energy at the asymptotic limit $t/L \to 0$. In the bending-dominated case we write the energy as

$$\begin{aligned}
\mathcal{F}(\mathbf{u}) = & \frac{D}{2}\int_0^L\int_0^L [\,\nu(\kappa_{11}+\kappa_{22})^2 + (1-\nu)(\kappa_{11}^2 + 2\kappa_{12}^2 + \kappa_{22}^2)\,]\,dxdy \\
+ & \frac{kD}{2t^2}\int_0^L\int_0^L (\rho_1^2 + \rho_2^2)\,dxdy \\
+ & \frac{6D}{t^2}\int_0^L\int_0^L [\,\nu(\beta_{11}+\beta_{22})^2 + (1-\nu)(\beta_{11}^2 + 2\beta_{12}^2 + \beta_{22}^2)\,]\,dxdy \\
- & \int_0^L\int_0^L (f_1 u + f_2 v + f_3 w)\,dxdy.
\end{aligned} \tag{5.2}$$

Here the coefficients D, k are again defined by Eq. (4.2), and we have assumed loading in terms of a given surface traction along the midsurface. In the membrane-dominated case we define the scaling parameter D differently and also reorganize the strain energy so that the dominant term comes first, writing

$$\begin{aligned}
\mathcal{F}(\mathbf{u}) = & \frac{D}{2}\int_0^L\int_0^L [\,\nu(\beta_{11}+\beta_{22})^2 + (1-\nu)(\beta_{11}^2 + 2\beta_{12}^2 + \beta_{22}^2)\,]\,dxdy \\
+ & \frac{kD}{24}\int_0^L\int_0^L (\rho_1^2 + \rho_2^2)\,dxdy \\
+ & \frac{Dt^2}{24}\int_0^L\int_0^L [\,\nu(\kappa_{11}+\kappa_{22})^2 + (1-\nu)(\kappa_{11}^2 + 2\kappa_{12}^2 + \kappa_{22}^2)\,]\,dxdy \\
- & \int_0^L\int_0^L (f_1 u + f_2 v + f_3 w)\,dxdy,
\end{aligned} \tag{5.3}$$

where now $D = Et/(1-\nu^2)$. We underline that the different scaling (and ordering) in Eqs. (5.2) and (5.3) is just for clarity. This will not affect the finite element error analysis, since our error indicator will be scaling-invariant anyway.

Note in this context that the exceptional stretching-dominated deformation of the circular arch mentioned in Section 3 is mathematically equivalent to the membrane state of an infinite cylindrical shell under axially constant loading. This is a "soft" membrane state in the sense that it relies on the specific angular shape of the load. Another, "hard" type of membrane state arises when the kinematic constraints along the edge of the shell (or at joints) are firm enough to prevent inextensional deformations under any loading [13]. Such deformation states are rather common in engineering shell structures, so the membrane state of a shell is more a rule than an exception.

Consider now the finite element approximation of the above problem. One of the simplest approaches is again to choose the bilinear element where each component of the displacement field is approximated independently. In the bending-dominated case, numerical modifications are then again necessary, since otherwise both the shear and the membrane energy terms in Eq. (5.2) cause error amplification by factor $\sim L/t$. In the lowest-oder shell elements available in the engineering literature, such modifications are indeed made, but often quite implicitly. In MITC4, the modifications of the membrane strains arise from the use of the so called *faceting* technique while assembling the stiffness matrix. The idea of faceting is to replace the shell midsurface locally by its isoparametric bilinear approximation when evaluating the element stiffness matrix. Such a "facet trick" is obviously an attempt to extend the successful "beam trick" of arch modelling, so we expect that this can again be understood as a purely numerical modification of the energy within a geometrically conforming shell model. Indeed, recent analysis by Malinen confirms this [3, 4]. In the present simplified context, the faceting corresponds (approximately, see [3, 4]) to the modification of the membrane strains as

$$\beta_{11} \hookrightarrow \Pi_h^x \beta_{11}, \quad \beta_{22} \hookrightarrow \Pi_h^y \beta_{22}, \quad \beta_{12} \hookrightarrow \Pi_h^{xy} \beta_{12}, \tag{5.4}$$

where Π_h^x, Π_h^y and Π_h^{xy} are the mixed-interpolation and elementwise averaging operators as defined above in the plate-bending problem. In addition to these modifications, MITC4 maintains the mixed interpolation of the shear strains as in the plate-bending problem, i.e., the modification (4.8) is performed as well.

We will refer to the above interpretation of MITC4 in the shallow shell model as MITC4-S. — We note that slightly different interpretations of MITC4 are possible, especially when modifying the membrane strain β_{12} [3]. These interpretations are just small variations of each other when the approximation of uniformly smooth displacement fields is considered, as we do here. Instead when the boundary layers are taken into account, somewhat larger differences arise, and it seems that the assumed averaging of β_{12} is actually a somewhat *better* numerical modification than the one contained in MITC4, see [3].

Let us now proceed to the error analysis of the MITC4-S, assuming a rectangular mesh on the domain. We use again the error indicator (2.11) based on the modified energy norm. Consider first the case of bending-dominated deformation. As in the analogous plate-bending problem, the main difficulty in that case is to bound the approximation error, and there the main difficulty is further to bound the error arising from the asymptotic part of the solution when $t/L \to 0$. In the bending-dominated deformation state of a shell, the asymptotic solution is the inextensional solution which satisfies the constraints

$$\beta_{ij}(\mathbf{u}) = 0, \quad \rho_i(\mathbf{u}) = 0, \quad i,j = 1,2. \tag{5.5}$$

The question then is, whether smooth fields $\mathbf{u}$ satisfying constraints (5.5) can be aproximated accurately by continuous piecewise bilinear fields $\tilde{\mathbf{u}}$ satisfying the weakened constraints

$$\Pi_h^x \beta_{11}(\tilde{\mathbf{u}}) = \Pi_h^y \beta_{22}(\tilde{\mathbf{u}}) = \Pi_h^{xy} \beta_{12}(\tilde{\mathbf{u}}) = 0, \quad \Pi_h^x \rho_1(\tilde{\mathbf{u}}) = \Pi_h^y \rho_2(\tilde{\mathbf{u}}) = 0. \tag{5.6}$$

When $\mathbf{u} = (u, v, w, \boldsymbol{\theta})$ satisfies Eq. (5.5), we can bound the approximation error with any $\tilde{\mathbf{u}} = (\tilde{u}, \tilde{v}, \tilde{w}, \tilde{\boldsymbol{\theta}}) \in U_h$ satisfying Eq. (5.6) as

$$e_a(\mathbf{u}) \ \le \ \frac{|\!|\!|\mathbf{u} - \tilde{\mathbf{u}}|\!|\!|_h}{|\!|\!|\mathbf{u}|\!|\!|} \ = \ \frac{\|\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}\|_b}{\|\boldsymbol{\theta}\|_b}, \tag{5.7}$$

where $\|\boldsymbol{\theta}\|_b$ is the square root of the bending energy.

When system (5.6) is written in terms of the nodal degrees of freedom of $\tilde{\mathbf{u}}$, we may understand this system as a finite difference approximation to Eq. (5.5). We take this observation as the starting point, so that our approximation error analysis is actually based on *finite difference* analysis. Compared with ordinary finite difference error analysis, however, there is one essential difference: Stability plays no role when bounding the approximation error here. In fact, system (5.6) turns out to be a rather unstable finite difference approximation of Eq. (5.5). However, this is no problem insofar as the approximation error is concerned. The stability becomes more a problem in the membrane state when bounding the consistency error, see below.

So far we have been able to carry out the approximation error analysis under constraints (5.6) only in the case where the boundary conditions at $y = 0, L$ are periodic and the mesh is uniform in the y -direction. For simplicity we also assume that the boundaries at $x = 0, L$ are free. These assumptions allow sharp error analysis based on continuous and discrete Fourier expansions [5]. The analysis shows that when choosing $\tilde{\mathbf{u}}$ in the best possible way, the bound obtained from Eq. (5.7) is indeed of the uniformly optimal order $\mathcal{O}(h/L)$ under the stated conditions. In case of elliptic shell geometry, the error bound is obtained under the same regularity assumptions as in the plate-bending problem, i.e., assuming that the second partial derivatives of θ_1 and θ_2 are square integrable. In parabolic and hyperbolic shell geometries an extra degree of regularity is required as a rule: The *third* partial derivatives of θ_1 and θ_2 need to be square integrable, except for the case where the characteristic lines of the shell (along which the tangential curvature vanishes) are all parallel with mesh lines. The result (which is sharp by the analysis) means that if $\mathbf{u}$ varies in a given characteristic length scale λ that is much smaller than L, then the approximation error in the parabolic and hyperbolic cases is magnified by factor $\sim L/\lambda$ as compared with the elliptic case, except for the mentioned specific situations. Thus the shell geometry has a mild but notable effect on the finite element error bound in the bending-dominated case.

The above strong assumptions allow sharp error analysis also in the membrane-dominated case. For the membrane state of deformation we scale the strain energy according Eq. (5.3) (for convenience) and assume that the exact solution satisfies

$$\|\mathbf{u}\| \sim \|\mathbf{u}\|_m, \tag{5.8}$$

where $\|\mathbf{u}\|_m$ is the square root of the membrane energy (the first term in Eq. (5.3)). In this case the main error term is the consistency error. Indeed, when Eq. (5.8) holds, the approximation error is of no concern, since there are no large coefficients in the scaled deformation energy as expressed by (5.3). The most natural finite element scheme in this case would then be the standard (unmodified) scheme where there arises no consistency error.

Let us pause here for a moment. We have now confronted the main challenge in the finite element modelling of shells. The challenge is to find a low-oder (here bilinear) formulation that is able to capture two *different deformation states* with the *same formulation*, i.e., without problem specific tuneups. This challenge is not met when modelling beams, arches or plates, since the deformation in those cases may be assumed bending-dominated a priori (in most cases, at least). Efficient finite element designs for these problems also use the assumption vitally, by typically leaving the assumed dominant part of the energy untouched in the numerical modifications. Indeed, the bending energy was not modified numerically in any of the lowest-order finite element approaches considered above. Why the change of the dominant part is risky in general is basically because of the danger of losing, or weakening, *stability*. When the stability is lost, the consistency error becomes infinite, unless it happens that the generalized load functional (2.14) is zero for the exact solution $\mathbf{u}$. In the more typical case where the stability is only weakened, the consistency error remains finite but can be severely amplified. The only chance to avoid such error growth is then the possibility that the load functional (2.14) happens to be small enough for every $\mathbf{v} \in U_h$, so that the consistency error remains at a tolerable level even when amplified due to weak stability. Such a compensation is typically based on strong regularity hypotheses on the exact solution $\mathbf{u}$ and often on specific assumptions on the finite element mesh such as mesh uniformity. — Recall that it was this consistency error anomaly that was the main cause of the failure of the first bilinear plate element formulation discussed in the previous section.

In the design of a "shell element" that attempts to approximate all deformation states of the shell, the risk of weakening the stability in the membrane state must be taken. Indeed, we know that the membrane energy term must be modified numerically, since otherwise there would arise error growth by factor $\sim L/t$ in the bending-dominated case. In the membrane state we then — unavoidably — modify the leading term. The complete stability loss due to such a modification can always be avoided by supplementing the energy functional with the additional *stabilizing* term

$$\mathcal{F}_h(\mathbf{u}) = \delta\,[\,\mathcal{A}(\mathbf{u},\mathbf{u}) - \mathcal{A}_h(\mathbf{u},\mathbf{u})\,], \tag{5.9}$$

where $\delta > 0$ is a dimensionless parameter. This extra modification of the strain energy always saves stability, since one has then

$$|||\mathbf{u}|||_h^2 \;=\; \mathcal{A}_h(\mathbf{u},\mathbf{u}) + \mathcal{F}_h(\mathbf{u}) \geq\; \min\{1,\delta\}\;|||\mathbf{u}|||^2. \tag{5.10}$$

Note that adding the term (5.9) is just an example of classical *regularization*, often used when solving (nearly) ill-posed problems. In the present case the parameter δ needs to be rather small, since otherwise the added term would cause a big consistency error in the bending state. Indeed, in the bending state the term (5.9) is, effectively, multiplied by factor L^2/t^2 due to the different scaling of the energy in that case. Thus to avoid parametric amplification of the consistency error due to the added stabilizing term, we should choose δ in such a way that

$$0 \;\leq\; \delta \;\leq\; c\left(\frac{t}{L}\right)^2, \tag{5.11}$$

where c is a constant. Under this assumption, the extra consistency error caused by stabilization is harmless in both deformation states. Below we assume Eq. (5.11), reserving the choice $\delta = 0$ for schemes that require no stabilization.

Due to Eqs. (5.10) and (5.11), the problem of possibly weakened stability remains after adding the term (5.9). In the case where $\delta \sim t^2/L^2$, the weakened stability may cause the consistency error in the mambrane state to be amplified in the worst case by facor $\sim L/t$, i.e., by the same factor that one wanted to avoid in the bending state. (Such "shell-bending" elements have actually been proposed in the mathematical literature.) To avoid such a backlash, one apparently needs to design the modifications of the membrane strains extremely carefully. In general, the modifications should be strong enough thinking of the bending state, but simultaneously weak enough thinking of the membrane state. The MITC4-S formulation is obviously an attempt to balance between such conflicting requirements. We have already seen that the formulation is successful in the bending state (under the assumed conditions), so it remains to see if the modifications are also weak enough so that the consistency error in the membrane state can be controlled. The first good sign here is that the MITC4-S requires no stabilization. Namely, stability analysis shows that after the modification (5.4), Eq. (5.10) is valid with $\delta \sim t^2/L^2$, despite that no stabilization was assumed. This result is based essentially on the fact that the modifications (5.4) do not affect the derivative terms of β_{ij} except for β_{12}. The mentioned stability result can then be deduced from the corresponding plane-elastic result stating that the elementwise averaging of β_{12} does not affect the stability of the plane-elastic bilinear element, except when the element aspect ratio is high, see [8, Theorem 6.1].

The consistency error analysis of MITC4-S in the membrane state of deformation was carried out in [6], where the above stability result was also proven. To allow a sharp stability and error analysis, the above hypotheses on the semiperiodicity of the boundary conditions and on the semiuniformity of the mesh were made also here. It was assumed additionally that the boundary lines at $x = 0, L$ are clamped, i.e., $\mathbf{u} = \mathbf{0}$ at $x = 0, L$. Under these conditions it turns out that the weak stability of the scheme can be somewhat compensated by sharp bounds of the load functional

(2.14). Extra regularity assumptions (compared with those needed in standard non-parametric analysis) are also needed here for the exact solution $\mathbf{u}$. Under such assumptions, the consistency error can be estimated as [6]

$$e_c(\mathbf{u}_h) \ \sim \ C_1 \frac{h}{L} \ + \ C_2 \frac{h^{1+s}}{tL^s}. \tag{5.12}$$

Here C_1, C_2 are constants depending on $\mathbf{u}$, and s is a parameter related to the regularity of $\mathbf{u}$, so that $s = 0$ means standard and $s > 0$ extra regularity in the mentioned sense. Compared with the optimal convergence rate $\mathcal{O}(h/L)$, estimate (5.12) predicts error magnification by factor

$$K \sim 1 + \frac{L}{t}\left(\frac{h}{L}\right)^s. \tag{5.13}$$

When $s > 0$ (the larger s the better), the error is thus roughly dampened by the extra factor $\sim (h/L)^s$ from the worst case with $s = 0$. (When the solution is assumed smooth in a smaller length scale $\lambda < L$, one should replace L by λ in Eq. (5.13).)

Summarizing the results of the error analysis we conclude that the MITC4 formulation, as we have interpreted it, does improve the performance of the standard bilinear element considerably, at least under the rather favourable hypotheses made. In the bending-dominated case the improvement is major, in fact nearly optimal, under the assumed conditions. In the membrane state the MITC4 formulation causes at best mild, at worst severe, parametric error amplification compared with the standard bilinear scheme, the amplification being the less the higher the regularity of the exact solution.

To what extent the above results truly rely on the extremely specific assumptions made is an open problem. In particular, it is not known how well the MITC4-S element works when extended to quadrilateral element shapes. A possible extension would be to leave Π_h^{xy} in Eq. (5.4) as an elementwise averaging operator as it is, and define the modifications of the diagonal membrane strains via mixed interpolation in the same way as for the shear. Whether this is still a correct interpretation of MITC4 is not clear. The true performance of such an algorithm on general quadrilateral meshes is even less clear.

6 Concluding remarks

We have discussed the mathematical reasoning and error analysis of the lowest-order linear and bilinear elements for thin structures, starting from beams and closing at shells. As the mathematical analysis reveals, the various successful formulations found in the engineering literature have a lot in common. Typically these may be understood as being based on the usual energy principle with relatively simple numerical modifications imposed on the critical parametric terms of the energy expression. These modifications have often deep roots that lead back to the early history, or even prehistory, of the finite element methodology in structural analysis.

Following the roots we find a number of different physical justifications for the ultimately numerical modifications.

For both the finite element designer and the finite element theorist, the most challenging of the thin structures is the shell. Any "shell element" may be understood mathematically as an attempt to use the same numerical formulation to model many different deformation states of a shell, including membrane-dominated and bending-dominated deformations and various boundary layer types. No comparable challenge is met when modelling simpler thin structures such as beams, arches or plates, since the deformation type in these problems is typically given, or at least assumed as a starting point. In view of the difficulties already met in these simpler problems, lowest-oder shell elements have to be taken as very ambitious attempts in a rather complex mathematical environment. The MITC4 considered here is perhaps one the best possible formulations. Very likely rather similar constructions are hidden under various other shell element trade marks. In the finite element theory so far, only the first steps have been taken to understand such formulations and to possibly find their mathematical limits.

In a recent article [27], Lee and Bathe propose a new benchmark program for evaluating the performance of shell elements. They suggest benchmark testing that isolates the various asymptotic categories of shell deformations, instead of the current practice of repeating somewhat formal tests on a narrow selection of known problems where the exact solution is often an unclear mixture of various deformation components. We strongly agree with such a program, which actually is rather parallel with the error analysis philosophy that we have presented. We would like to add that it is perhaps the time also to open the various "shell elements" in the commercial codes for mutual comparisons and mathematical error analysis. If not the combined effort of the engineer and the mathematician can bring out the ultimate dream element for shells, it could achieve another important goal. It could raise the art of finite element modelling of shells from occultism to science.

References

[1] E.N. Dvorkin and K.J. Bathe, A continuum based four-node shell element for general nonlinear analysis, *Engineering Computations*, **1**, 77-88 (1984).

[2] K.J. Bathe and E.N. Dvorkin, A formulation of general shell elements – the use of mixed interpolation of tensorial components, *Int. J. Numer. Methods Engrg.*, **22**, 697-722 (1986).

[3] M. Malinen, On the classical shell model underlying bilinear degenerated shell finite elements, *Int. J. Numer. Methods Engrg.*, **52**, 389-416 (2001).

[4] M. Malinen, The classical shell model underlying the bilinear degenerated 3D FEM, Preprint TKK-Lo-31, Laboratory for Mechanics of Materials, Helsinki University of Technology (2000), to appear in *Int. J. Numer. Methods Engrg.*

[5] V. Havu and J. Pitkäranta, Analysis of a bilinear finite element for shallow shells I: Approximation of inextensional deformations, Preprint A430, Institue

of Mathematics, Helsinki University of Technology (2000), to appear in *Math. Comp.*

[6] V. Havu and J. Pitkäranta, Analysis of a bilinear finite element for shallow shells II: Consistency error, Preprint A433, Institue of Mathematics, Helsinki University of Technology (2001), to appear in *Math. Comp.*

[7] H.C. Martin, *Introduction to Matrix Methods of Structural Analysis*, McGraw-Hill, New York (1966).

[8] J. Pitkäranta, The first locking-free plane-elastic finite element: historia mathematica, *Comput. Mathods Appl. Mech. Engrg.* **190**, 1323-1366 (2000).

[9] R.H. MacNeal, A simple quadrilateral shell element, *Computers and Structures* **8**, 175-183 (1978).

[10] T.J.R. Hughes and T.E. Tezduyar, Finite elements based upon Mindlin plate theory with particular reference to the four-node bilinear isoparametric element, *J. Appl. Mech.* 587-596 (1981).

[11] K.J. Bathe and E.N. Dvorkin, A four-node plate bending element based on Mindlin/Reissner plate theory and mixed interpolation, *Int. J. Numer. Methods Engrg.* **21**, 367-383 (1985).

[12] K.J. Bathe and F. Brezzi, On the convergence of a four-node plate bending element based on Mindlin-Reissner plate bending theory and a mixed interpolation, in J.R. Whiteman (ed.), *Proc. of the Conference on Mathematics of Finite Elements and Applications V*, 491-503, Academic Press, New York (1985).

[13] 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.* **133**, 157-182 (1996).

[14] J. Pitkäranta, A.-M. Matache and C. Schwab, Fourier mode analysis of layers in shallow shell deformations, *Comput. Methods Appl. Mech. Engrg.* **190**, 2943-2975 (2001).

[15] M.J. Turner, R.W. Clough, H.C. Martin and L.J. Topp, Stiffness and deflection analysis of complex structures, *J. Aeronaut. Sci.* **23**, 805-823 (1956).

[16] D.N. Arnold, Discretization by finite elements of a model parameter dependent problem, *Numer.Math.* **37**, 405-421 (1981).

[17] F. Kikuchi, On the validity of the finite element analysis of circular arches represented by an assemblage of beam elements, *Comput. Methods Appl. Mech. Engrg.* **5**, 253-276 (1975).

[18] J. Pitkäranta, The problem of membrane locking in finite element analysis of cylindrical shells, *Numer. Math.* **61**, 523-542 (1992).

[19] D. Chapelle, A locking-free approximation of curved rods by straight beam elements, *Numer. Math.* **77**, 299-322 (1997).

[20] D. Malkus and T.J.R. Hughes, Mixed finite element methods – reduced and selective integration techniques: A unification of concepts, *Comput. Methods Appl. Mech. Engrg.* **15**, 63-81 (1978).

[21] T.J.R. Hughes, R.L. Taylor and W. Kanoknukulchai, A simple and efficient finite element for plate bending, *Int. J. Numer. Methods Engrg.* **11**, 1529-1543 (1977).

[22] C. Johnson and J. Pitkäranta, Analysis of some mixed finite element methods related to reduced integration, *Math. Comp.* **38**, 375-400 (1982).

[23] J. Pitkäranta and R. Stenberg, Error bounds for the approximation of the Stokes problem using bilinear/constant elements on irregular quadrilateral meshes, in J.R. Whiteman (ed.), *Proc. of the Conference on Mathematics of Finite Elements and Applications V*, 325-334, Academic Press, New york (1985).

[24] D.N. Arnold and R.S. Falk, Asymptotic analysis of the boundary layer for the Reissner-Mindlin plate model, *SIAM J. Math. Anal.* **27**, 486-514 (1996).

[25] J. Pitkäranta and M. Suri, Design principles and error analysis for reduced-shear plate-bending finite elements, *Numer. Math.* **75**, 223-266 (1996).

[26] J. Pitkäranta and M. Suri, Upper and lower error bounds for plate-bending finite elements, *Numer. Math.* **84**, 611-648 (2000).

[27] P.-S. Lee and K.J. Bathe, On the asymptotic behavior of shell structures and the evaluation in finite element solutions, *Computers and Structures* **80**, 235-255 (2002).

(continued from the back cover)

A456 Ville Havu , Harri Hakula , Tomi Tuominen
A benchmark study of elliptic and hyperbolic shells of revolution
January 2003

A455 Yaroslav V. Kurylev , Matti Lassas , Erkki Somersalo
Maxwell's Equations with Scalar Impedance: Direct and Inverse Problems
January 2003

A454 Timo Eirola , Marko Huhtanen , Jan von Pfaler
Solution methods for R-linear problems in C^n
October 2002

A453 Marko Huhtanen
Aspects of nonnormality for iterative methods
September 2002

A452 Kalle Mikkola
Infinite-Dimensional Linear Systems, Optimal Control and Algebraic Riccati Equations
October 2002

A451 Marko Huhtanen
Combining normality with the FFT techniques
September 2002

A450 Nikolai Yu. Bakaev
Resolvent estimates of elliptic differential and finite element operators in pairs of function spaces
August 2002

A449 Juhani Pitkäranta
Mathematical and historical reflections on the lowest order finite element models for thin structures
May 2002

A448 Teijo Arponen
Numerical solution and structural analysis of differential-algebraic equations
May 2002

HELSINKI UNIVERSITY OF TECHNOLOGY INSTITUTE OF MATHEMATICS
RESEARCH REPORTS

The list of reports is continued inside. Electronical versions of the reports are available at *http://www.math.hut.fi/reports/* .

A461 Tuomas Hytönen
Vector-valued wavelets and the Hardy space $H^1(R^n;X)$
April 2003

A460 Jan von Pfaler , Timo Eirola
Numerical Taylor expansions for invariant manifolds
April 2003

A459 Timo Salin
The quenching problem for the N-dimensional ball
April 2003

A458 Tuomas Hytönen
Translation-invariant Operators on Spaces of Vector-valued Functions
April 2003

A457 Timo Salin
On a Refined Asymptotic Analysis for the Quenching Problem
March 2003

ISBN 951-22-6001-8
ISSN 0784-3143
Espoo, 2002