Chapter 9
Mixed Finite Element Methods

Ferdinando Auricchio, Franco Brezzi and Carlo Lovadina
Università di Pavia and IMAT-C.N.R, Pavia, Italy

1 INTRODUCTION

Finite element method is a well-known and highly effective technique for the computation of approximate solutions of complex boundary value problems. Started in the fifties with milestone papers in a structural engineering context (see e.g. references in Chapter 1 of Zienkiewicz and Taylor (2000a) as well as classical references such as Turner *et al.* (1956) and Clough (1965)), the method has been extensively developed and studied in the last 50 years (Bathe, 1996; Brezzi and Fortin, 1991; Becker, Carey and Oden, 1981; Brenner and Scott, 1994; Crisfield, 1986; Hughes, 1987; Johnson, 1992; Ottosen and Petersson, 1992; Quarteroni and Valli, 1994; Reddy, 1993; Wait and Mitchell, 1985) and it is currently used also for the solution of complex nonlinear problems (Bathe, 1996; Bonet and Wood, 1997; Belytschko, Liu and Moran, 2000; Crisfield, 1991; Crisfield, 1997; Simo and Hughes, 1998; Simo, 1999; Zienkiewicz and Taylor, 2000b; Zienkiewicz and Taylor, 2000c).

Within such a broad approximation method, we focus on the often-called *mixed finite element methods*, where in our terminology the word ‘mixed’ indicates the fact that the problem discretization typically results in a linear algebraic system of the general form

$$\begin{bmatrix} \mathbf{A} & \mathbf{B}^{\mathrm{T}} \\ \mathbf{B} & \mathbf{0} \end{bmatrix} \begin{Bmatrix} \mathbf{x} \\ \mathbf{y} \end{Bmatrix} = \begin{Bmatrix} \mathbf{f} \\ \mathbf{g} \end{Bmatrix} \tag{1}$$

with **A** and **B** matrices and with **x**, **y**, **f**, and **g** vectors. Also, on mixed finite elements, the bibliography is quite large, ranging from classical contributions (Atluri, Gallagher and Zienkiewicz, 1983; Carey and Oden, 1983; Strang and Fix, 1973; Zienkiewicz *et al.*, 1983) to more recent references (Bathe, 1996; Belytschko, Liu and Moran, 2000; Bonet and Wood, 1997; Brezzi and Fortin, 1991; Hughes, 1987; Zienkiewicz and Taylor, 2000a; Zienkiewicz and Taylor, 2000c). An impressive amount of work has been devoted to a number of different stabilization techniques, virtually for all applications in which mixed formulations are involved. Their treatment is, however, beyond the scope of this chapter, and we will just say a few words on the general idea in Section 4.2.5.

In particular, the chapter is organized as follows. Section 2 sketches out the fact that several physical problem formulations share the same algebraic structure (1), once a discretization is introduced. Section 3 presents a simple, algebraic version of the abstract theory that rules most applications of mixed finite element methods. Section 4 gives several examples of efficient mixed finite element methods. Finally, in Section 5 we give some hints on how to perform a stability and error analysis, focusing on a representative problem (i.e. the Stokes equations).

Encyclopedia of Computational Mechanics, Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes. Volume 1: *Fundamentals*. © 2004 John Wiley & Sons, Ltd. ISBN: 0-470-84699-2.

2 FORMULATIONS

The goal of the present section is to point out that a quite large set of physical problem formulations shares the same algebraic structure (1), once a discretization is introduced.

To limit the discussion, we focus on *steady state* field problems defined in a domain $\Omega \subset \mathbb{R}^d$, with d the Euclidean space dimension. Moreover, we start from the simplest class of physical problems, that is, the one associated to diffusion mechanisms. Classical problems falling in this frame and frequently encountered in engineering are heat conduction, distribution of electrical or magnetic potentials, irrotational flow of ideal fluids, torsion or bending of cylindrical beams.

After addressing the thermal diffusion, as representative of the whole class, we move to more complex problems, such as the steady state flow of an incompressible Newtonian fluid and the mechanics of elastic bodies. For each problem, we briefly describe the local *differential equations* and possible *variational formulations*.

Before proceeding, we need to comment on the adopted notation. In general, we indicate scalar fields with nonbold lower-case roman or nonbold lower-case greek letters (such as a, α, b, β), vector fields with bold lower-case roman letters (such as $\mathbf{a}$, $\mathbf{b}$), second-order tensors with bold lower-case greek letters or bold upper-case roman letters (such as $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\mathbf{A}$, $\mathbf{B}$), fourth-order tensors with upper-case blackboard roman letters (such as $\mathbb{D}$). We however reserve the letters $\mathbb{A}$ and $\mathbb{B}$ for 'composite' matrices (see e.g. equation (31)). Moreover, we indicate with $\mathbf{0}$ the null vector, with $\mathbf{I}$ the identity second-order tensor and with $\mathbb{I}$ the identity fourth-order tensor.

Whenever necessary or useful, we may use the standard indicial notation to represent vectors or tensors. Accordingly, in a Euclidean space with base vectors $\mathbf{e}_i$, a vector $\mathbf{a}$, a second-order tensor $\boldsymbol{\alpha}$, and a fourth-order tensor $\mathbb{D}$ have the following components

$$\begin{aligned} \mathbf{a}|_i = a_i = \mathbf{a}\cdot\mathbf{e}_i, \quad \boldsymbol{\alpha}|_{ij} = \alpha_{ij} = \mathbf{e}_i\cdot(\boldsymbol{\alpha}\mathbf{e}_j) \\ \mathbb{D}|_{ijkl} = \mathbb{D}_{ijkl} = (\mathbf{e}_i\otimes\mathbf{e}_j):[\mathbb{D}(\mathbf{e}_k\otimes\mathbf{e}_l)] \end{aligned} \qquad (2)$$

where $\cdot$, $\otimes$, and $:$ indicate respectively the scalar vector product, the (second-order) tensorial vector product, and the scalar (second-order) tensor product. Sometimes, the scalar vector product will be also indicated as $\mathbf{a}^{\mathrm{T}}\mathbf{b}$, where the superscript T indicates transposition.

During the discussion, we also introduce standard differential operators such as gradient and divergence, indicated respectively as '∇' and 'div', and acting either on scalar, vector, or tensor fields. In particular, we have

$$\begin{aligned} \nabla a|_i = a_{,i}, \quad \nabla\mathbf{a}|_{ij} = a_{i,j} \\ \operatorname{div}\mathbf{a} = a_{i,i}, \quad \operatorname{div}\boldsymbol{\alpha}|_i = \alpha_{ij,j} \end{aligned} \qquad (3)$$

where repeated subscript indices imply summation and where the subscript comma indicates derivation, that is, $a_{,i} = \partial a/\partial x_i$.

Finally, given for example, a scalar field a, a vector field $\mathbf{a}$, and a tensor field $\boldsymbol{\alpha}$, we indicate with δa, $\delta\mathbf{a}$, $\delta\boldsymbol{\alpha}$ the corresponding *variation fields* and with a^h, $\mathbf{a}^h$, $\boldsymbol{\alpha}^h$ the corresponding interpolations, expressed in general as

$$a^h = N_k^a \hat{a}_k, \quad \mathbf{a}^h = \mathbf{N}_k^{\mathbf{a}} \hat{a}_k, \quad \boldsymbol{\alpha}^h = \mathbf{N}_k^{\boldsymbol{\alpha}} \hat{\alpha}_k \qquad (4)$$

where N_k^a, $\mathbf{N}_k^{\mathbf{a}}$, and $\mathbf{N}_k^{\boldsymbol{\alpha}}$ are a set of interpolation functions (i.e. the so-called *shape functions*), while $\hat{a}_k$ and $\hat{\alpha}_k$ are a set of interpolation parameters (i.e. the so-called *degrees of freedom*); clearly, N_k^a, $\mathbf{N}_k^{\mathbf{a}}$, and $\mathbf{N}_k^{\boldsymbol{\alpha}}$ are respectively scalar, vector, and tensor predefined (assigned) fields, while $\hat{a}_k$ and $\hat{\alpha}_k$ are scalar quantities, representing the *effective unknowns* of the approximated problems. With the adopted notation, it is now simple to evaluate the differential operators (3) on the interpolated fields, that is,

$$\begin{aligned} \nabla a^h = \left(\nabla N_k^a\right)\hat{a}_k, \quad \nabla\mathbf{a}^h = \left(\nabla\mathbf{N}_k^{\mathbf{a}}\right)\hat{a}_k \\ \operatorname{div}\mathbf{a}^h = \left(\operatorname{div}\mathbf{N}_k^{\mathbf{a}}\right)\hat{a}_k, \quad \operatorname{div}\boldsymbol{\alpha}^h = \left(\operatorname{div}\mathbf{N}_k^{\boldsymbol{\alpha}}\right)\hat{\alpha}_k \end{aligned} \qquad (5)$$

or in indicial notation

$$\begin{aligned} \nabla a^h|_i = N_{k,i}^a \hat{a}_k, \quad \nabla\mathbf{a}^h|_{ij} = \mathbf{N}_k^{\mathbf{a}}|_{i,j}\hat{a}_k \\ \operatorname{div}\mathbf{a}^h = \mathbf{N}_k^{\mathbf{a}}|_{i,i}\hat{a}_k, \quad \operatorname{div}\boldsymbol{\alpha}^h|_i = \mathbf{N}_k^{\boldsymbol{\alpha}}|_{ij,j}\hat{\alpha}_k \end{aligned} \qquad (6)$$

2.1 Thermal diffusion

The physical problem
Indicating with θ the body temperature, $\mathbf{e}$ the temperature gradient, $\mathbf{q}$ the heat flux, and with b the assigned heat source per unit volume, a steady state thermal problem in a domain Ω can be formulated as a (θ, $\mathbf{e}$, $\mathbf{q}$) *three field problem* as follows:

$$\begin{cases} \operatorname{div}\mathbf{q} + b = 0 & \text{in } \Omega \\ \mathbf{q} = -\mathbf{D}\mathbf{e} & \text{in } \Omega \\ \mathbf{e} = \nabla\theta & \text{in } \Omega \end{cases} \qquad (7)$$

which are respectively the balance equation, the constitutive equation, the compatibility equation.

In particular, we assume a linear constitutive equation (known as Fourier law), where $\mathbf{D}$ is the conductivity material-dependent second-order tensor; in the simple case of thermally isotropic material, $\mathbf{D} = k\mathbf{I}$ with k the isotropic thermal conductivity.

Equation (7) is completed by proper boundary conditions. For simplicity, we consider only the case of trivial essential conditions on the whole domain boundary, that is,

$$\theta = 0 \quad \text{on} \quad \partial\Omega \tag{8}$$

This position is clearly very restrictive from a physical point of view but it is still adopted since it simplifies the forthcoming discussion, at the same time without limiting our numerical considerations.

As classically done, the three field problem (7) can be simplified eliminating the temperature gradient **e**, obtaining a $(\theta, \mathbf{q})$ *two field problem*

$$\begin{cases} \operatorname{div} \mathbf{q} + b = 0 & \text{in} \quad \Omega \\ \mathbf{q} = -\mathbf{D}\nabla\theta & \text{in} \quad \Omega \end{cases} \tag{9}$$

and the two field problem (9) can be further simplified eliminating the thermal flux **q** (or eliminating the fields **e** and **q** directly from equation (7)), obtaining a θ *single field problem*

$$-\operatorname{div}(\mathbf{D}\nabla\theta) + b = 0 \quad \text{in} \quad \Omega \tag{10}$$

For the case of an isotropic and homogeneous body, this last equation specializes as follows

$$-k\Delta\theta + b = 0 \quad \text{in} \quad \Omega \tag{11}$$

where Δ is the standard Laplace operator.

Variational principles

The single field equation (10) can be easily derived starting from the *potential energy* functional

$$\Pi(\theta) = \frac{1}{2}\int_\Omega [\nabla\theta \cdot \mathbf{D}\nabla\theta]\, \mathrm{d}\Omega + \int_\Omega \theta b \, \mathrm{d}\Omega \tag{12}$$

Requiring the stationarity of potential (12), we obtain

$$\mathrm{d}\Pi(\theta)[\delta\theta] = \int_\Omega [(\nabla\delta\theta) \cdot \mathbf{D}\nabla\theta]\, \mathrm{d}\Omega + \int_\Omega [\delta\theta b]\, \mathrm{d}\Omega = 0 \tag{13}$$

where $\delta\theta$ indicates a possible variation of the temperature field θ and $\mathrm{d}\Pi(\theta)[\delta\theta]$ indicates the potential variation evaluated at θ in the direction $\delta\theta$. Since functional (12) is convex, we may note that the stationarity requirement is equivalent to a minimization.

Recalling equation (4), we may now introduce an interpolation for the temperature field in the form

$$\theta \approx \theta^h = N_k^\theta \hat{\theta}_k \tag{14}$$

as well as a similar approximation for the corresponding variation field, such that equation (13) can be rewritten in matricial form as follows

$$\mathbf{A}\hat{\boldsymbol{\theta}} = \mathbf{f} \tag{15}$$

with

$$\begin{cases} \mathbf{A}|_{ij} = \displaystyle\int_\Omega [\nabla N_i^\theta \cdot \mathbf{D}\nabla N_j^\theta]\, \mathrm{d}\Omega, \quad \hat{\boldsymbol{\theta}}|_j = \hat{\theta}_j \\ \mathbf{f}|_i = -\displaystyle\int_\Omega \left[N_i^\theta b\right] \mathrm{d}\Omega \end{cases} \tag{16}$$

Besides the integral form (13) associated to the single field equation (10), it is also possible to associate an integral form to the two field equation (9) starting now from the more general *Hellinger–Reissner functional*

$$\Pi^{\mathrm{HR}}(\theta, \mathbf{q}) = -\frac{1}{2}\int_\Omega \left[\mathbf{q} \cdot \mathbf{D}^{-1}\mathbf{q}\right] \mathrm{d}\Omega - \int_\Omega \left[\mathbf{q} \cdot \nabla\theta\right] \mathrm{d}\Omega + \int_\Omega \theta b \, \mathrm{d}\Omega \tag{17}$$

Requiring the stationarity of functional (17), we obtain

$$\begin{cases} \mathrm{d}\Pi^{\mathrm{HR}}(\theta, \mathbf{q})[\delta\mathbf{q}] = -\displaystyle\int_\Omega \left[\delta\mathbf{q} \cdot \mathbf{D}^{-1}\mathbf{q}\right] \mathrm{d}\Omega \\ \qquad\qquad - \displaystyle\int_\Omega \left[\delta\mathbf{q} \cdot \nabla\theta\right] \mathrm{d}\Omega = 0 \\ \mathrm{d}\Pi^{\mathrm{HR}}(\theta, \mathbf{q})[\delta\theta] = -\displaystyle\int_\Omega \left[(\nabla\delta\theta) \cdot \mathbf{q}\right] \mathrm{d}\Omega \\ \qquad\qquad + \displaystyle\int_\Omega [\delta\theta b]\, \mathrm{d}\Omega = 0 \end{cases} \tag{18}$$

which is now equivalent to the search of a saddle point. Changing sign to both equations and introducing the approximation

$$\begin{cases} \theta \approx \theta^h = N_k^\theta \hat{\theta}_k \\ \mathbf{q} \approx \mathbf{q}^h = \mathbf{N}_k^{\mathbf{q}} \hat{q}_k \end{cases} \tag{19}$$

as well as a similar approximation for the corresponding variation fields, equation (18) can be rewritten in matricial form as follows

$$\begin{bmatrix} \mathbf{A} & \mathbf{B}^{\mathrm{T}} \\ \mathbf{B} & \mathbf{0} \end{bmatrix} \begin{Bmatrix} \hat{\mathbf{q}} \\ \hat{\boldsymbol{\theta}} \end{Bmatrix} = \begin{Bmatrix} \mathbf{0} \\ \mathbf{g} \end{Bmatrix} \tag{20}$$

where

$$\begin{cases} \mathbf{A}|_{ij} = \displaystyle\int_\Omega [\mathbf{N}_i^{\mathbf{q}} \cdot \mathbf{D}^{-1}\mathbf{N}_j^{\mathbf{q}}]\, \mathrm{d}\Omega, \quad \hat{\mathbf{q}}|_j = \hat{q}_j \\ \mathbf{B}|_{rj} = \displaystyle\int_\Omega [\nabla N_r^\theta \cdot \mathbf{N}_j^{\mathbf{q}}]\, \mathrm{d}\Omega, \quad \hat{\boldsymbol{\theta}}|_r = \hat{\theta}_r \\ \mathbf{g}|_r = \displaystyle\int_\Omega [N_r^\theta b]\, \mathrm{d}\Omega \end{cases} \tag{21}$$

Starting from the Hellinger–Reissner functional (17) previously addressed, the following *modified Hellinger–Reissner* functional can be also generated

$$\Pi^{\mathrm{HR,m}}(\theta, \mathbf{q}) = -\frac{1}{2}\int_\Omega \left[\mathbf{q}\cdot\mathbf{D}^{-1}\mathbf{q}\right]\,\mathrm{d}\Omega + \int_\Omega \left[\theta\,\mathrm{div}\,\mathbf{q}\right]\,\mathrm{d}\Omega + \int_\Omega \theta b\,\mathrm{d}\Omega \tag{22}$$

and, requiring its stationarity, we obtain

$$\begin{cases} \mathrm{d}\Pi^{\mathrm{HR,m}}(\theta, \mathbf{q})[\delta\mathbf{q}] = -\displaystyle\int_\Omega \left[\delta\mathbf{q}\cdot\mathbf{D}^{-1}\mathbf{q}\right]\,\mathrm{d}\Omega \\ \qquad\qquad + \displaystyle\int_\Omega \left[\mathrm{div}\,(\delta\mathbf{q})\ \theta\right]\,\mathrm{d}\Omega = 0 \\ \mathrm{d}\Pi^{\mathrm{HR,m}}(\theta, \mathbf{q})[\delta\theta] = \displaystyle\int_\Omega \left[\delta\theta\,\mathrm{div}\,\mathbf{q}\right]\,\mathrm{d}\Omega + \int_\Omega [\delta\theta b]\,\mathrm{d}\Omega = 0 \end{cases} \tag{23}$$

which is again equivalent to the search of a saddle point. Changing sign to both equations and introducing again field approximation (19), equation (23) can be rewritten in matricial form as equation (20), with the difference that now

$$\mathbf{B}|_{rj} = -\int_\Omega \left[N_r^\theta\ \mathrm{div}\left(\mathbf{N}_j^{\mathbf{q}}\right)\right]\,\mathrm{d}\Omega \tag{24}$$

Similarly, we may also associate an integral form to the three field equation (7) starting from the even more general *Hu–Washizu functional*

$$\Pi^{\mathrm{HW}}(\theta, \mathbf{e}, \mathbf{q}) = \frac{1}{2}\int_\Omega [\mathbf{e}\cdot\mathbf{De}]\,\mathrm{d}\Omega + \int_\Omega \left[\mathbf{q}\cdot(\mathbf{e}-\nabla\theta)\right]\,\mathrm{d}\Omega + \int_\Omega \theta b\,\mathrm{d}\Omega \tag{25}$$

Requiring the stationarity of functional (25), we obtain

$$\begin{cases} \mathrm{d}\Pi^{\mathrm{HW}}(\theta, \mathbf{e}, \mathbf{q})[\delta\mathbf{e}] = \displaystyle\int_\Omega [\delta\mathbf{e}\cdot\mathbf{De}]\,\mathrm{d}\Omega \\ \qquad\qquad + \displaystyle\int_\Omega \left[\delta\mathbf{e}\cdot\mathbf{q}\right]\,\mathrm{d}\Omega = 0 \\ \mathrm{d}\Pi^{\mathrm{HW}}(\theta, \mathbf{e}, \mathbf{q})[\delta\mathbf{q}] = \displaystyle\int_\Omega \left[\delta\mathbf{q}\cdot(\mathbf{e}-\nabla\theta)\right]\,\mathrm{d}\Omega = 0 \\ \mathrm{d}\Pi^{\mathrm{HW}}(\theta, \mathbf{e}, \mathbf{q})[\delta\theta] = -\displaystyle\int_\Omega \left[(\nabla\delta\theta)\cdot\mathbf{q}\right]\,\mathrm{d}\Omega \\ \qquad\qquad + \displaystyle\int_\Omega [\delta\theta b]\,\mathrm{d}\Omega = 0 \end{cases} \tag{26}$$

which is equivalent to searching a saddle point. Introducing the following approximation

$$\begin{cases} \theta \approx \theta^h = N_k^\theta \hat{\theta}_k \\ \mathbf{e} \approx \mathbf{e}^h = \mathbf{N}_k^{\mathbf{e}} \hat{e}_k \\ \mathbf{q} \approx \mathbf{q}^h = \mathbf{N}_k^{\mathbf{q}} \hat{q}_k \end{cases} \tag{27}$$

as well as a similar approximation for the corresponding variation fields, equation (26) can be rewritten in matricial form as follows:

$$\begin{bmatrix} \mathbf{A} & \mathbf{B}^{\mathrm{T}} & \mathbf{0} \\ \mathbf{B} & \mathbf{0} & \mathbf{C}^{\mathrm{T}} \\ \mathbf{0} & \mathbf{C} & \mathbf{0} \end{bmatrix} \begin{Bmatrix} \hat{\mathbf{e}} \\ \hat{\mathbf{q}} \\ \hat{\boldsymbol{\theta}} \end{Bmatrix} = \begin{Bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{h} \end{Bmatrix} \tag{28}$$

where

$$\begin{cases} \mathbf{A}|_{ij} = \displaystyle\int_\Omega [\mathbf{N}_i^{\mathbf{e}}\cdot\mathbf{DN}_j^{\mathbf{e}}]\,\mathrm{d}\Omega, \quad \hat{\mathbf{e}}|_j = \hat{e}_j \\ \mathbf{B}|_{rj} = \displaystyle\int_\Omega [\nabla\mathbf{N}_r^{\mathbf{q}}\cdot\mathbf{N}_j^{\mathbf{e}}]\,\mathrm{d}\Omega, \quad \hat{\mathbf{q}}|_r = \hat{q}_r \\ \mathbf{C}|_{sr} = -\displaystyle\int_\Omega [\nabla N_s^\theta\cdot\mathbf{N}_r^{\mathbf{q}}]\,\mathrm{d}\Omega, \quad \hat{\boldsymbol{\theta}}|_s = \hat{\theta}_s \\ \mathbf{h}|_s = -\displaystyle\int_\Omega \left[N_s^\theta b\right]\,\mathrm{d}\Omega \end{cases} \tag{29}$$

For later considerations, we note that equation (28) can be also rewritten as

$$\begin{bmatrix} \mathbb{A} & \mathbb{B}^{\mathrm{T}} \\ \mathbb{B} & \mathbf{0} \end{bmatrix} \begin{Bmatrix} \mathbf{x} \\ \mathbf{y} \end{Bmatrix} = \begin{Bmatrix} \mathbf{0} \\ \mathbf{h} \end{Bmatrix} \tag{30}$$

where we made the following simple identifications

$$\mathbb{A} = \begin{bmatrix} \mathbf{A} & \mathbf{B}^{\mathrm{T}} \\ \mathbf{B} & \mathbf{0} \end{bmatrix}, \quad \mathbb{B} = \{\mathbf{0}, \mathbf{C}\}$$

$$\mathbf{x} = \begin{Bmatrix} \hat{\mathbf{e}} \\ \hat{\mathbf{q}} \end{Bmatrix}, \quad \mathbf{y} = \hat{\boldsymbol{\theta}} \tag{31}$$

Examples of specific choices for the interpolating functions (14), (19), or (27) respectively within the single field, two field, and three field formulations can be found in standard textbooks (Bathe, 1996; Ottosen and Petersson, 1992; Brezzi and Fortin, 1991; Hughes, 1987; Zienkiewicz and Taylor, 2000a) or in the literature.

2.2 Stokes equations

The physical problem
Indicating with $\mathbf{u}$ the fluid velocity, $\boldsymbol{\varepsilon}$ the symmetric part of the velocity gradient, $\boldsymbol{\sigma}$ the stress, p a pressure-like quantity, and with $\mathbf{b}$ the assigned body load per unit volume,

the steady state flow of an incompressible Newtonian fluid can be formulated as a $(\mathbf{u}, \boldsymbol{\varepsilon}, \boldsymbol{\sigma}, p)$ *four field problem* as follows:

$$\begin{cases} \operatorname{div}\boldsymbol{\sigma} + \mathbf{b} = \mathbf{0} & \text{in } \Omega \\ \boldsymbol{\sigma} = 2\mu\boldsymbol{\varepsilon} - p\mathbf{1} & \text{in } \Omega \\ \boldsymbol{\varepsilon} = \nabla^s\mathbf{u} & \text{in } \Omega \\ \operatorname{div}\mathbf{u} = 0 & \text{in } \Omega \end{cases} \tag{32}$$

which are respectively the balance, the constitutive, the compatibility, and the incompressibility constraint equations. In particular, ∇^s indicates the symmetric part of the gradient, that is, in a more explicit form,

$$\boldsymbol{\varepsilon} = \nabla^s\mathbf{u} = \tfrac{1}{2}\left[\nabla\mathbf{u} + (\nabla\mathbf{u})^{\mathrm{T}}\right] \tag{33}$$

while the constitutive equation relates the stress $\boldsymbol{\sigma}$ to the symmetric part of the velocity gradient $\boldsymbol{\varepsilon}$ through a material constant μ known as viscosity, and a volumetric pressure-like scalar contribution p.

This set of equations is completed by proper boundary conditions. As for the thermal problem, we prescribe trivial essential conditions on the whole domain boundary, that is,

$$\mathbf{u} = \mathbf{0} \quad \text{on} \quad \partial\Omega \tag{34}$$

As classically done, equation (32) can be simplified eliminating $\boldsymbol{\varepsilon}$ and $\boldsymbol{\sigma}$, obtaining a $(\mathbf{u},\ p)$ *two field problem*

$$\begin{cases} \mu\Delta\mathbf{u} - \nabla p + \mathbf{b} = \mathbf{0} & \text{in } \Omega \\ \operatorname{div}\mathbf{u} = 0 & \text{in } \Omega \end{cases} \tag{35}$$

Variational principles

Equation (35) can be derived starting from the *potential energy* functional

$$\Pi(\mathbf{u}) = \frac{1}{2}\mu\int_\Omega [\nabla\mathbf{u} : \nabla\mathbf{u}]\ \mathrm{d}\Omega - \int_\Omega [\mathbf{b}\cdot\mathbf{u}]\ \mathrm{d}\Omega \tag{36}$$

where now $\mathbf{u}$ is a function satisfying the constraint, that is, such that $\operatorname{div}\mathbf{u} = 0$.

To remove the constraint on $\mathbf{u}$, we can modify the variational principle introducing the functional

$$L(\mathbf{u},\ p) = \frac{1}{2}\mu\int_\Omega [\nabla\mathbf{u} : \nabla\mathbf{u}]\ \mathrm{d}\Omega - \int_\Omega [\mathbf{b}\cdot\mathbf{u}]\ \mathrm{d}\Omega \\ - \int_\Omega \left[p \operatorname{div}\mathbf{u}\right]\ \mathrm{d}\Omega \tag{37}$$

where p now plays the role of *Lagrange multiplier*.

Requiring the stationarity of functional (37), we obtain

$$\begin{cases} \mathrm{d}L(\mathbf{u},\ p)[\delta\mathbf{u}] = \mu\displaystyle\int_\Omega [(\nabla\delta\mathbf{u}) : \nabla\mathbf{u}]\ \mathrm{d}\Omega - \int_\Omega [\delta\mathbf{u}\cdot\mathbf{b}]\ \mathrm{d}\Omega \\ \qquad\qquad - \displaystyle\int_\Omega [\operatorname{div}(\delta\mathbf{u})\ \ \pi]\,\mathrm{d}\Omega = 0 \\ \mathrm{d}L(\mathbf{u},\ p)[\delta p] = -\displaystyle\int_\Omega \left[\delta p \operatorname{div}\mathbf{u}\right]\ \mathrm{d}\Omega = 0 \end{cases} \tag{38}$$

which is equivalent to the search of a saddle point. Introducing the following approximation

$$\begin{cases} \mathbf{u} \approx \mathbf{u}^h = \mathbf{N}_k^{\mathbf{u}}\hat{u}_k \\ p \approx p^h = N_k^p\hat{p}_k \end{cases} \tag{39}$$

as well as a similar approximation for the corresponding variation fields, equation (38) can be rewritten as follows

$$\begin{bmatrix} \mathbf{A} & \mathbf{B}^{\mathrm{T}} \\ \mathbf{B} & \mathbf{0} \end{bmatrix} \begin{Bmatrix} \hat{\mathbf{u}} \\ \hat{\mathbf{p}} \end{Bmatrix} = \begin{Bmatrix} \mathbf{f} \\ \mathbf{0} \end{Bmatrix} \tag{40}$$

where

$$\begin{cases} \mathbf{A}|_{ij} = \mu\displaystyle\int_\Omega \left[\nabla\mathbf{N}_i^{\mathbf{u}} : \nabla\mathbf{N}_j^{\mathbf{u}}\right]\ \mathrm{d}\Omega, \quad \hat{\mathbf{u}}|_j = \hat{u}_j \\ \mathbf{B}|_{rj} = -\displaystyle\int_\Omega \left[N_r^p \operatorname{div}\left(\mathbf{N}_j^{\mathbf{u}}\right)\right]\ \mathrm{d}\Omega, \quad \hat{\mathbf{p}}|_r = \hat{p}_r \\ \mathbf{f}|_i = \displaystyle\int_\Omega \left[\mathbf{N}_i^{\mathbf{u}}\cdot\mathbf{b}\right]\ \mathrm{d}\Omega \end{cases} \tag{41}$$

Examples of specific choices for the interpolating functions (39) can be found in standard textbooks (Bathe, 1996; Brezzi and Fortin, 1991; Hughes, 1987; Quarteroni and Valli, 1994; Zienkiewicz and Taylor, 2000a) or in the literature.

2.3 Elasticity

The physical problem

Indicating with $\mathbf{u}$ the body displacement, $\boldsymbol{\varepsilon}$ the strain, $\boldsymbol{\sigma}$ the stress, and with $\mathbf{b}$ the assigned body load per unit volume, the steady state equations for a deformable solid under the assumption of small displacement gradients can be formulated as a $(\mathbf{u}, \boldsymbol{\varepsilon}, \boldsymbol{\sigma})$ *three field problem* as follows

$$\begin{cases} \operatorname{div}\boldsymbol{\sigma} + \mathbf{b} = \mathbf{0} & \text{in } \Omega \\ \boldsymbol{\sigma} = \mathbb{D}\boldsymbol{\varepsilon} & \text{in } \Omega \\ \boldsymbol{\varepsilon} = \nabla^s\mathbf{u} & \text{in } \Omega \end{cases} \tag{42}$$

which are respectively the balance, the constitutive, and the compatibility equations.

In particular, we assume a linear constitutive equation, where $\mathbb{D}$ is the elastic material-dependent fourth-order tensor; in the simple case of a mechanically isotropic material, $\mathbb{D}$ specializes as

$$\mathbb{D} = 2\mu\mathbb{I} + \lambda \mathbf{I} \otimes \mathbf{I} \tag{43}$$

and the constitutive equation can be rewritten as

$$\boldsymbol{\sigma} = 2\mu\boldsymbol{\varepsilon} + \lambda \ \mathrm{tr}\,(\boldsymbol{\varepsilon}) \ \mathbf{I} \tag{44}$$

where $\mathrm{tr}\,(\boldsymbol{\varepsilon}) = \mathbf{I} : \boldsymbol{\varepsilon}$. This set of equations is completed by proper boundary conditions. As previously done, we prescribe trivial essential conditions on the whole domain boundary, that is,

$$\mathbf{u} = \mathbf{0} \quad \text{on} \quad \partial\Omega \tag{45}$$

This position is once more very restrictive from a physical point of view but it is still adopted since it simplifies the forthcoming discussion, at the same time without limiting our numerical considerations.

The three field problem (42) can be simplified eliminating the strain $\boldsymbol{\varepsilon}$, obtaining a $(\mathbf{u}, \boldsymbol{\sigma})$ *two field problem*

$$\begin{cases} \mathrm{div}\,\boldsymbol{\sigma} + \mathbf{b} = \mathbf{0} & \text{in } \Omega \\ \boldsymbol{\sigma} = \mathbb{D}\nabla^s\mathbf{u} & \text{in } \Omega \end{cases} \tag{46}$$

and the *two field problem* (46) can be simplified eliminating the stress $\boldsymbol{\sigma}$ (or eliminating $\boldsymbol{\varepsilon}$ and $\boldsymbol{\sigma}$ directly from equation (42)), obtaining a $\mathbf{u}$ *single field problem*

$$\mathrm{div}\,\left(\mathbb{D}\nabla^s\mathbf{u}\right) + \mathbf{b} = \mathbf{0} \quad \text{in} \quad \Omega \tag{47}$$

In the case of an isotropic and homogeneous body, this last equation specializes as follows:

$$2\mu \ \mathrm{div}\left(\nabla^s\mathbf{u}\right) + \lambda\nabla\,(\,\mathrm{div}\,\mathbf{u}) + \mathbf{b} = \mathbf{0} \quad \text{in} \quad \Omega \tag{48}$$

Variational principles
The single field equation (47) can be easily derived starting from the potential energy functional

$$\Pi(\mathbf{u}) = \frac{1}{2}\int_\Omega \left[\nabla^s\mathbf{u} : \mathbb{D}\nabla^s\mathbf{u}\right] \mathrm{d}\Omega - \int_\Omega [\mathbf{b}\cdot\mathbf{u}] \ \mathrm{d}\Omega \tag{49}$$

Requiring the stationarity of potential (49), we obtain

$$\begin{aligned} \mathrm{d}\Pi(\mathbf{u})[\delta\mathbf{u}] = &\int_\Omega \left[\left(\nabla^s\delta\mathbf{u}\right) : \mathbb{D}\nabla^s\mathbf{u}\right] \mathrm{d}\Omega \\ &- \int_\Omega [\delta\mathbf{u}\cdot\mathbf{b}] \ \mathrm{d}\Omega = 0 \end{aligned} \tag{50}$$

where $\delta\mathbf{u}$ indicates a possible variation of the displacement field $\mathbf{u}$. Since functional (49) is convex, we may note that the stationarity requirement is equivalent to a minimization. Recalling the notation introduced in equation (4), we may now introduce an interpolation for the displacement field in the form

$$\mathbf{u} \approx \mathbf{u}^h = \mathbf{N}^{\mathbf{u}}_k \hat{u}_k \tag{51}$$

as well as a similar approximation for the variation field, such that equation (50) can be rewritten as follows:

$$\mathbf{A}\hat{\mathbf{u}} = \mathbf{f} \tag{52}$$

where

$$\begin{cases} \mathbf{A}|_{ij} = \displaystyle\int_\Omega \left[\nabla^s\mathbf{N}^{\mathbf{u}}_i : \mathbb{D}\nabla^s\mathbf{N}^{\mathbf{u}}_j\right] \mathrm{d}\Omega, \quad \hat{\mathbf{u}}|_j = \hat{u}_j \\ \mathbf{f}|_i = \displaystyle\int_\Omega \left[\mathbf{N}^{\mathbf{u}}_i \cdot \mathbf{b}\right] \mathrm{d}\Omega \end{cases} \tag{53}$$

Besides the integral form (50) associated to the single field equation (47), it is also possible to associate an integral form to the two field equation (46) starting now from the more general Hellinger–Reissner functional

$$\begin{aligned} \Pi^{\mathrm{HR}}(\mathbf{u},\boldsymbol{\sigma}) = &-\frac{1}{2}\int_\Omega \left[\boldsymbol{\sigma} : \mathbb{D}^{-1}\boldsymbol{\sigma}\right] \mathrm{d}\Omega + \int_\Omega [\boldsymbol{\sigma} : \nabla\mathbf{u}] \ \mathrm{d}\Omega \\ &- \int_\Omega [\mathbf{b}\cdot\mathbf{u}] \ \mathrm{d}\Omega \end{aligned} \tag{54}$$

Requiring the stationarity of functional (54), we obtain

$$\begin{cases} \mathrm{d}\Pi^{\mathrm{HR}}(\mathbf{u},\boldsymbol{\sigma})[\delta\boldsymbol{\sigma}] = -\displaystyle\int_\Omega \left[\delta\boldsymbol{\sigma} : \mathbb{D}^{-1}\boldsymbol{\sigma}\right] \mathrm{d}\Omega \\ \qquad\qquad + \displaystyle\int_\Omega [\delta\boldsymbol{\sigma} : \nabla\mathbf{u}] \ \mathrm{d}\Omega = 0 \\ \mathrm{d}\Pi^{\mathrm{HR}}(\mathbf{u},\boldsymbol{\sigma})[\delta\mathbf{u}] = \displaystyle\int_\Omega [(\nabla\delta\mathbf{u}) : \boldsymbol{\sigma}] \ \mathrm{d}\Omega \\ \qquad\qquad - \displaystyle\int_\Omega [\delta\mathbf{u}\cdot\mathbf{b}] \ \mathrm{d}\Omega = 0 \end{cases} \tag{55}$$

which is now equivalent to the search of a saddle point. Changing sign to both equations and introducing the approximation

$$\begin{cases} \mathbf{u} \approx \mathbf{u}^h = \mathbf{N}^{\mathbf{u}}_k \hat{u}_k \\ \boldsymbol{\sigma} \approx \boldsymbol{\sigma}^h = \mathbf{N}^{\boldsymbol{\sigma}}_k \hat{\sigma}_k \end{cases} \tag{56}$$

as well as a similar approximation for the corresponding variation fields, equation (55) can be rewritten in matricial form as follows

$$\begin{bmatrix} \mathbf{A} & \mathbf{B}^{\mathrm{T}} \\ \mathbf{B} & \mathbf{0} \end{bmatrix} \begin{Bmatrix} \hat{\boldsymbol{\sigma}} \\ \hat{\mathbf{u}} \end{Bmatrix} = \begin{Bmatrix} \mathbf{0} \\ \mathbf{g} \end{Bmatrix} \tag{57}$$

where

$$\begin{cases} \mathbf{A}|_{ij} = \displaystyle\int_\Omega [\mathbf{N}_i^{\boldsymbol{\sigma}} : \mathbb{D}^{-1}\mathbf{N}_j^{\boldsymbol{\sigma}}]\,\mathrm{d}\Omega, & \hat{\boldsymbol{\sigma}}|_j = \hat{\sigma}_j \\ \mathbf{B}|_{rj} = -\displaystyle\int_\Omega [\nabla \mathbf{N}_r^{\mathbf{u}} : \mathbf{N}_j^{\boldsymbol{\sigma}}]\,\mathrm{d}\Omega, & \hat{\mathbf{u}}|_r = \hat{u}_r \\ \mathbf{g}|_r = -\displaystyle\int_\Omega [\mathbf{N}_r^{\mathbf{u}} \cdot \mathbf{b}]\,\mathrm{d}\Omega & \end{cases} \tag{58}$$

Starting from equation (54) the following modified Hellinger–Reissner functional can be also generated

$$\Pi^{\mathrm{HR,m}}(\mathbf{u}, \boldsymbol{\sigma}) = -\frac{1}{2}\int_\Omega \left[\boldsymbol{\sigma} : \mathbb{D}^{-1}\boldsymbol{\sigma}\right]\,\mathrm{d}\Omega - \int_\Omega [\mathbf{u} \cdot \operatorname{div}\boldsymbol{\sigma}]\,\mathrm{d}\Omega - \int_\Omega [\mathbf{b} \cdot \mathbf{u}]\,\mathrm{d}\Omega \tag{59}$$

and, requiring its stationarity, we obtain

$$\begin{cases} \mathrm{d}\Pi^{\mathrm{HR,m}}(\mathbf{u}, \boldsymbol{\sigma})[\delta\boldsymbol{\sigma}] = -\displaystyle\int_\Omega \left[\delta\boldsymbol{\sigma} : \mathbb{D}^{-1}\boldsymbol{\sigma}\right]\,\mathrm{d}\Omega \\ \qquad\qquad - \displaystyle\int_\Omega [\operatorname{div}(\delta\boldsymbol{\sigma}) \cdot \mathbf{u}]\,\mathrm{d}\Omega = 0 \\ \mathrm{d}\Pi^{\mathrm{HR,m}}(\mathbf{u}, \boldsymbol{\sigma})[\delta\mathbf{u}] = -\displaystyle\int_\Omega [\delta\mathbf{u} \cdot \operatorname{div}\boldsymbol{\sigma}]\,\mathrm{d}\Omega \\ \qquad\qquad - \displaystyle\int_\Omega [\delta\mathbf{u} \cdot \mathbf{b}]\,\mathrm{d}\Omega = 0 \end{cases} \tag{60}$$

which is again equivalent to the search of a saddle point. Changing sign to both equations and introducing again field approximation (56), equation (60) can rewritten in matricial form as equation (57), with the difference that now

$$\mathbf{B}|_{rj} = \int_\Omega \left[\mathbf{N}_r^{\mathbf{u}} \cdot \operatorname{div}\left(\mathbf{N}_j^{\boldsymbol{\sigma}}\right)\right]\,\mathrm{d}\Omega \tag{61}$$

Similarly, we may also associate an integral form to three field equation (42) starting from the even more general Hu–Washizu functional

$$\Pi^{\mathrm{HW}}(\mathbf{u}, \boldsymbol{\varepsilon}, \boldsymbol{\sigma}) = \frac{1}{2}\int_\Omega [\boldsymbol{\varepsilon} : \mathbb{D}\boldsymbol{\varepsilon}]\,\mathrm{d}\Omega - \int_\Omega \left[\boldsymbol{\sigma} : \left(\boldsymbol{\varepsilon} - \nabla^s\mathbf{u}\right)\right]\,\mathrm{d}\Omega - \int_\Omega [\mathbf{b} \cdot \mathbf{u}]\,\mathrm{d}\Omega \tag{62}$$

Requiring the stationarity of functional (62), we obtain

$$\begin{cases} \mathrm{d}\Pi^{\mathrm{HW}}(\mathbf{u}, \boldsymbol{\varepsilon}, \boldsymbol{\sigma})[\delta\boldsymbol{\varepsilon}] = \displaystyle\int_\Omega [\delta\boldsymbol{\varepsilon} : \mathbb{D}\boldsymbol{\varepsilon}]\,\mathrm{d}\Omega \\ \qquad\qquad - \displaystyle\int_\Omega [\delta\boldsymbol{\varepsilon} : \boldsymbol{\sigma}]\,\mathrm{d}\Omega = 0 \\ \mathrm{d}\Pi^{\mathrm{HW}}(\mathbf{u}, \boldsymbol{\varepsilon}, \boldsymbol{\sigma})[\delta\boldsymbol{\sigma}] = -\displaystyle\int_\Omega \left[\delta\boldsymbol{\sigma} : \left(\boldsymbol{\varepsilon} - \nabla^s\mathbf{u}\right)\right]\,\mathrm{d}\Omega = 0 \\ \mathrm{d}\Pi^{\mathrm{HW}}(\mathbf{u}, \boldsymbol{\varepsilon}, \boldsymbol{\sigma})[\delta\mathbf{u}] = \displaystyle\int_\Omega \left[\left(\nabla^s\delta\mathbf{u}\right) : \boldsymbol{\sigma}\right]\,\mathrm{d}\Omega \\ \qquad\qquad - \displaystyle\int_\Omega [\delta\mathbf{u} : \mathbf{b}]\,\mathrm{d}\Omega = 0 \end{cases} \tag{63}$$

which is again equivalent to the search of a saddle point. Introducing the following approximation

$$\begin{cases} \mathbf{u} \approx \mathbf{u}^h = \mathbf{N}_k^{\mathbf{u}}\hat{u}_k \\ \boldsymbol{\varepsilon} \approx \boldsymbol{\varepsilon}^h = \mathbf{N}_k^{\boldsymbol{\varepsilon}}\hat{\varepsilon}_k \\ \boldsymbol{\sigma} \approx \boldsymbol{\sigma}^h = \mathbf{N}_k^{\boldsymbol{\sigma}}\hat{\sigma}_k \end{cases} \tag{64}$$

as well as a similar approximation for the variation fields, equation (63) can be rewritten as follows

$$\begin{bmatrix} \mathbf{A} & \mathbf{B}^{\mathrm{T}} & \mathbf{0} \\ \mathbf{B} & \mathbf{0} & \mathbf{C}^{\mathrm{T}} \\ \mathbf{0} & \mathbf{C} & \mathbf{0} \end{bmatrix} \begin{Bmatrix} \hat{\boldsymbol{\varepsilon}} \\ \hat{\boldsymbol{\sigma}} \\ \hat{\mathbf{u}} \end{Bmatrix} = \begin{Bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{h} \end{Bmatrix} \tag{65}$$

where

$$\begin{cases} \mathbf{A}|_{ij} = \displaystyle\int_\Omega \left[\mathbf{N}_i^{\boldsymbol{\varepsilon}} : \mathbb{D}\mathbf{N}_j^{\boldsymbol{\varepsilon}}\right]\,\mathrm{d}\Omega, & \hat{\boldsymbol{\varepsilon}}|_j = \hat{\varepsilon}_j \\ \mathbf{B}|_{rj} = -\displaystyle\int_\Omega \left[\mathbf{N}_r^{\boldsymbol{\sigma}} : \mathbf{N}_j^{\boldsymbol{\varepsilon}}\right]\,\mathrm{d}\Omega, & \hat{\boldsymbol{\sigma}}|_r = \hat{\sigma}_r \\ \mathbf{C}|_{sr} = \displaystyle\int_\Omega \left[\nabla^s\mathbf{N}_s^{\mathbf{u}} : \mathbf{N}_r^{\boldsymbol{\sigma}}\right]\,\mathrm{d}\Omega, & \hat{\mathbf{u}}|_s = \hat{u}_s \\ \mathbf{h}|_s = \displaystyle\int_\Omega \left[\mathbf{N}_s^{\mathbf{u}} \cdot \mathbf{b}\right]\,\mathrm{d}\Omega & \end{cases} \tag{66}$$

For later consideration, we note that equation (65) can be rewritten as

$$\begin{bmatrix} \mathbb{A} & \mathbb{B}^{\mathrm{T}} \\ \mathbb{B} & \mathbf{0} \end{bmatrix} \begin{Bmatrix} \mathbf{x} \\ \mathbf{y} \end{Bmatrix} = \begin{Bmatrix} \mathbf{0} \\ \mathbf{h} \end{Bmatrix} \tag{67}$$

where we made the following simple identifications

$$\mathbb{A} = \begin{bmatrix} \mathbf{A} & \mathbf{B}^{\mathrm{T}} \\ \mathbf{B} & \mathbf{0} \end{bmatrix}, \quad \mathbb{B} = \{\mathbf{0}, \mathbf{C}\}$$

$$\mathbf{x} = \begin{Bmatrix} \hat{\boldsymbol{\varepsilon}} \\ \hat{\boldsymbol{\sigma}} \end{Bmatrix}, \quad \mathbf{y} = \hat{\mathbf{u}} \tag{68}$$

Examples of specific choices for the interpolating functions (51), (56), or (64) respectively within the single field, the two field, and the three field formulations can be found in standard textbooks (Bathe, 1996; Brezzi and Fortin, 1991; Hughes, 1987; Zienkiewicz and Taylor, 2000a) or in the literature.

Toward incompressible elasticity
It is interesting to observe that the strain $\boldsymbol{\varepsilon}$, the stress $\boldsymbol{\sigma}$ and the symmetric gradient of the displacement $\nabla^s \mathbf{u}$ can be easily decomposed respectively in a deviatoric (traceless) part and a volumetric (trace-related) part. In particular, recalling that we indicate with d the Euclidean space dimension, we may set

$$\begin{cases} \boldsymbol{\varepsilon} = \mathbf{e} + \dfrac{\theta}{d}\mathbf{I} & \text{with } \theta = \ \operatorname{tr}(\boldsymbol{\varepsilon}) \\ \boldsymbol{\sigma} = \mathbf{s} + p\mathbf{I} & \text{with } p = \dfrac{\operatorname{tr}(\boldsymbol{\sigma})}{d} \\ \nabla^s \mathbf{u} = \overline{\nabla^s \mathbf{u}} + \dfrac{\operatorname{div}\mathbf{u}}{d}\mathbf{I} & \text{with } \operatorname{div}(\mathbf{u}) \ = \ \operatorname{tr}(\nabla^s \mathbf{u}) \end{cases} \tag{69}$$

where θ, p, and $\operatorname{div}(\mathbf{u})$ are the volumetric (trace-related) quantities, while $\mathbf{e}$, $\mathbf{s}$, and $\overline{\nabla^s \mathbf{u}}$ are the deviatoric (or traceless) quantities, that is,

$$\operatorname{tr}(\mathbf{e}) \ = \ \operatorname{tr}(\mathbf{s}) \ = \ \operatorname{tr}\left(\overline{\nabla^s \mathbf{u}}\right) \ = 0 \tag{70}$$

Adopting these deviatoric-volumetric decompositions and limiting the discussion for simplicity of notation to the case of an isotropic material, the three field Hu–Washizu functional (62) can be rewritten as

$$\begin{aligned} \Pi^{\mathrm{HW,m}}(\mathbf{u}, \mathbf{e}, \theta, \mathbf{s}, p) &= \frac{1}{2}\int_\Omega [2\mu \mathbf{e} : \mathbf{e} + k\theta^2]\, \mathrm{d}\Omega \\ &- \int_\Omega [\mathbf{s} : (\mathbf{e} - \overline{\nabla^s \mathbf{u}})]\, \mathrm{d}\Omega - \int_\Omega [p(\theta - \ \operatorname{div}\mathbf{u})]\, \mathrm{d}\Omega \\ &- \int_\Omega [\mathbf{b} \cdot \mathbf{u}]\, \mathrm{d}\Omega \end{aligned} \tag{71}$$

where we introduce the bulk modulus $k = \lambda + 2\mu/d$. If we now require a strong (pointwise) satisfaction of the deviatoric compatibility condition $\mathbf{e} = \overline{\nabla^s \mathbf{u}}$ (obtained from the stationarity of functional (71) with respect to $\mathbf{s}$) as well as a strong (pointwise) satisfaction of the volumetric constitutive equation $p = k\theta$ (obtained from the stationarity of functional(71) with respect to θ), we end up with the following simpler modified Hellinger–Reissner functional

$$\begin{aligned} \Pi^{\mathrm{HR,m}}(\mathbf{u}, p) &= \frac{1}{2}\int_\Omega [2\mu \overline{\nabla^s \mathbf{u}} : \overline{\nabla^s \mathbf{u}}]\, \mathrm{d}\Omega - \frac{1}{2}\int_\Omega [\frac{1}{k}p^2]\, \mathrm{d}\Omega \\ &+ \int_\Omega [p \operatorname{div}\mathbf{u}]\, \mathrm{d}\Omega - \int_\Omega [\mathbf{b} \cdot \mathbf{u}]\, \mathrm{d}\Omega \end{aligned} \tag{72}$$

It is interesting to observe that taking the variation of functional (72) with respect to p, we obtain the correct relation between the pressure p and the volumetric component of the displacement gradient, that is,

$$p = k \operatorname{div}\mathbf{u} = \left(\lambda + \frac{2}{d}\mu\right) \operatorname{div}\mathbf{u} \tag{73}$$

For the case of incompressibility ($\lambda \to \infty$ and $k \to \infty$), functional (72) reduces to the following form

$$\begin{aligned} \Pi^{\mathrm{HR,m}}(\mathbf{u}, p) &= \frac{1}{2}\int_\Omega [2\mu \overline{\nabla^s \mathbf{u}} : \overline{\nabla^s \mathbf{u}}]\, \mathrm{d}\Omega + \int_\Omega [p \operatorname{div}\mathbf{u}]\, \mathrm{d}\Omega \\ &- \int_\Omega [\mathbf{b} \cdot \mathbf{u}]\, \mathrm{d}\Omega \end{aligned} \tag{74}$$

which resembles the potential energy functional (49) for the case of an isotropic material with the addition of the incompressibility constraint $\operatorname{div}\mathbf{u} = 0$ and with the difference that the quadratic term now involves only the deviatoric part of the symmetric displacement gradient and not the whole symmetric displacement gradient.

Requiring the stationarity of functional (74), we obtain

$$\begin{cases} \mathrm{d}\Pi^{\mathrm{HR,m}}(\mathbf{u}, p)[\delta\mathbf{u}] = \displaystyle\int_\Omega [2\mu \overline{\nabla^s \delta\mathbf{u}} : \overline{\nabla^s \mathbf{u}}]\, \mathrm{d}\Omega \\ \qquad + \displaystyle\int_\Omega [\ \operatorname{div}(\delta\mathbf{u}) \ \ p]\, \mathrm{d}\Omega - \int_\Omega [\delta\mathbf{u} \cdot \mathbf{b}]\, \mathrm{d}\Omega = 0 \\ \mathrm{d}\Pi^{\mathrm{HR,m}}(\mathbf{u}, p)[\delta p] = \displaystyle\int_\Omega [\delta p \operatorname{div}\mathbf{u}]\, \mathrm{d}\Omega = 0 \end{cases} \tag{75}$$

Introducing the following approximation

$$\begin{cases} \mathbf{u} \approx \mathbf{u}^h = \mathbf{N}_k^{\mathbf{u}} \hat{u}_k \\ p \approx p^h = N_k^p \hat{p}_k \end{cases} \tag{76}$$

as well as a similar approximation for the corresponding variation fields, equation (75) can be rewritten as follows

$$\begin{bmatrix} \mathbf{A} & \mathbf{B}^{\mathrm{T}} \\ \mathbf{B} & \mathbf{0} \end{bmatrix} \begin{Bmatrix} \hat{\mathbf{u}} \\ \hat{\mathbf{p}} \end{Bmatrix} = \begin{Bmatrix} \mathbf{f} \\ \mathbf{0} \end{Bmatrix} \tag{77}$$

where

$$\begin{cases} \mathbf{A}|_{ij} = 2\mu \displaystyle\int_\Omega \left[\overline{\nabla \mathbf{N}_i^{\mathbf{u}}} : \overline{\nabla \mathbf{N}_j^{\mathbf{u}}}\right] \mathrm{d}\Omega, \quad \hat{\mathbf{u}}|_j = \hat{u}_j \\ \mathbf{B}|_{rj} = \displaystyle\int_\Omega \left[N_r^p \ \operatorname{div}\left(\mathbf{N}_j^{\mathbf{u}}\right)\right] \mathrm{d}\Omega, \quad \hat{\mathbf{p}}|_r = \hat{p}_r \\ \mathbf{f}|_i = \displaystyle\int_\Omega \left[\mathbf{N}_i^{\mathbf{u}} \cdot \mathbf{b}\right] \mathrm{d}\Omega \end{cases} \tag{78}$$

It is interesting to observe that this approach may result in an unstable discrete formulation since the volumetric components of the symmetric part of the displacement gradient may not be controlled. Examples of specific choices for the interpolating functions (76) can be found in standard textbooks (Hughes, 1987; Zienkiewicz and Taylor, 2000a) or in the literature.

A different stable formulation can be easily obtained as in the case of Stokes problem. In particular, we may start from the potential energy functional (49), which for an isotropic material specializes as

$$\Pi(\mathbf{u}) = \frac{1}{2}\int_\Omega \left[2\mu\left(\nabla^s\mathbf{u} : \nabla^s\mathbf{u}\right) + \lambda\,(\operatorname{div}\mathbf{u})^2\right] \mathrm{d}\Omega - \int_\Omega [\mathbf{b}\cdot\mathbf{u}]\, \mathrm{d}\Omega \tag{79}$$

Introducing now the pressure-like field $\pi = \lambda \operatorname{div}\mathbf{u}$, we can rewrite functional (79) as

$$\Pi^m(\mathbf{u},\pi) = \frac{1}{2}\int_\Omega \left[2\mu\left(\nabla^s\mathbf{u} : \nabla^s\mathbf{u}\right) - \frac{1}{\lambda}\pi^2\right] \mathrm{d}\Omega + \int_\Omega [\pi \operatorname{div}\mathbf{u}]\, \mathrm{d}\Omega - \int_\Omega [\mathbf{b}\cdot\mathbf{u}]\, \mathrm{d}\Omega \tag{80}$$

We may note that π is a pressure-like quantity, different, however, from the physical pressure p, previously introduced. In fact, π is the Lagrangian multiplier associated to the incompressibility constraint and it can related to the physical pressure p recalling relation (73)

$$p = k \operatorname{div}\mathbf{u} = \pi + \frac{2}{d}\mu \operatorname{div}\mathbf{u} \tag{81}$$

For the incompressible case ($\lambda \to \infty$), functional (80) reduces to the following form:

$$\Pi^m(\mathbf{u},\pi) = \frac{1}{2}\int_\Omega \left[2\mu\nabla^s\mathbf{u} : \nabla^s\mathbf{u}\right] \mathrm{d}\Omega + \int_\Omega [\pi \operatorname{div}\mathbf{u}]\, \mathrm{d}\Omega - \int_\Omega [\mathbf{b}\cdot\mathbf{u}]\, \mathrm{d}\Omega \tag{82}$$

Taking the variation of (82) and introducing the following approximation

$$\begin{cases} \mathbf{u} \approx \mathbf{u}^h = \mathbf{N}^{\mathbf{u}}_k \hat{u}_k \\ \pi \approx \pi^h = N^\pi_k \hat{\pi}_k \end{cases} \tag{83}$$

as well as a similar approximation for the corresponding variation fields, we obtain a discrete problem of the following form:

$$\begin{bmatrix} \mathbf{A} & \mathbf{B}^{\mathrm{T}} \\ \mathbf{B} & \mathbf{0} \end{bmatrix} \begin{Bmatrix} \hat{\mathbf{u}} \\ \hat{\boldsymbol{\pi}} \end{Bmatrix} = \begin{Bmatrix} \mathbf{f} \\ \mathbf{0} \end{Bmatrix} \tag{84}$$

where

$$\begin{cases} \mathbf{A}|_{ij} = 2\mu \displaystyle\int_\Omega \left[\nabla^s\mathbf{N}^{\mathbf{u}}_i : \nabla^s\mathbf{N}^{\mathbf{u}}_j\right] \mathrm{d}\Omega, \quad \hat{\mathbf{u}}|_j = \hat{u}_j \\ \mathbf{B}|_{rj} = \displaystyle\int_\Omega \left[N^\pi_r \operatorname{div}\left(\mathbf{N}^{\mathbf{u}}_j\right)\right] \mathrm{d}\Omega, \quad \hat{\boldsymbol{\pi}}|_r = \hat{\pi}_r \\ \mathbf{f}|_i = \displaystyle\int_\Omega \left[\mathbf{N}^{\mathbf{u}}_i \cdot \mathbf{b}\right] \mathrm{d}\Omega \end{cases} \tag{85}$$

It is interesting to observe that, in general, this approach results in a stable discrete formulation since the volumetric components of the symmetric part of the displacement gradient are now controlled. Examples of specific choices for the interpolating functions (83) can be found in standard textbooks (Bathe, 1996; Brezzi and Fortin, 1991; Hughes, 1987; Zienkiewicz and Taylor, 2000a) or in the literature.

Enhanced strain formulation

Starting from the work of Simo and Rifai (1990), recently, a lot of attention has been paid to the so-called *enhanced strain formulation*, which can be variationally deduced for example, from the Hu–Washizu formulation (62). As a first step, the method describes the strain $\boldsymbol{\varepsilon}$ as the sum of a *compatible contribution*, $\nabla^s\mathbf{u}$, and of an *incompatible contribution*, $\tilde{\boldsymbol{\varepsilon}}$, that is,

$$\boldsymbol{\varepsilon} = \nabla^s\mathbf{u} + \tilde{\boldsymbol{\varepsilon}} \tag{86}$$

Using this position into the Hu–Washizu formulation (62), we obtain the following functional

$$\Pi^{\mathrm{enh}}(\mathbf{u},\tilde{\boldsymbol{\varepsilon}},\boldsymbol{\sigma}) = \frac{1}{2}\int_\Omega \left[\left(\nabla^s\mathbf{u} + \tilde{\boldsymbol{\varepsilon}}\right) : \mathbb{D}\left(\nabla^s\mathbf{u} + \tilde{\boldsymbol{\varepsilon}}\right)\right] \mathrm{d}\Omega - \int_\Omega \left[\boldsymbol{\sigma} : \tilde{\boldsymbol{\varepsilon}}\right] \mathrm{d}\Omega - \int_\Omega [\mathbf{b}\cdot\mathbf{u}]\, \mathrm{d}\Omega \tag{87}$$

Requiring the stationarity of the functional and introducing the following approximation

$$\begin{cases} \mathbf{u} \approx \mathbf{u}^h = \mathbf{N}^{\mathbf{u}}_k \hat{u}_k \\ \tilde{\boldsymbol{\varepsilon}} \approx \tilde{\boldsymbol{\varepsilon}}^h = \mathbf{N}^{\tilde{\boldsymbol{\varepsilon}}}_k \hat{\tilde{\varepsilon}}_k \\ \boldsymbol{\sigma} \approx \boldsymbol{\sigma}^h = \mathbf{N}^{\boldsymbol{\sigma}}_k \hat{\sigma}_k \end{cases} \tag{88}$$

as well as a similar approximation for the variation fields, we obtain the following discrete problem:

$$\begin{bmatrix} \mathbf{A} & \mathbf{B}^{\mathrm{T}} & \mathbf{0} \\ \mathbf{B} & \mathbf{C} & \mathbf{D}^{\mathrm{T}} \\ \mathbf{0} & \mathbf{D} & \mathbf{0} \end{bmatrix} \begin{Bmatrix} \hat{\mathbf{u}} \\ \hat{\boldsymbol{\varepsilon}} \\ \hat{\boldsymbol{\sigma}} \end{Bmatrix} = \begin{Bmatrix} \mathbf{f} \\ \mathbf{0} \\ \mathbf{0} \end{Bmatrix} \tag{89}$$

where

$$\begin{cases} \mathbf{A}|_{ij} = \int_\Omega \left[\nabla^s \mathbf{N}_i^{\mathbf{u}} : \mathbb{D}\nabla^s \mathbf{N}_j^{\mathbf{u}}\right] \mathrm{d}\Omega, \quad \hat{\mathbf{u}}|_j = \hat{u}_j \\ \mathbf{B}|_{rj} = \int_\Omega \left[\mathbf{N}_r^{\tilde{\boldsymbol{\varepsilon}}} : \mathbb{D}\nabla^s \mathbf{N}_j^{\mathbf{u}}\right] \mathrm{d}\Omega, \quad \hat{\tilde{\boldsymbol{\varepsilon}}}|_s = \hat{\tilde{\varepsilon}}_s \\ \mathbf{C}|_{rs} = \int_\Omega \left[\mathbf{N}_r^{\tilde{\boldsymbol{\varepsilon}}} : \mathbb{D}\mathbf{N}_s^{\tilde{\boldsymbol{\varepsilon}}}\right] \mathrm{d}\Omega, \quad \hat{\boldsymbol{\sigma}}|_r = \hat{\sigma}_r \\ \mathbf{D}|_{jr} = -\int_\Omega \left[\mathbf{N}_j^{\boldsymbol{\sigma}} : \mathbf{N}_r^{\boldsymbol{\varepsilon}}\right] \mathrm{d}\Omega, \quad \mathbf{f}|_j = \int_\Omega \left[\mathbf{N}_j^{\mathbf{u}} \cdot \mathbf{b}\right] \mathrm{d}\Omega \end{cases} \tag{90}$$

For later consideration, we note that equation (89) can be rewritten as

$$\begin{bmatrix} \mathbb{A} & \mathbb{B}^{\mathrm{T}} \\ \mathbb{B} & \mathbf{0} \end{bmatrix} \begin{Bmatrix} \mathbf{x} \\ \mathbf{y} \end{Bmatrix} = \begin{Bmatrix} \mathbf{d} \\ \mathbf{0} \end{Bmatrix} \tag{91}$$

where we made the following simple identifications:

$$\mathbb{A} = \begin{bmatrix} \mathbf{A} & \mathbf{B}^{\mathrm{T}} \\ \mathbf{B} & \mathbf{C} \end{bmatrix}, \quad \mathbb{B} = \{\mathbf{0}, \mathbf{D}\}$$
$$\mathbf{x} = \begin{Bmatrix} \hat{\mathbf{u}} \\ \hat{\tilde{\boldsymbol{\varepsilon}}} \end{Bmatrix}, \quad \mathbf{d} = \begin{Bmatrix} \mathbf{f} \\ \mathbf{0} \end{Bmatrix}, \quad \mathbf{y} = \hat{\boldsymbol{\sigma}} \tag{92}$$

Examples of specific choices for the interpolating functions can be found in standard textbooks (Zienkiewicz and Taylor, 2000a) or in the literature.

However, the most widely adopted enhanced strain formulation also requires the incompatible part of the strain to be orthogonal to the stress $\boldsymbol{\sigma}$

$$\int_\Omega \left[\boldsymbol{\sigma} : \tilde{\boldsymbol{\varepsilon}}\right] \mathrm{d}\Omega = 0 \tag{93}$$

If we use conditions (86) and (93) into the Hu–Washizu formulation (62), we obtain the following simplified functional:

$$\Pi^{\mathrm{enh}}(\mathbf{u}, \tilde{\boldsymbol{\varepsilon}}) = \frac{1}{2}\int_\Omega \left[\left(\nabla^s \mathbf{u} + \tilde{\boldsymbol{\varepsilon}}\right) : \mathbb{D}\left(\nabla^s \mathbf{u} + \tilde{\boldsymbol{\varepsilon}}\right)\right] \mathrm{d}\Omega - \int_\Omega [\mathbf{b} \cdot \mathbf{u}]\, \mathrm{d}\Omega \tag{94}$$

which closely resembles a standard displacement-based incompatible approach. Examples of specific choices for the interpolating functions involved in this simplified enhanced formulation can be found in standard textbooks (Zienkiewicz and Taylor, 2000a) or in the literature.

3 STABILITY OF SADDLE-POINTS IN FINITE DIMENSIONS

3.1 Solvability and stability

The examples discussed in Section 2 clearly show that, after discretization, several formulations typically lead to linear algebraic systems of the general form

$$\begin{bmatrix} \mathbf{A} & \mathbf{B}^{\mathrm{T}} \\ \mathbf{B} & \mathbf{0} \end{bmatrix} \begin{Bmatrix} \mathbf{x} \\ \mathbf{y} \end{Bmatrix} = \begin{Bmatrix} \mathbf{f} \\ \mathbf{g} \end{Bmatrix} \tag{95}$$

where $\mathbf{A}$ and $\mathbf{B}$ are respectively an $n \times n$ matrix and an $m \times n$ matrix, while $\mathbf{x}$ and $\mathbf{y}$ are respectively an $n \times 1$ vector and $m \times 1$ vector, as well as $\mathbf{f}$ and $\mathbf{g}$. Discretizations leading to such a system are often indicated as mixed finite element methods and in the following, we present a simple, algebraic version of the abstract theory that rules most applications of mixed methods.

Our first need is clearly to express in proper form *solvability* conditions for linear systems of type (95) in terms of the properties of the matrices $\mathbf{A}$ and $\mathbf{B}$. By solvability we mean that for every right-hand side $\mathbf{f}$ and $\mathbf{g}$ system, (95) has a unique solution. It is well known that this property holds *if and only if* the $(n+m) \times (n+m)$ matrix

$$\begin{bmatrix} \mathbf{A} & \mathbf{B}^{\mathrm{T}} \\ \mathbf{B} & \mathbf{0} \end{bmatrix} \tag{96}$$

is *nonsingular*, that is, if and only if its determinant is different from zero.

In order to have a good numerical method, however, solvability is not enough. An additional property that we also require is stability. We want to see this property with a little more detail. For a solvable finite-dimensional linear system, we always have continuous dependence of the solution upon the data. This means that there exists a constant c such that for every set of vectors $\mathbf{x}, \mathbf{y}, \mathbf{f}, \mathbf{g}$ satisfying (95) we have

$$\|\mathbf{x}\| + \|\mathbf{y}\| \leq c(\|\mathbf{f}\| + \|\mathbf{g}\|) \tag{97}$$

This property implies solvability. Indeed, if we assume that (97) holds for every set of vectors $\mathbf{x}, \mathbf{y}, \mathbf{f}, \mathbf{g}$ satisfying (95), then, whenever $\mathbf{f}$ and $\mathbf{g}$ are both zero, $\mathbf{x}$ and $\mathbf{y}$ must also be equal to zero. This is another way of saying that the homogeneous system has only the trivial solution, which implies that the determinant of the matrix (96) is different from zero, and hence the system is solvable.

However, Formula (97) deserves another very important comment. Actually, we did not specify the norms adopted for $\mathbf{x}, \mathbf{y}, \mathbf{f}, \mathbf{g}$. We had the right to do so, since in finite dimension all norms are equivalent. Hence, the change of one norm with another would only result in a change of the numerical value of the constant c, but it would not change the basic fact that such a constant exists. However, in dealing with linear systems resulting from the discretization of a partial differential equation we face a slightly different situation. In fact, if we want to analyze the behaviour of a given method when the meshsize becomes smaller and smaller, we must ideally consider a sequence of linear systems whose dimension increases and approaches infinity when the meshsize tends to zero. As it is well known (and it can be also easily verified), the constants involved in the equivalence of different norms depend on the dimension of the space. For instance, in $\mathbb{R}^n$, the two norms

$$\|\mathbf{x}\|_1 := \sum_{i=1}^{n} |x_i| \qquad \text{and} \qquad \|\mathbf{x}\|_2 := \left(\sum_{i=1}^{n} |x_i|^2\right)^{1/2} \tag{98}$$

are indeed equivalent, in the sense that there exist two positive constants c_1 and c_2 such that

$$c_1\|\mathbf{x}\|_1 \le \|\mathbf{x}\|_2 \le c_2\|\mathbf{x}\|_1 \tag{99}$$

for all $\mathbf{x}$ in $\mathbb{R}^n$. However, it can be rather easily checked that the *best* constants one can choose in (99) are

$$\|\mathbf{x}\|_2 \le \|\mathbf{x}\|_1 \le \sqrt{n}\|\mathbf{x}\|_2 \tag{100}$$

In particular, the first inequality becomes an equality, for instance, when x_1 is equal to 1 and all the other x_i's are zero, while the second inequality becomes an equality, for instance, when all the x_i are equal to 1.

When considering a sequence of problems with increasing dimension, we have to take into account that n and m become unbounded. It is then natural to ask if, for a given choice of the norms $\|\mathbf{x}\|$, $\|\mathbf{y}\|$, $\|\mathbf{f}\|$, and $\|\mathbf{g}\|$, it is possible to find a constant c independent of the meshsize (say, h), that is, a constant c that makes (97) hold true for all meshsizes.

However, even if inequality (97) holds with a constant c independent of h, it will not provide a good concept of stability unless the four norms are properly chosen (see Remark 18). This is going to be our next task.

3.2 Assumptions on the norms

We start denoting by $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{F}$, $\mathbf{G}$ respectively the spaces of vectors $\mathbf{x}$, $\mathbf{y}$, $\mathbf{f}$, $\mathbf{g}$. Then, we assume what follows.

1. The spaces $\mathbf{X}$ and $\mathbf{Y}$ are equipped with norms $\|\cdot\|_X$ and $\|\cdot\|_Y$ for which the matrices $\mathbf{A}$ and $\mathbf{B}$ satisfy the continuity conditions: *there exist two constants M_a and M_b, independent of the meshsize, such that for all* $\mathbf{x}$ *and* $\mathbf{z}$ *in* $\mathbf{X}$ *and for all* $\mathbf{y}$ *in* $\mathbf{Y}$

$$\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{z} \le M_a\|\mathbf{x}\|_X\|\mathbf{z}\|_X \quad \text{and} \quad \mathbf{x}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{y} \le M_b\|\mathbf{x}\|_X\|\mathbf{y}\|_Y \tag{101}$$

Moreover, we suppose there exist symmetric positive definite matrices $\mathbf{M}^x$ and $\mathbf{M}^y$, respectively of dimensions $n \times n$ and $m \times m$, such that

$$\|\mathbf{x}\|_X^2 = \mathbf{x}^{\mathrm{T}}\mathbf{M}^x\mathbf{x} \qquad \forall\ \mathbf{x} \in \mathbf{X} \tag{102}$$

and

$$\|\mathbf{y}\|_Y^2 = \mathbf{y}^{\mathrm{T}}\mathbf{M}^y\mathbf{y} \qquad \forall\ \mathbf{y} \in \mathbf{Y} \tag{103}$$

2. The spaces $\mathbf{F}$ and $\mathbf{G}$ are equipped with norms $\|\cdot\|_F$ and $\|\cdot\|_G$ defined as the dual norms of $\|\cdot\|_X$ and $\|\cdot\|_Y$, that is,

$$\|\mathbf{f}\|_F := \sup_{\mathbf{x}\in\mathbf{X}\setminus\{\mathbf{0}\}} \frac{\mathbf{x}^{\mathrm{T}}\mathbf{f}}{\|\mathbf{x}\|_X} \quad \text{and} \quad \|\mathbf{g}\|_G := \sup_{\mathbf{y}\in\mathbf{Y}\setminus\{\mathbf{0}\}} \frac{\mathbf{y}^{\mathrm{T}}\mathbf{g}}{\|\mathbf{y}\|_Y} \tag{104}$$

It is worth noting that

- assumptions (102) to (103) mean that the norms for $\mathbf{X}$ and $\mathbf{Y}$ are both induced by an inner product or, in other words, the norms at hand are *hilbertian* (as it happens in most of the applications);
- for every $\mathbf{x}$ and $\mathbf{f}$ in $\mathbb{R}^n$ and for every $\mathbf{y}$ and $\mathbf{g}$ in $\mathbb{R}^m$, we have

$$\mathbf{x}^{\mathrm{T}}\mathbf{f} \le \|\mathbf{x}\|_X\|\mathbf{f}\|_F \qquad \text{and} \qquad \mathbf{y}^{\mathrm{T}}\mathbf{g} \le \|\mathbf{y}\|_Y\|\mathbf{g}\|_G \tag{105}$$

- combining the continuity condition (101) on $||\cdot||_X$ and $||\cdot||_Y$ with the dual norm definition (105), for every $\mathbf{x} \in \mathbf{X}$ and for every $\mathbf{y} \in \mathbf{Y}$, we have the following relations:

$$\|\mathbf{A}\mathbf{x}\|_F = \sup_{\mathbf{z}\in\mathbf{X}\setminus\{\mathbf{0}\}} \frac{\mathbf{z}^{\mathrm{T}}\mathbf{A}\mathbf{x}}{\|\mathbf{z}\|_X} \le M_a\|\mathbf{x}\|_X \tag{106}$$

$$\|\mathbf{B}\mathbf{x}\|_G = \sup_{\mathbf{z}\in\mathbf{Y}\setminus\{\mathbf{0}\}} \frac{\mathbf{z}^{\mathrm{T}}\mathbf{B}\mathbf{x}}{\|\mathbf{z}\|_X} \le M_b\|\mathbf{x}\|_X \tag{107}$$

$$\|\mathbf{B}^{\mathrm{T}}\mathbf{y}\|_F = \sup_{\mathbf{z}\in\mathbf{X}\setminus\{\mathbf{0}\}} \frac{\mathbf{z}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{y}}{\|\mathbf{z}\|_X} \le M_b\|\mathbf{y}\|_Y \tag{108}$$

- if $\mathbf{A}$ is symmetric and positive semidefinite, then for every $\mathbf{x}, \mathbf{z} \in \mathbf{X}$

$$|\mathbf{z}^{\mathrm{T}}\mathbf{A}\mathbf{x}| \le (\mathbf{z}^{\mathrm{T}}\mathbf{A}\mathbf{z})^{1/2}(\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x})^{1/2} \tag{109}$$

so that (106) can be improved to

$$\|\mathbf{Ax}\|_F \leq \sup_{\mathbf{z}\in\mathbf{X}\setminus\{0\}} \frac{\mathbf{z}^{\mathrm{T}}\mathbf{Ax}}{\|\mathbf{z}\|_X} \leq M_a^{1/2}(\mathbf{x}^{\mathrm{T}}\mathbf{Ax})^{1/2} \qquad (110)$$

We are now ready to introduce a precise definition of stability.

Stability definition. *Given a numerical method, that produces a sequence of matrices* **A** *and* **B** *when applied to a given sequence of meshes (with the meshsize h going to zero), we choose norms* $\|\cdot\|_X$ *and* $\|\cdot\|_Y$ *that satisfy the continuity condition (101), and dual norms* $\|\cdot\|_F$ *and* $\|\cdot\|_G$ *according to (104). Then, we say that* **the method is stable** *if there exists a constant c,* **independent of the mesh**, *such that for all vectors* **x**, **y**, **f**, **g** *satisfying the general system (95), it holds*

$$\|\mathbf{x}\|_X + \|\mathbf{y}\|_Y \leq c(\|\mathbf{f}\|_F + \|\mathbf{g}\|_G) \qquad (111)$$

Having now a precise definition of stability, we can look for suitable assumptions on the matrices **A** and **B** that may provide the stability result (111). In particular, to guarantee stability condition (111), we need to introduce two assumptions involving such matrices. The first assumption, the so-called *inf–sup* condition, involves only the matrix **B** and it will be used throughout the whole section. To illustrate the second assumption we will first focus on a simpler but less general case that involves a ‘strong’ requirement on the matrix **A**. Among the problems presented in Section 2, this requirement is verified in practice only for the Stokes problem. Then, we shall tackle a more complex and clearly more general case, corresponding to a ‘weak’ requirement on the matrix **A**, suited for instance for discretizations of the mixed formulations of thermal diffusion problems.

Later on we shall deal with some additional complications that occur for instance, in the $(\mathbf{u}, \pi)$-formulation of nearly incompressible elasticity (cf. (80)). Finally, we shall briefly discuss more complicated problems, omitting the proofs for simplicity.

3.3 A requirement on the B matrix: the *inf–sup* condition

The basic assumption that we are going to use, throughout the whole section, deals with the matrix **B**. We assume the following:

***Inf–sup* condition**. *There exists a positive constant* β, **independent of the meshsize** *h, such that:*

$$\forall\, \mathbf{y} \in \mathbf{Y} \quad \exists \mathbf{x} \in \mathbf{X} \setminus \{\mathbf{0}\} \text{ such that } \mathbf{x}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{y} \geq \beta\|\mathbf{x}\|_X\|\mathbf{y}\|_Y \qquad (112)$$

Condition (112) requires the existence of a positive constant β, independent of h, such that for every $\mathbf{y} \in \mathbf{Y}$ we can find a suitable $\mathbf{x} \in \mathbf{X}$, different from **0** (and depending on **y**), such that (112) holds.

Remark 1. To better understand the meaning of (112), it might be useful to see when it fails. We thus consider the following $m \times n$ pseudodiagonal matrix ($m < n$)

$$\mathbf{B} = \begin{bmatrix} \vartheta_1 & 0 & \cdot & \cdot & \cdot & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & \vartheta_2 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & 0 & \vartheta_m & 0 & \cdot & \cdot & \cdot & \cdot & 0 \end{bmatrix} \qquad (113)$$

with $0 \leq \vartheta_1 \leq \vartheta_2 \leq \cdot\cdot \leq \vartheta_m \leq 1$. To fix ideas, we suppose that both $\mathbf{X} \equiv \mathbb{R}^n$ and $\mathbf{Y} \equiv \mathbb{R}^m$ are equipped with the standard Euclidean norms, which coincide with the corresponding dual norms on **F** and **G** (cf. (104)). If $\vartheta_1 = 0$, choosing $\mathbf{y} = (1, 0, \ldots, 0)^{\mathrm{T}} \neq \mathbf{0}$, we have $\mathbf{B}^{\mathrm{T}}\mathbf{y} = \mathbf{0}$. Therefore, for *every* $\mathbf{x} \in \mathbf{X}$, we get $\mathbf{x}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{y} = 0$ and condition (112) cannot hold since β must be positive. We then infer that condition (112) requires that

$$\text{no } \ \mathbf{y} \neq \mathbf{0} \text{ satisfies } \ \mathbf{B}^{\mathrm{T}}\mathbf{y} = \mathbf{0}$$

which, by definition, means that $\mathbf{B}^{\mathrm{T}}$ is *injective*. However, the injectivity of $\mathbf{B}^{\mathrm{T}}$ is not sufficient for the fulfillment of condition (112). Indeed, for $0 < \vartheta_1 \leq \vartheta_2 \leq \cdots \leq \vartheta_m \leq 1$, the matrix $\mathbf{B}^{\mathrm{T}}$ is injective and we have, still choosing $\mathbf{y} = (1, 0, \ldots, 0)^{\mathrm{T}}$,

$$\mathbf{B}^{\mathrm{T}}\mathbf{y} = (\vartheta_1, 0, \ldots, 0)^{\mathrm{T}} \neq \mathbf{0} \qquad (114)$$

Since for every $\mathbf{x} \in \mathbf{X}$ it holds

$$\mathbf{x}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{y} = \vartheta_1 x_1 \leq \vartheta_1 ||\mathbf{x}||_X = \vartheta_1 ||\mathbf{x}||_X ||\mathbf{y}||_Y \qquad (115)$$

we obtain that the constant β in (112) is forced to satisfy

$$0 < \beta \leq \vartheta_1 \qquad (116)$$

As a consequence, if $\vartheta_1 > 0$ tends to zero with the meshsize h, the matrix $\mathbf{B}^{\mathrm{T}}$ is still injective but condition (112) fails, because β, on top of being positive, must be independent of h. Noting that (see (114))

$$\frac{\|\mathbf{B}^{\mathrm{T}}\mathbf{y}\|_F}{\|\mathbf{y}\|_Y} = \vartheta_1 \qquad (117)$$

we then deduce that condition (112) requires that for $\mathbf{y} \neq \mathbf{0}$

$$\text{the vector } \ \mathbf{B}^{\mathrm{T}}\mathbf{y} \text{ is not ‘too small’ with respect to } \ \mathbf{y}$$

which is a property stronger than the injectivity of the matrix $\mathbf{B}^{\mathrm{T}}$. We will see in Proposition 1 that all these considerations on the particular matrix **B** in (113) does extend to the general case.

We now rewrite condition (112) in different equivalent forms, which will also make clear the reason why it is called *inf–sup condition*.

Since, by assumption, **x** is different from zero, condition (112) can equivalently be written as

$$\forall\ \mathbf{y} \in \mathbf{Y} \quad \exists \mathbf{x} \in \mathbf{X}\backslash\{\mathbf{0}\} \quad \text{such that} \quad \frac{\mathbf{x}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{y}}{\|\mathbf{x}\|_X} \geq \beta\|\mathbf{y}\|_Y \tag{118}$$

This last form (118) highlights that given $\mathbf{y} \in \mathbf{Y}$, the most suitable $\mathbf{x} \in \mathbf{X}$ is the one that makes the left-hand side of (118) as big as possible. Hence, the best we can do is to take the *supremum* of the left-hand side, when **x** varies among all possible $\mathbf{x} \in \mathbf{X}$ different from **0**. Hence, we may equivalently require that

$$\forall\ \mathbf{y} \in \mathbf{Y} \quad \sup_{\mathbf{x}\in\mathbf{X}\backslash\{\mathbf{0}\}} \frac{\mathbf{x}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{y}}{\|\mathbf{x}\|_X} \geq \beta\|\mathbf{y}\|_Y \tag{119}$$

In a sense, we got rid of the task of choosing **x**. However, condition (119) still depends on **y** and it clearly holds for $\mathbf{y} = \mathbf{0}$. Therefore, we can concentrate on the **y**'s that are different from **0**; in particular, for $\mathbf{y} \neq 0$ condition (119) can be also written as

$$\sup_{\mathbf{x}\in\mathbf{X}\backslash\{\mathbf{0}\}} \frac{\mathbf{x}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{y}}{\|\mathbf{x}\|_X\|\mathbf{y}\|_Y} \geq \beta \tag{120}$$

The worst possible **y** is therefore the one that makes the left-hand side of (120) as small as possible. If we want (120) to hold for every $\mathbf{y} \in \mathbf{Y}$ we might as well consider the worst case, looking directly at the *infimum* of the left-hand side of (120) among all possible **y**'s, requiring that

$$\inf_{\mathbf{y}\in\mathbf{Y}\backslash\{\mathbf{0}\}} \sup_{\mathbf{x}\in\mathbf{X}\backslash\{\mathbf{0}\}} \frac{\mathbf{x}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{y}}{\|\mathbf{x}\|_X\|\mathbf{y}\|_Y} \geq \beta \tag{121}$$

The advantage of formulation (121), if any, is that we got rid of the dependency on **y** as well. Indeed, condition (121) is now a condition on the matrix **B**, on the spaces **X** and **Y** (together with their norms) as well as on the crucial constant β.

Let us see now the relationship of the *inf–sup* condition with a basic property of the matrix **B**.

Proposition 1. *The inf–sup condition (112) is equivalent to require that*

$$\beta\|\mathbf{y}\|_Y \leq \|\mathbf{B}^{\mathrm{T}}\mathbf{y}\|_F \qquad \forall\ \mathbf{y} \in \mathbf{Y} \tag{122}$$

Therefore, in particular, the inf–sup condition implies that the matrix $\mathbf{B}^{\mathrm{T}}$ *is injective.*

Proof. Assume that the *inf–sup* condition (112) holds, and let **y** be any vector in **Y**. By the equivalent form (119) and using definition (104) of the dual norm $\|\cdot\|_F$, we have

$$\beta\|\mathbf{y}\|_Y \leq \sup_{\mathbf{x}\in\mathbf{X}\backslash\{\mathbf{0}\}} \frac{\mathbf{x}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{y}}{\|\mathbf{x}\|_X} = \|\mathbf{B}^{\mathrm{T}}\mathbf{y}\|_F \tag{123}$$

and therefore (122) holds true. Moreover, the matrix $\mathbf{B}^{\mathrm{T}}$ is injective since (122) shows that $\mathbf{y} \neq \mathbf{0}$ implies $\mathbf{B}^{\mathrm{T}}\mathbf{y} \neq \mathbf{0}$.

Assume conversely that (122) holds. Using again the definition (104) of the dual norm $\|\cdot\|_F$, we have

$$\beta\|\mathbf{y}\|_Y \leq \|\mathbf{B}^{\mathrm{T}}\mathbf{y}\|_F = \sup_{\mathbf{x}\in\mathbf{X}\backslash\{\mathbf{0}\}} \frac{\mathbf{x}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{y}}{\|\mathbf{x}\|_X} \tag{124}$$

which implies the *inf–sup* condition in the form (119). □

Remark 2. Whenever the $m \times n$ matrix **B** satisfies the *inf–sup* condition, the injectivity of $\mathbf{B}^{\mathrm{T}}$ implies that $n \geq m$. We point out once again (cf. Remark 1) that the injectivity of $\mathbf{B}^{\mathrm{T}}$ is not sufficient for the fulfillment of the *inf–sup* condition.

Additional relationships between the *inf–sup* and other properties of the matrix **B** will be presented later on in Section 3.5.

3.4 A 'strong' condition on the A matrix. Ellipticity on the whole space — Stokes

As we shall see in the sequel, the *inf–sup* condition is a necessary condition for having stability of problems of the general form (95). In order to have sufficient conditions, we now introduce a further assumption on the matrix **A**. As discussed at the end of Section 3.2, we start considering a strong condition on the matrix **A**. More precisely, we assume the following:

Ellipticity condition. *There exists a positive constant* α, *independent of the meshsize h, such that*

$$\alpha\|\mathbf{x}\|_X^2 \leq \mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} \qquad \forall\ \mathbf{x} \in \mathbf{X} \tag{125}$$

We first notice that from (101) and (125) it follows that

$$\alpha \leq M_a \tag{126}$$

We have now the following theorem.

Theorem 1. *Let* $\mathbf{x}, \mathbf{y}, \mathbf{f}, \mathbf{g}$ *satisfy the general system of equations (95). Moreover, assume that* **A** *is symmetric and* *that the continuity conditions (101), the dual norm assumptions (105), the inf–sup (112) and the ellipticity requirement (125) are all satisfied. Then, we have*

$$\|\mathbf{x}\|_X \leq \frac{1}{\alpha}\|\mathbf{f}\|_F + \frac{M_a^{1/2}}{\alpha^{1/2}\beta}\|\mathbf{g}\|_G \tag{127}$$

$$\|\mathbf{y}\|_Y \leq \frac{2M_a^{1/2}}{\alpha^{1/2}\beta}\|\mathbf{f}\|_F + \frac{M_a}{\beta^2}\|\mathbf{g}\|_G \tag{128}$$

Proof. We shall prove the result by splitting $\mathbf{x} = \mathbf{x}_f + \mathbf{x}_g$ and $\mathbf{y} = \mathbf{y}_f + \mathbf{y}_g$ defined as the solutions of

$$\begin{cases} \mathbf{A}\mathbf{x}_f + \mathbf{B}^{\mathrm{T}}\mathbf{y}_f = \mathbf{f} \\ \mathbf{B}\mathbf{x}_f = \mathbf{0} \end{cases} \tag{129}$$

and

$$\begin{cases} \mathbf{A}\mathbf{x}_g + \mathbf{B}^{\mathrm{T}}\mathbf{y}_g = \mathbf{0} \\ \mathbf{B}\mathbf{x}_g = \mathbf{g} \end{cases} \tag{130}$$

We proceed in several steps.

• *Step 1 – Estimate of* $\mathbf{x}_f$ *and* $\mathbf{A}\mathbf{x}_f$ We multiply the first equation of (129) to the left by $\mathbf{x}_f^{\mathrm{T}}$ and we notice that $\mathbf{x}_f^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{y}_f \equiv \mathbf{y}^{\mathrm{T}}\mathbf{B}\mathbf{x}_f = 0$ (by the second equation). Hence,

$$\mathbf{x}_f^{\mathrm{T}}\mathbf{A}\mathbf{x}_f = \mathbf{x}^{\mathrm{T}}\mathbf{f} \tag{131}$$

and, using the ellipticity condition (125), relation (131), and the first of the dual norm estimates (105), we have

$$\alpha\|\mathbf{x}_f\|_X^2 \leq \mathbf{x}_f^{\mathrm{T}}\mathbf{A}\mathbf{x}_f = \mathbf{x}^{\mathrm{T}}\mathbf{f} \leq \|\mathbf{x}_f\|_X\|\mathbf{f}\|_F \tag{132}$$

giving immediately

$$\|\mathbf{x}_f\|_X \leq \frac{1}{\alpha}\|\mathbf{f}\|_F \tag{133}$$

and

$$\mathbf{x}_f^{\mathrm{T}}\mathbf{A}\mathbf{x}_f \leq \frac{1}{\alpha}\|\mathbf{f}\|_F^2 \tag{134}$$

Therefore, using (110) we also get

$$\|\mathbf{A}\mathbf{x}_f\|_F \leq \frac{M_a^{1/2}}{\alpha^{1/2}}\|\mathbf{f}\|_F \tag{135}$$

• *Step 2 – Estimate of* $\mathbf{y}_f$ We use now the *inf–sup* condition (112) with $\mathbf{y} = \mathbf{y}_f$. We obtain that there exists $\widetilde{\mathbf{x}} \in \mathbf{X}$ such that $\widetilde{\mathbf{x}}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{y}_f \geq \beta\|\widetilde{\mathbf{x}}\|_X\|\mathbf{y}_f\|_Y$. Multiplying the first equation of (129) by $\widetilde{\mathbf{x}}^{\mathrm{T}}$ and using the first of the dual norm estimates (105), we have

$$\begin{aligned}\beta\|\widetilde{\mathbf{x}}\|_X\|\mathbf{y}_f\|_Y &\leq \widetilde{\mathbf{x}}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{y}_f = \widetilde{\mathbf{x}}^{\mathrm{T}}(\mathbf{f} - \mathbf{A}\mathbf{x}_f) \\ &\leq \|\widetilde{\mathbf{x}}\|_X\|\mathbf{f} - \mathbf{A}\mathbf{x}_f\|_F \end{aligned}\tag{136}$$

We now use the fact that in the *inf–sup* condition (112) we had $\widetilde{\mathbf{x}} \neq 0$, so that in the above equation (136) we can simplify by its norm. Then, using (135) and (126), we obtain

$$\begin{aligned}\|\mathbf{y}_f\|_Y &\leq \frac{1}{\beta}\|\mathbf{f} - \mathbf{A}\mathbf{x}_f\|_F \leq \left(\frac{1}{\beta} + \frac{M_a^{1/2}}{\alpha^{1/2}\beta}\right)\|\mathbf{f}\|_F \\ &\leq \frac{2M_a^{1/2}}{\alpha^{1/2}\beta}\|\mathbf{f}\|_F \end{aligned}\tag{137}$$

• *Step 3 – Estimate of* $\mathbf{x}_g^{\mathrm{T}}\mathbf{A}\mathbf{x}_g$ *by* $\|\mathbf{y}_g\|_Y$ We multiply the first equation of (130) by $\mathbf{x}_g^{\mathrm{T}}$. Using the second equation of (130) and the second of the dual norm estimates (105), we have

$$\mathbf{x}_g^{\mathrm{T}}\mathbf{A}\mathbf{x}_g = -\mathbf{x}_g^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{y}_g \equiv \mathbf{y}_g^{\mathrm{T}}\mathbf{B}\mathbf{x}_g = \mathbf{y}_g^{\mathrm{T}}\mathbf{g} \leq \|\mathbf{y}_g\|_Y\|\mathbf{g}\|_G \tag{138}$$

• *Step 4 – Estimate of* $\|\mathbf{y}_g\|_Y$ *by* $(\mathbf{x}_g^{\mathrm{T}}\mathbf{A}\mathbf{x}_g)^{1/2}$ We proceed as in *Step 2*. Using the *inf–sup* condition (112) with $\mathbf{y} = \mathbf{y}_g$ we get a new vector, that we call again $\widetilde{\mathbf{x}}$, such that $\widetilde{\mathbf{x}}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{y}_g \geq \beta\|\widetilde{\mathbf{x}}\|_X\|\mathbf{y}_g\|_Y$. This relation, the first equation of (130), and the continuity property (109) yield

$$\begin{aligned}\beta\|\widetilde{\mathbf{x}}\|_X\|\mathbf{y}_g\|_Y &\leq \widetilde{\mathbf{x}}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{y}_g = -\widetilde{\mathbf{x}}^{\mathrm{T}}\mathbf{A}\mathbf{x}_g \\ &\leq M_a^{1/2}\|\widetilde{\mathbf{x}}\|_X(\mathbf{x}_g^{\mathrm{T}}\mathbf{A}\mathbf{x}_g)^{1/2} \end{aligned}\tag{139}$$

giving

$$\|\mathbf{y}_g\|_Y \leq \frac{M_a^{1/2}}{\beta}(\mathbf{x}_g^{\mathrm{T}}\mathbf{A}\mathbf{x}_g)^{1/2} \tag{140}$$

• *Step 5 – Estimate of* $\|\mathbf{x}_g\|_X$ *and* $\|\mathbf{y}_g\|_Y$ We first combine (138) and (140) to obtain

$$\|\mathbf{y}_g\|_Y \leq \frac{M_a}{\beta^2}\|\mathbf{g}\|_G \tag{141}$$

Moreover, using the ellipticity assumption (125) in (138) and inserting (141), we have

$$\alpha\|\mathbf{x}_g\|_X^2 \leq \mathbf{x}_g^{\mathrm{T}}\mathbf{A}\mathbf{x}_g \leq \|\mathbf{y}_g\|_Y\|\mathbf{g}\|_G \leq \frac{M_a}{\beta^2}\|\mathbf{g}\|_G^2 \tag{142}$$

which can be rewritten as

$$\|\mathbf{x}_g\|_X \leq \frac{M_a^{1/2}}{\alpha^{1/2}\beta}\|\mathbf{g}\|_G \tag{143}$$

The final estimate follows then by simply collecting the separate estimates (133), (137), (143), and (141). □

A straightforward consequence of Theorem 1 and Remark 4 is the following stability result (cf. (111)):

Corollary 1. *Assume that a numerical method produces a sequence of matrices* **A** *and* **B** *for which both the inf–sup condition (112) and the ellipticity condition (125) are satisfied. Then the method is stable.*

Remark 3. In certain applications, it might happen that the constants α and β either depend on h (and tend to zero as h tends to zero) or have a fixed value that is however very small. It is therefore important to keep track of the possible degeneracy of the constants in our estimates when α and/or β are very small. In particular, it is relevant to know whether our stability constants degenerate, say, as $1/\beta$, or $1/\beta^2$, or other powers of $1/\beta$ (and, similarly, of $1/\alpha$). In this respect, we point out that the behavior indicated in (127) and (128) is optimal. This means that we cannot hope to find a better proof giving a better behavior of the constants in terms of powers of $1/\alpha$ and $1/\beta$. Indeed, consider the system

$$\begin{bmatrix} 2 & \sqrt{a} & b \\ \sqrt{a} & a & 0 \\ b & 0 & 0 \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \\ y \end{Bmatrix} = \begin{Bmatrix} f_1 \\ f_2 \\ g \end{Bmatrix} \qquad 0 < a, b \ll 1 \tag{144}$$

whose solution is

$$x_1 = \frac{g}{b}, \quad x_2 = \frac{f_2}{a} - \frac{g}{a^{1/2}b}, \quad y = \frac{f_1}{b} - \frac{f_2}{a^{1/2}b} - \frac{g}{b^2} \tag{145}$$

Since the constants α and β are given by

$$\alpha = \frac{2 + a - \sqrt{a^2 + 4}}{2} = \frac{4a}{2\left(2 + a + \sqrt{a^2 + 4}\right)} \approx a$$

and

$$\beta = b$$

we see from (145) that there are cases in which the actual stability constants behave exactly as predicted by the theory.

Remark 4. We point out that the symmetry condition on the matrix **A** is not necessary. Indeed, with a slightly different (and even simpler) proof one can prove stability without the symmetry assumption. The dependence of the stability constant upon α and β is however worse, as it can be seen in the following example. Considering the system

$$\begin{bmatrix} 1 & -1 & b \\ 1 & a & 0 \\ b & 0 & 0 \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \\ y \end{Bmatrix} = \begin{Bmatrix} f_1 \\ f_2 \\ g \end{Bmatrix} \qquad 0 < a, b \ll 1 \tag{146}$$

one easily obtains

$$x_1 = \frac{g}{b}, \quad x_2 = \frac{f_2}{a} - \frac{g}{ab}, \quad y = \frac{f_1}{b} + \frac{f_2}{ab} - \frac{(1 + a)g}{ab^2} \tag{147}$$

Since $\alpha = a$ and $\beta = b$, from (147) we deduce that the bounds of Theorem 1 cannot hold when **A** is not symmetric.

As announced in the title of the section the situation in which **A** is elliptic in the whole space is typical (among others) of the Stokes problem, as presented in (22) to (24). Indeed, denoting the interpolating functions for **u** and p by $\mathbf{N}_i^{\mathbf{u}}$ and N_r^p respectively (cf. (39)), if we set

$$\|\hat{\mathbf{u}}\|_X^2 := \mu \int_\Omega \left|\nabla(\mathbf{N}_i^{\mathbf{u}} \hat{u}_i)\right|^2 \, \mathrm{d}\Omega \tag{148}$$

and

$$\|\hat{\mathbf{p}}\|_Y^2 := \int_\Omega \left|N_r^p \hat{p}_r\right|^2 \, \mathrm{d}\Omega \tag{149}$$

we can easily see that conditions (101) are verified with $M_a = 1$ and $M_b = \sqrt{(d/\mu)}$ respectively. Clearly the ellipticity property (125) is also verified with $\alpha = 1$, no matter what is the choice of the mesh and of the interpolating functions. On the other hand, the *inf–sup* Property (112) is much less obvious, as we are going to see in Section 4, and finite element choices have to be specially tailored in order to satisfy it.

3.5 The *inf–sup* condition and the lifting operator

In this section, we shall see that the *inf–sup* condition is related to another important property of the matrix **B**. Before proceeding, we recall that an $m \times n$ matrix **B** is surjective if for every $\mathbf{g} \in \mathbb{R}^m$, there exists $\mathbf{x}_g \in \mathbb{R}^n$ such that $\mathbf{B}\mathbf{x}_g = \mathbf{g}$.

We have the following Proposition.

Proposition 2. *The inf–sup condition (112) is equivalent to require the existence of a lifting operator* $\mathbf{L}\colon \mathbf{g} \to \mathbf{x}_g = \mathbf{L}\mathbf{g}$ *such that, for every* $\mathbf{g} \in \mathbb{R}^m$, *it holds*

$$\begin{cases} \mathbf{B}\mathbf{x}_g = \mathbf{g} \\ \beta\|\mathbf{x}_g\|_X \equiv \beta\|\mathbf{L}\mathbf{g}\|_X \le \|\mathbf{g}\|_G \equiv \|\mathbf{B}\mathbf{x}_g\|_G \end{cases} \tag{150}$$

Therefore, in particular, the inf–sup condition implies that the matrix **B** *is surjective.*

Proof. We begin by recalling that there exists a symmetric $(n \times n)$ positive definite matrix $\mathbf{M}^x$ such that (cf. (102))

$$\mathbf{x}^{\mathrm{T}}\mathbf{M}^x\mathbf{x} = \|\mathbf{x}\|_X^2 \tag{151}$$

It is clear that the choice $\mathbf{A} \equiv \mathbf{M}^x$ easily satisfies the first of the continuity conditions (101) with $M_a = 1$, as well as the ellipticity condition (125) with $\alpha = 1$. Given $\mathbf{g}$, if the *inf–sup* condition holds, we can, therefore, use Theorem 1 and find a unique solution $(\widetilde{\mathbf{x}}_g, \widetilde{\mathbf{y}}_g)$ of the following auxiliary problem:

$$\begin{cases} \mathbf{M}^x\widetilde{\mathbf{x}}_g + \mathbf{B}^{\mathrm{T}}\widetilde{\mathbf{y}}_g = \mathbf{0}, \\ \mathbf{B}\widetilde{\mathbf{x}}_g = \mathbf{g} \end{cases} \tag{152}$$

We can now use estimate (143) from *Step 5* of the proof of Theorem 1, recalling that in our case, $\alpha = M_a = 1$ since we are using the matrix $\mathbf{M}^x$ instead of $\mathbf{A}$. We obtain

$$\beta\|\widetilde{\mathbf{x}}_g\|_X \le \|\mathbf{g}\|_G \tag{153}$$

It is then clear that setting $\mathbf{Lg} := \widetilde{\mathbf{x}}_g$ (the first part of the solution of the auxiliary problem (152)) we have that estimate (150) in our statement holds true.

Assume conversely that we have the existence of a continuous lifting $\mathbf{L}$ satisfying (150). First we recall that there exists a symmetric $(m \times m)$ positive definite matrix $\mathbf{M}^y$ such that (cf. (103))

$$\mathbf{y}^{\mathrm{T}}\mathbf{M}^y\mathbf{y} = \|\mathbf{y}\|_Y^2 \tag{154}$$

Then, for a given $\mathbf{y} \in \mathbf{Y}$, we set first $\mathbf{g} := \mathbf{M}^y\mathbf{y}$ (so that $\mathbf{y}^{\mathrm{T}}\mathbf{g} = \|\mathbf{y}\|_Y^2$) and then we define $\mathbf{x}_g := \mathbf{Lg}$ (so that $\mathbf{Bx}_g = \mathbf{g}$). Hence,

$$\mathbf{x}_g^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{y} \equiv \mathbf{y}^{\mathrm{T}}\mathbf{B}\mathbf{x}_g = \mathbf{y}^{\mathrm{T}}\mathbf{g} = \mathbf{y}^{\mathrm{T}}\mathbf{M}^y\mathbf{y} = \|\mathbf{y}\|_Y^2 \tag{155}$$

On the other hand, it is easy to see that using (150) we have

$$\beta\|\mathbf{x}_g\|_X \le \|\mathbf{g}\|_G \le \|\mathbf{y}\|_Y \tag{156}$$

where the last inequality is based on the choice $\mathbf{g} = \mathbf{M}^y\mathbf{y}$ and the use of (108) with $\mathbf{M}^y$ in the place of $\mathbf{B}^{\mathrm{T}}$. Hence, for every $\mathbf{y} \in \mathbf{Y}$, different from zero, we constructed $\mathbf{x} = \mathbf{x}_g \in \mathbf{X}$, different from zero, which, joining (155) and (156), satisfies

$$\mathbf{x}_g^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{y} = \|\mathbf{y}\|_Y^2 \ge \beta\|\mathbf{x}_g\|_X\|\mathbf{y}\|_Y \tag{157}$$

that is, the *inf–sup* condition in its original form (112). □

3.6 A ‘weak’ condition on the A matrix. Ellipticity on the kernel — thermal diffusion

We now consider that, together with the *inf–sup* condition on $\mathbf{B}$, the condition on $\mathbf{A}$ is weaker than the full Ellipticity (125). In particular, we require the ellipticity of $\mathbf{A}$ to hold only in a subspace $\mathbf{X}_0$ of the whole space $\mathbf{X}$, with $\mathbf{X}_0$ defined as follows:

$$\mathbf{X}_0 := \mathrm{Ker}(\mathbf{B}) \equiv \{\mathbf{x} \in \mathbf{X} \ \text{ such that } \ \mathbf{Bx} = \mathbf{0}\} \tag{158}$$

More precisely, we require the following:

***Elker* condition.** *There exists a positive constant α_0, independent of the meshsize h, such that*

$$\alpha_0\|\mathbf{x}\|_X^2 \le \mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} \qquad \forall\, \mathbf{x} \in \mathbf{X}_0 \tag{159}$$

The above condition is often called *elker* since it requires the ellipticity on the kernel. Moreover, from (101) and (159), we get

$$\alpha_0 \le M_a \tag{160}$$

The following theorem generalizes Theorem 1. For the sake of completeness, we present here the proof in the case of a matrix $\mathbf{A}$ that is not necessarily symmetric.

Theorem 2. *Let $\mathbf{x} \in \mathbf{X}$ and $\mathbf{y} \in \mathbf{Y}$ satisfy system (1) and assume that the continuity conditions (101), the dual norm assumptions (105), the inf–sup (112), and the elker condition (159) are satisfied. Then, we have*

$$\|\mathbf{x}\|_X \le \frac{1}{\alpha_0}\|\mathbf{f}\|_F + \frac{2M_a}{\alpha_0\beta}\|\mathbf{g}\|_G \tag{161}$$

$$\|\mathbf{y}\|_Y \le \frac{2M_a}{\alpha_0\beta}\|\mathbf{f}\|_F + \frac{2M_a^2}{\alpha_0\beta^2}\|\mathbf{g}\|_G \tag{162}$$

Proof. We first set $\mathbf{x}_g := \mathbf{Lg}$ where $\mathbf{L}$ is the lifting operator defined by Proposition 2. We also point out the following estimates on $\mathbf{x}_g$: from the continuity of the lifting $\mathbf{L}$ (150), we have

$$\beta\|\mathbf{x}_g\|_X \le \|\mathbf{g}\|_G \tag{163}$$

and using (106) and (163), we obtain

$$\|\mathbf{A}\mathbf{x}_g\|_F \le M_a\|\mathbf{x}_g\|_X \le \frac{M_a}{\beta}\|\mathbf{g}\|_G \tag{164}$$

Then, we set

$$\mathbf{x}_0 := \mathbf{x} - \mathbf{x}_g = \mathbf{x} - \mathbf{Lg} \tag{165}$$

and we notice that $\mathbf{x}_0 \in \mathbf{X}_0$. Moreover, $(\mathbf{x}_0, \mathbf{y})$ solves the linear system

$$\begin{cases} \mathbf{A}\mathbf{x}_0 + \mathbf{B}^{\mathrm{T}}\mathbf{y} = \mathbf{f} - \mathbf{A}\mathbf{x}_g, \\ \mathbf{B}\mathbf{x}_0 = \mathbf{0} \end{cases} \tag{166}$$

We can now proceed as in *Steps 1* and *2* of the proof of Theorem 1 (as far as we do not use (110), since we gave up the symmetry assumption). We note that our weaker assumption *elker* (159) is sufficient for allowing the first step in (132). Proceeding as in the first part of *Step 1*, and using (164) at the end, we get

$$\|\mathbf{x}_0\|_X \le \frac{1}{\alpha_0}\|\mathbf{f}-\mathbf{A}\mathbf{x}_g\|_F \le \frac{1}{\alpha_0}\left(\|\mathbf{f}\|_F + \frac{M_a}{\beta}\|\mathbf{g}\|_G\right) \quad (167)$$

This allows to reconstruct the estimate on $\mathbf{x}$:

$$\begin{aligned}\|\mathbf{x}\|_X = \|\mathbf{x}_0+\mathbf{x}_g\|_X &\le \frac{1}{\alpha_0}\|\mathbf{f}\|_F + \left(\frac{M_a}{\alpha_0\beta}+\frac{1}{\beta}\right)\|\mathbf{g}\|_G \\ &\le \frac{1}{\alpha_0}\|\mathbf{f}\|_F + \frac{2M_a}{\alpha_0\beta}\|\mathbf{g}\|_G \qquad (168)\end{aligned}$$

where we have used (160) in the last inequality. Combining (106) and (168), we also have

$$\|\mathbf{A}\mathbf{x}\|_F \le M_a\|\mathbf{x}\|_X \le \frac{M_a}{\alpha_0}\|\mathbf{f}\|_F + \frac{2M_a^2}{\alpha_0\beta}\|\mathbf{g}\|_G \qquad (169)$$

which is weaker than (135) since we could not use the symmetry assumption. Then, we proceed as in *Step 2* to obtain, as in (137)

$$\beta\|\mathbf{y}\|_Y \le \|\mathbf{f}-\mathbf{A}\mathbf{x}\|_F \qquad (170)$$

and using the above estimate (169) on $\mathbf{Ax}$ in (170), we obtain

$$\begin{aligned}\|\mathbf{y}\|_Y &\le \left(\frac{1}{\beta}+\frac{M_a}{\alpha_0\beta}\right)\|\mathbf{f}\|_F + \frac{2M_a^2}{\alpha_0\beta^2}\|\mathbf{g}\|_G \\ &\le \frac{2M_a}{\alpha_0\beta}\|\mathbf{f}\|_F + \frac{2M_a^2}{\alpha_0\beta^2}\|\mathbf{g}\|_G \qquad (171)\end{aligned}$$

and the proof is concluded. □

A straightforward consequence of Theorem 2 is the following stability result (cf. (111)):

Corollary 2. *Assume that a numerical method produces a sequence of matrices* $\mathbf{A}$ *and* $\mathbf{B}$ *for which both the inf–sup condition (112) and the elker condition (159) are satisfied. Then the method is stable.*

Remark 5. In the spirit of Remark 3, we notice that the dependence of the stability constants from α_0 and β is optimal, as shown by the previous example (146), for which $\alpha_0 = a$ and $\beta = b$. It is interesting to notice that just adding the assumption that $\mathbf{A}$ is symmetric will not improve the bounds. Indeed, considering the system

$$\begin{bmatrix}1 & 1 & b\\ 1 & a & 0\\ b & 0 & 0\end{bmatrix}\begin{Bmatrix}x_1\\ x_2\\ y\end{Bmatrix} = \begin{Bmatrix}f_1\\ f_2\\ g\end{Bmatrix} \qquad 0 < a, b \ll 1 \quad (172)$$

one easily obtains

$$x_1 = \frac{g}{b}, \quad x_2 = \frac{f_2}{a} - \frac{g}{ab}, \quad y = \frac{f_1}{b} - \frac{f_2}{ab} + \frac{(1-a)g}{ab^2} \qquad (173)$$

Since $\alpha_0 = a$ and $\beta = b$, system (172) shows the same behavior as the bounds of Theorem 2 (and not better), even though $\mathbf{A}$ is symmetric. In order to get back the better bounds found in Theorem 1, we have to assume that $\mathbf{A}$, on top of satisfying the ellipticity in the kernel (159), is symmetric *and* positive semidefinite in the whole $\mathbb{R}^n$ (a property that the matrix $\mathbf{A}$ in (172) does not have for $0 \le a < 1$). This is because, in order to improve the bounds, one has to use (110) that, indeed, requires $\mathbf{A}$ to be symmetric and positive semidefinite.

As announced in the title of the section, the situation in which $\mathbf{A}$ is elliptic only in the kernel of $\mathbf{B}$ is typical (among others) of the mixed formulation of thermal problems, as presented in (22) to (24). As in (19), we denote the interpolating functions for θ and $\mathbf{q}$ by N_r^θ and $\mathbf{N}_i^{\mathbf{q}}$, respectively, and we set

$$\begin{aligned}\|\hat{\mathbf{q}}\|_X^2 :=& \int_\Omega \left[\left(\mathbf{N}_i^{\mathbf{q}}\hat{q}_i\right)\cdot\mathbf{D}^{-1}\left(\mathbf{N}_j^{\mathbf{q}}\hat{q}_j\right)\right]\mathrm{d}\Omega \\ &+ \frac{\ell^2}{k_*}\int_\Omega |\mathrm{div}\,(\mathbf{N}_i^{\mathbf{q}}\hat{q}_i)|^2\,\mathrm{d}\Omega \qquad (174)\end{aligned}$$

$$\|\hat{\theta}\|_Y^2 := \int_\Omega |N_r^\theta\hat{\theta}_r|^2\,\mathrm{d}\Omega \qquad (175)$$

where ℓ represents some characteristic length of the domain Ω (for instance its diameter) and k_* represent some characteristic value of the thermal conductivity (for instance, its average).

We can easily see that the continuity conditions (101) are verified with $M_a = 1$ and $M_b = \ell^{-1}\sqrt{k_*}$ respectively. On the other hand, the full ellipticity property (125) is verified only with a constant α that behaves, in most cases, like $\alpha \simeq h^2$, where h is a measure of the mesh size. Indeed, the norm of $\hat{\mathbf{q}}$ contains the derivatives of the interpolating functions, while the term $\hat{\mathbf{q}}^{\mathrm{T}}\mathbf{A}\hat{\mathbf{q}}$ does not, as it can be seen in (21). On the other hand, we are obliged to add the divergence term in the definition (174) of the norm of $\hat{\mathbf{q}}$: otherwise, we cannot have a uniform bound for M_b when the meshsize goes to zero, precisely for the same reason as before. Indeed, the term $\hat{\boldsymbol{\theta}}^{\mathrm{T}}\mathbf{B}\hat{\mathbf{q}}$ contains the derivatives of the interpolating functions $\mathbf{N}_i^{\mathbf{q}}$ (see (21)), and the first part of $\|\hat{\mathbf{q}}\|_X$ does not. One can object that the constant M_b does not show up in the stability estimates. It does, however, come into play in the *error estimates*, as we are going to see in Section 5.

It follows from this analysis that, keeping the norms as in (174) and (175), the *elker* property (159) holds, in practical cases, only if the kernel of $\mathbf{B}$ is made of free-divergence vectors. In that case, we would actually have $\alpha_0 = 1$, no matter what is the choice of the mesh and of the interpolating functions.

On the other hand, the *inf–sup* property (112) is still difficult and it depends heavily on the choices of the interpolating functions. As we are going to see in the next section, the need to satisfy both the *elker* and the *inf–sup* condition poses serious limitations on the choice of the approximations. Apart from some special one-dimensional cases, there is no hope that these two properties can hold at the same time unless the finite element spaces have been *designed* for that. However, this work has been already done and there are several families of finite element spaces that can be profitably used for these problems. We also note that the *elker* condition (or, more precisely, the requirement that the kernel of $\mathbf{B}$ is made only of free-divergence vectors) poses some difficulties in the choice of the element, but in most applications it constitutes a very desirable conservation property for the discrete solutions.

3.7 Perturbation of the problem — nearly incompressible elasticity

We now consider a possible variant of our general form (95). Namely, we assume that we have, together with the matrices $\mathbf{A}$ and $\mathbf{B}$, a third matrix $\mathbf{C}$, that we assume to be an $(m \times m)$ matrix, and we consider the general form

$$\begin{bmatrix} \mathbf{A} & \mathbf{B}^{\mathrm{T}} \\ \mathbf{B} & -\mathbf{C} \end{bmatrix} \begin{Bmatrix} \mathbf{x} \\ \mathbf{y} \end{Bmatrix} = \begin{Bmatrix} \mathbf{f} \\ \mathbf{g} \end{Bmatrix} \tag{176}$$

For simplicity, we assume that the matrix $\mathbf{C}$ is given by $\mathbf{C} = \varepsilon \mathbf{M}^y$, where the matrix $\mathbf{M}^y$ is attached to the norm $\|\cdot\|_Y$ as in (103). Clearly, the results will apply, almost unchanged, to a symmetric positive definite matrix having maximum and minimum eigenvalue of order ε. We have the following result.

Theorem 3. *Let* $\mathbf{x} \in \mathbf{X}$ *and* $\mathbf{y} \in \mathbf{Y}$ *satisfy the system*

$$\begin{cases} \mathbf{A}\mathbf{x} + \mathbf{B}^{\mathrm{T}}\mathbf{y} = \mathbf{f} \\ \mathbf{B}\mathbf{x} - \varepsilon\mathbf{M}^y\mathbf{y} = \mathbf{g} \end{cases} \tag{177}$$

Assume that $\mathbf{A}$ *is symmetric and positive semidefinite, and that the continuity condition (101), the dual norm assumptions (105), the inf–sup (112) and the elker condition (159) are satisfied. Then, we have*

$$\|\mathbf{x}\|_X \le \frac{\beta^2 + 4\varepsilon M_a}{\alpha_0\beta^2}\|\mathbf{f}\|_F + \frac{2M_a^{1/2}}{\alpha_0^{1/2}\beta}\|\mathbf{g}\|_G \tag{178}$$

and

$$\|\mathbf{y}\|_Y \le \frac{2M_a^{1/2}}{\alpha_0^{1/2}\beta}\|\mathbf{f}\|_F + \frac{4M_a}{M_a\varepsilon + \beta^2}\|\mathbf{g}\|_G \tag{179}$$

Proof. The proof can be performed with arguments similar to the ones used in the previous stability proofs, but using more technicalities. For simplicity, we are going to give only a sketch, treating separately the two cases $\mathbf{f} = \mathbf{0}$ and $\mathbf{g} = \mathbf{0}$.

- *The case* $\mathbf{f} = \mathbf{0}$. We set $\widetilde{\mathbf{x}} = \mathbf{L}(\mathbf{g} + \varepsilon\mathbf{M}^y\mathbf{y})$ and $\mathbf{x}_0 = \mathbf{x} - \widetilde{\mathbf{x}}$. Proceeding exactly as in the proof of Theorem 1 (*Step 4*), we obtain inequality (140):

$$\|\mathbf{y}\|_Y \le \frac{M_a^{1/2}}{\beta}(\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x})^{1/2} \tag{180}$$

Then, we multiply the first equation of (176) times $\mathbf{x}^{\mathrm{T}}$ and substitute the value of $\mathbf{y}$ obtained from the second equation. We have

$$\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} + \frac{1}{\varepsilon}\left[(\mathbf{M}^y)^{-1}(\mathbf{B}\mathbf{x} - \mathbf{g})\right]^{\mathrm{T}}\mathbf{B}\mathbf{x} = \mathbf{0} \tag{181}$$

Using the fact that $\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} > 0$, we easily deduce that

$$\|\mathbf{B}\mathbf{x}\|_G \le \|\mathbf{g}\|_G \tag{182}$$

This implies

$$\|\widetilde{\mathbf{x}}\|_X \le \frac{1}{\beta}\|\mathbf{B}\mathbf{x}\|_G \le \frac{1}{\beta}\|\mathbf{g}\|_G \tag{183}$$

We now multiply the first equation times $\mathbf{x}_0^{\mathrm{T}}$, and we have $\mathbf{x}_0^{\mathrm{T}}\mathbf{A}\mathbf{x} = 0$. We can then use (109) to get

$$\mathbf{x}_0^{\mathrm{T}}\mathbf{A}\mathbf{x}_0 = -\mathbf{x}_0^{\mathrm{T}}\mathbf{A}\widetilde{\mathbf{x}} \le (\mathbf{x}_0^{\mathrm{T}}\mathbf{A}\mathbf{x}_0)^{1/2}(\widetilde{\mathbf{x}}^{\mathrm{T}}\mathbf{A}\widetilde{\mathbf{x}})^{1/2} \tag{184}$$

Simplifying by $(\mathbf{x}_0^{\mathrm{T}}\mathbf{A}\mathbf{x}_0)^{1/2}$ and using (183), we obtain

$$\mathbf{x}_0^{\mathrm{T}}\mathbf{A}\mathbf{x}_0 \le \widetilde{\mathbf{x}}^{\mathrm{T}}\mathbf{A}\widetilde{\mathbf{x}} \le M_a\|\widetilde{\mathbf{x}}\|_X^2 \le \frac{M_a}{\beta^2}\|\mathbf{g}\|_G^2 \tag{185}$$

Using $\mathbf{x} = \mathbf{x}_0 + \widetilde{\mathbf{x}}$, and then again (109) and (183), we obtain

$$\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} \le \frac{4M_a}{\beta^2}\|\mathbf{g}\|_G^2 \tag{186}$$

that inserted in (180) gives an estimate for $\mathbf{y}$

$$\|\mathbf{y}\|_Y \le \frac{2M_a}{\beta^2}\|\mathbf{g}\|_G \tag{187}$$

On the other hand, using the *elker* condition (159), estimates (183), (185), and (160) we have

$$\|\mathbf{x}\|_X \le \|\mathbf{x}_0\|_X + \|\widetilde{\mathbf{x}}\|_X \le \left(\frac{M_a^{1/2}}{\alpha_0^{1/2}\beta} + \frac{1}{\beta}\right)\|\mathbf{g}\|_G$$
$$= \frac{M_a^{1/2} + \alpha_0^{1/2}}{\alpha_0^{1/2}\beta}\|\mathbf{g}\|_G \le \frac{2M_a^{1/2}}{\alpha_0^{1/2}\beta}\|\mathbf{g}\|_G \qquad (188)$$

However, we note that using the second equation we might have another possible estimate for $\mathbf{y}$:

$$\|\mathbf{y}\|_Y \le \frac{1}{\varepsilon}\|\mathbf{B}\mathbf{x} - \mathbf{g}\|_G \le \frac{2}{\varepsilon}\|\mathbf{g}\|_G \qquad (189)$$

We can combine (187) and (189) into

$$\|\mathbf{y}\|_Y \le \min\left\{\frac{2}{\varepsilon}, \frac{2M_a}{\beta^2}\right\}\|\mathbf{g}\|_G \le \frac{4M_a}{M_a\varepsilon + \beta^2}\|\mathbf{g}\|_G \qquad (190)$$

• *The case* $\mathbf{g} = \mathbf{0}$. We set this time $\widetilde{\mathbf{x}} = \mathbf{L}(\varepsilon\mathbf{M}^y\mathbf{y})$ and again $\mathbf{x}_0 := \mathbf{x} - \widetilde{\mathbf{x}}$. From (150), we have as usual

$$\|\widetilde{\mathbf{x}}\|_X \le \frac{1}{\beta}\|\mathbf{B}\widetilde{\mathbf{x}}\|_G \equiv \frac{1}{\beta}\|\mathbf{B}\mathbf{x}\|_G \qquad (191)$$

Multiplying the first equation by $\mathbf{x}_0^{\mathrm{T}}$, we have $\mathbf{x}_0^{\mathrm{T}}\mathbf{A}\mathbf{x} = \mathbf{x}_0\mathbf{f}$ that gives, using (159) and (109)

$$\mathbf{x}_0^{\mathrm{T}}\mathbf{A}\mathbf{x}_0 \le \frac{1}{\alpha_0^{1/2}}\|\mathbf{f}\|_F(\mathbf{x}_0^{\mathrm{T}}\mathbf{A}\mathbf{x}_0)^{1/2} + (\mathbf{x}_0^{\mathrm{T}}\mathbf{A}\mathbf{x}_0)^{1/2}(\widetilde{\mathbf{x}}^{\mathrm{T}}\mathbf{A}\widetilde{\mathbf{x}})^{1/2} \qquad (192)$$

and finally,

$$(\mathbf{x}_0^{\mathrm{T}}\mathbf{A}\mathbf{x}_0)^{1/2} \le \frac{1}{\alpha_0^{1/2}}\|\mathbf{f}\|_F + (\widetilde{\mathbf{x}}^{\mathrm{T}}\mathbf{A}\widetilde{\mathbf{x}})^{1/2} \qquad (193)$$

In particular, using once more, (109), (193), and (191), we obtain

$$|\mathbf{x}_0^{\mathrm{T}}\mathbf{A}\widetilde{\mathbf{x}}| \le \frac{1}{\alpha_0^{1/2}}\|\mathbf{f}\|_F(\widetilde{\mathbf{x}}^{\mathrm{T}}\mathbf{A}\widetilde{\mathbf{x}})^{1/2} + \widetilde{\mathbf{x}}^{\mathrm{T}}\mathbf{A}\widetilde{\mathbf{x}}$$
$$\le \frac{M_a^{1/2}}{\alpha_0^{1/2}\beta}\|\mathbf{f}\|_F\|\mathbf{B}\mathbf{x}\|_G + \widetilde{\mathbf{x}}^{\mathrm{T}}\mathbf{A}\widetilde{\mathbf{x}} \qquad (194)$$

Take now the product of the first equation times $\widetilde{\mathbf{x}}^{\mathrm{T}}$ and using $\mathbf{y} = \varepsilon^{-1}(\mathbf{M}^y)^{-1}\mathbf{B}\mathbf{x}$ from the second equation, we have $\widetilde{\mathbf{x}}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{y} = \varepsilon^{-1}\widetilde{\mathbf{x}}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}(\mathbf{M}^y)^{-1}\mathbf{B}\mathbf{x} = \varepsilon^{-1}\|\mathbf{B}\mathbf{x}\|_G^2$. Hence,

$$\widetilde{\mathbf{x}}^{\mathrm{T}}\mathbf{A}\mathbf{x} + \frac{1}{\varepsilon}\|\mathbf{B}\mathbf{x}\|_G^2 = \widetilde{\mathbf{x}}^{\mathrm{T}}\mathbf{f} \le \frac{1}{\beta}\|\mathbf{f}\|_F\|\mathbf{B}\mathbf{x}\|_G \qquad (195)$$

Using $\widetilde{\mathbf{x}}^{\mathrm{T}}\mathbf{A}\mathbf{x} = \widetilde{\mathbf{x}}^{\mathrm{T}}\mathbf{A}\widetilde{\mathbf{x}} + \widetilde{\mathbf{x}}^{\mathrm{T}}\mathbf{A}\mathbf{x}_0$ and the estimate (194) in (195), we deduce

$$\frac{1}{\varepsilon}\|\mathbf{B}\mathbf{x}\|_G^2 \le \frac{1}{\beta}\|\mathbf{f}\|_F\|\mathbf{B}\mathbf{x}\|_G + \frac{M_a^{1/2}}{\alpha_0^{1/2}\beta}\|\mathbf{f}\|_F\|\mathbf{B}\mathbf{x}\|_G \qquad (196)$$

that finally gives

$$\|\mathbf{B}\mathbf{x}\|_G \le \varepsilon\left(\frac{1}{\beta} + \frac{M_a^{1/2}}{\alpha_0^{1/2}\beta}\right)\|\mathbf{f}\|_F \le \frac{2\varepsilon M_a^{1/2}}{\alpha_0^{1/2}\beta}\|\mathbf{f}\|_F \qquad (197)$$

which is a crucial step in our proof. Indeed, from (197) and the second equation, we obtain our estimate for $\mathbf{y}$

$$\|\mathbf{y}\|_Y \le \frac{1}{\varepsilon}\|\mathbf{B}\mathbf{x}\|_G \le \frac{2M_a^{1/2}}{\alpha_0^{1/2}\beta}\|\mathbf{f}\|_F \qquad (198)$$

From (191) and (197), we have

$$\|\widetilde{\mathbf{x}}\|_X \le \frac{1}{\beta}\|\mathbf{B}\mathbf{x}\|_G \le \frac{2\varepsilon M_a^{1/2}}{\alpha_0^{1/2}\beta^2}\|\mathbf{f}\|_F \qquad (199)$$

Finally, from (159), (193), and (199), we obtain

$$\|\mathbf{x}_0\|_X \le \frac{1}{\alpha_0^{1/2}}(\mathbf{x}_0^{\mathrm{T}}\mathbf{A}\mathbf{x}_0)^{1/2} \le \left(\frac{1}{\alpha_0} + \frac{2\varepsilon M_a}{\alpha_0\beta^2}\right)\|\mathbf{f}\|_F$$
$$= \frac{\beta^2 + 2\varepsilon M_a}{\alpha_0\beta^2}\|\mathbf{f}\|_F \qquad (200)$$

which together with (199) gives us the estimate for $\mathbf{x}$

$$\|\mathbf{x}\|_X \le \left(\frac{2\varepsilon M_a^{1/2}}{\alpha_0^{1/2}\beta^2} + \frac{2\varepsilon M_a + \beta^2}{\alpha_0\beta^2}\right)\|\mathbf{f}\|_F \le \frac{4\varepsilon M_a + \beta^2}{\alpha_0\beta^2}\|\mathbf{f}\|_F \qquad (201)$$

Collecting (190), (188), (198), and (201), we have the result. □

Remark 6. We notice that the dependence of the stability constants upon α_0 and β in Theorem 3 are optimal, as shown by the system

$$\begin{bmatrix} 2a & \sqrt{a} & -\sqrt{a} & 0 & 0 \\ \sqrt{a} & 2 & 1 & b & 0 \\ -\sqrt{a} & 1 & 2 & 0 & b \\ 0 & b & 0 & -\varepsilon & 0 \\ 0 & 0 & b & 0 & -\varepsilon \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \\ x_3 \\ y_1 \\ y_2 \end{Bmatrix} = \begin{Bmatrix} 2f \\ 0 \\ 0 \\ 0 \\ 2g \end{Bmatrix}$$
$$0 < a, b, \varepsilon \ll 1 \qquad (202)$$

Indeed, we have $\alpha_0 = 2a$, $\beta = b$, and the solution is given by

$$x_1 = \frac{f(b^2 + \varepsilon)}{ab^2} + \frac{g}{a^{1/2}b}, \quad x_2 = -\frac{f\varepsilon}{a^{1/2}b^2} - \frac{3g\varepsilon}{b(3\varepsilon + b^2)},$$

$$x_3 = \frac{f\varepsilon}{a^{1/2}b^2} + \frac{g(3\varepsilon + 2b^2)}{b(3\varepsilon + b^2)},$$
$$y_1 = -\frac{f}{a^{1/2}b} - \frac{3g}{3\varepsilon + b^2}, \quad y_2 = \frac{f}{a^{1/2}b} - \frac{3g}{3\varepsilon + b^2}$$

Remark 7. It is also worth noticing that assuming full ellipticity of the matrix $\mathbf{A}$ as in (125) (instead of ellipticity only in the kernel as we did here) would improve the estimates for $\mathbf{x}$. In particular, we could obtain estimates that do not degenerate when β goes to zero, as far as ε remains strictly positive. For the case $\mathbf{f} = 0$, this is immediate from the estimate of $\mathbf{y}$ (190): from the first equation, we have easily

$$\|\mathbf{x}\| \le \frac{1}{\alpha} M_b \|\mathbf{y}\|_Y \le \frac{4M_a M_b}{\alpha(M_a\varepsilon + \beta^2)} \|\mathbf{g}\|_G \tag{203}$$

In the case $\mathbf{g} = \mathbf{0}$, we can combine the two equations to get

$$\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} + \varepsilon\|\mathbf{y}\|_Y^2 = \mathbf{x}^{\mathrm{T}}\mathbf{f} \tag{204}$$

that gives (always using (125))

$$\|\mathbf{x}\|_X \le \frac{1}{\alpha}\|\mathbf{f}\|_F \tag{205}$$

that then gives

$$\|\mathbf{y}\|_Y \le \frac{1}{\varepsilon}\|\mathbf{B}\mathbf{x}\|_G \le \frac{M_b}{\varepsilon\alpha}\|\mathbf{f}\|_F \tag{206}$$

This could be combined with (198) into

$$\begin{aligned}\|\mathbf{y}\|_Y &\le \min\left\{\frac{M_b}{\varepsilon\alpha}, \frac{2M_a^{1/2}}{\alpha^{1/2}\beta}\right\}\|\mathbf{f}\|_F \\ &\le \frac{4M_a^{1/2}M_b}{2M_a^{1/2}\alpha\varepsilon + \alpha^{1/2}\beta M_b}\|\mathbf{f}\|_F\end{aligned} \tag{207}$$

Collecting the two cases we have

$$\|\mathbf{x}\|_X \le \frac{1}{\alpha}\|\mathbf{f}\|_F + \frac{4M_a M_b}{\alpha^{1/2}(M_a\varepsilon + \beta^2)}\|\mathbf{g}\|_G \tag{208}$$

and

$$\|\mathbf{y}\|_Y \le \frac{4M_a^{1/2}M_b}{2M_a^{1/2}\alpha\varepsilon + \alpha^{1/2}\beta M_b}\|\mathbf{f}\|_F + \frac{4M_a}{M_a\varepsilon + \beta^2}\|\mathbf{g}\|_G \tag{209}$$

which do not degenerate for β going to zero.

As announced in the title of the section, systems of the type (176) occur, for instance, in the so-called $(\mathbf{u}, \pi)$ formulation of nearly incompressible elasticity. Sometimes they are also obtained by penalizing systems of the original type (95) in order to obtain a partial cure in cases in which β is zero or tending to zero with the meshsize (as it could happen, for instance, for a discretization of Stokes problem that does not satisfy the *inf–sup* condition), in the spirit of Remark 7. Indeed, the $(\mathbf{u}, \pi)$ formulation of nearly incompressible elasticity, in the case of an isotropic and homogeneous body, could be seen, mathematically, as a perturbation of the Stokes system with $\varepsilon = 1/\lambda$, and the elements to be used are essentially the same.

3.8 Composite matrices

In a certain number of applications, one has to deal with formulations of mixed type where more than two fields are involved. These give rise to matrices that are naturally split as 3 × 3 or 4 × 4 (or more) block matrices. For the sake of completeness, we show how the previous theory can often apply almost immediately to these more general cases. As an example, we consider matrices of the type

$$\begin{bmatrix} \mathbf{A} & \mathbf{B}^{\mathrm{T}} & \mathbf{0} \\ \mathbf{B} & \mathbf{0} & \mathbf{C}^{\mathrm{T}} \\ \mathbf{0} & \mathbf{C} & \mathbf{0} \end{bmatrix} \begin{Bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{y}_1 \end{Bmatrix} = \begin{Bmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \\ \mathbf{g} \end{Bmatrix} \tag{210}$$

Matrices of the form (210) are found (among several other applications) in the discretization of formulations of Hu–Washizu type. However, in particular, for elasticity problems, there are no good examples of finite element discretizations of the Hu–Washizu principle that satisfy the following two requirements at the same time: not reducing more or less immediately (in the linear case) to known discretizations of the minimum potential energy or of the Hellinger–Reissner principle, and having been proved to be stable and optimally convergent in a sound mathematical way. Actually, the only way, so far, has been using *stabilized formulations* (see for instance Behr, Franca and Tezduyar (1993)) that we decided to avoid here. Still, we hope that the following brief discussion could also be useful for the possible development of good Hu–Washizu elements in the future.

Coming back to the analysis of (210), we already observed that systems of this type can be reconduced to the general form (95)

$$\begin{bmatrix} \mathbb{A} & \mathbb{B}^{\mathrm{T}} \\ \mathbb{B} & \mathbf{0} \end{bmatrix} \begin{Bmatrix} \mathbf{x} \\ \mathbf{y} \end{Bmatrix} = \begin{Bmatrix} \mathbf{f} \\ \mathbf{g} \end{Bmatrix} \tag{211}$$

after making the following simple identifications:

$$\mathbb{A} = \begin{bmatrix} \mathbf{A} & \mathbf{B}^{\mathrm{T}} \\ \mathbf{B} & \mathbf{0} \end{bmatrix}, \quad \mathbb{B} = \{\mathbf{0}, \mathbf{C}\}$$
$$\mathbf{x} = \begin{Bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{Bmatrix}, \quad \mathbf{y} = \mathbf{y}_1 \tag{212}$$

The stability of system (211) can then be studied using the previous analysis. Sometimes it is, however, more convenient to reach the compact form (211) with a different identification:

$$\mathbb{A} = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad \mathbb{B} = \{\mathbf{B}, \mathbf{C}^{\mathrm{T}}\}$$

$$\mathbf{x} = \begin{Bmatrix} \mathbf{x}_1 \\ \mathbf{y}_1 \end{Bmatrix}, \quad \mathbf{y} = \mathbf{x}_2 \tag{213}$$

Indeed, in this case, the matrix $\mathbb{A}$ is much simpler. In particular, as it happens quite often in practice, when the original matrix $\mathbf{A}$ in (210) is symmetric and positive semidefinite, the same properties will be shared by $\mathbb{A}$. We are not going to repeat the theory of the above sections for the extended systems (210). We will just point out the meaning of conditions *elker* and *inf–sup*, applied to the system (212) to (213), in terms of the original matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$.

The kernel of $\mathbb{B}$, as given in (213), is made of the pairs $(\mathbf{x}_1, \mathbf{y}_1)$ such that

$$\mathbf{B}\mathbf{x}_1 + \mathbf{C}^{\mathrm{T}}\mathbf{y}_1 = \mathbf{0} \tag{214}$$

These include, in particular, all the pairs $(\mathbf{0}, \mathbf{y}_1)$, where $\mathbf{y}_1$ is in the kernel of $\mathbf{C}^{\mathrm{T}}$:

$$\mathrm{Ker}(\mathbf{C}^{\mathrm{T}}) := \{\mathbf{y}_1 | \text{ such that } \mathbf{C}^{\mathrm{T}}\mathbf{y}_1 = \mathbf{0}\} \tag{215}$$

There is no hope that the matrix $\mathbb{A}$, as defined in (213), can be elliptic on those pairs. Hence, we must require that those pairs are actually reduced to the pair $(\mathbf{0}, \mathbf{0})$, that is, we must require that

$$\mathrm{Ker}(\mathbf{C}^{\mathrm{T}}) = \{\mathbf{0}\} \tag{216}$$

This does not settle the matter of *elker*, since there are many other pairs $(\mathbf{x}_1, \mathbf{y}_1)$ satisfying (214). As $\mathbb{A}$ acts only on the $\mathbf{x}_1$ variables, we must characterize the vectors $\mathbf{x}_1$ such that $(\mathbf{x}_1, \mathbf{y}_1)$ satisfies (214) for some $\mathbf{y}_1$. These are

$$\mathcal{K} := \{\mathbf{x}_1 | \text{ such that } \mathbf{z}^{\mathrm{T}}\mathbf{B}\mathbf{x}_1 = 0 \quad \forall\, \mathbf{z} \in \mathrm{Ker}(\mathbf{C})\} \tag{217}$$

Hence we have the following result: condition *elker* will hold, for the system (212) to (213) if and only if

$$\exists \widetilde{\alpha} > 0 \text{ such that } \widetilde{\alpha}\|\mathbf{x}_1\|^2 \le \mathbf{x}_1^{\mathrm{T}}\mathbf{A}\mathbf{x}_1 \quad \forall\, \mathbf{x}_1 \in \mathcal{K} \tag{218}$$

On the other hand, it is not difficult to see that condition *inf–sup* for (212) to (213) reads

$$\exists \widetilde{\beta} > 0 \text{ such that } \sup_{(\mathbf{x}_1, \mathbf{y}_1)} \frac{\mathbf{x}_2^{\mathrm{T}}\mathbf{B}\mathbf{x}_1 + \mathbf{x}_2^{\mathrm{T}}\mathbf{C}^{\mathrm{T}}\mathbf{y}_1}{\|\mathbf{x}_1\| + \|\mathbf{y}_1\|} \ge \widetilde{\beta}\|\mathbf{x}_2\| \quad \forall\, \mathbf{x}_2 \tag{219}$$

It is clear that a *sufficient* condition would be to have the *inf–sup* condition to hold for at least one of the two matrices $\mathbf{B}$, $\mathbf{C}^{\mathrm{T}}$. In many applications, however, this is too strong a requirement. A weaker condition (although stronger than (219)) can be written as

$$\exists \widetilde{\beta} > 0 \text{ such that } \widetilde{\beta}\|\mathbf{x}_2\| \le \|\mathbf{C}\mathbf{x}_2\| + \|\mathbf{B}^{\mathrm{T}}\mathbf{x}_2\| \qquad \forall\, \mathbf{x}_2 \tag{220}$$

More generally, many variations are possible, according to the actual structure of the matrices at play.

4 APPLICATIONS

In this section, we give several examples of efficient mixed finite element methods, focusing our attention mostly on the thermal problem (Section 4.1) and on the Stokes equation (Section 4.2). For simplicity, we mainly consider triangular elements, while we briefly discuss their possible extensions to quadrilateral geometries and to three-dimensional cases. Regarding Stokes equation, we point out (as already mentioned) that the same discretization spaces can be profitably used to treat the nearly incompressible elasticity problem, within the context of the $(\mathbf{u}, \pi)$ formulation (80). We also address a brief discussion on elements for the elasticity problem in the framework of the Hellinger–Reissner principle (Section 4.3).

We finally remark that, for all the schemes that we are going to present, a rigorous stability and convergence analysis has been established, even though we will not detail the proofs.

4.1 Thermal diffusion

We consider the thermal diffusion problem described in Section 2.1 in the framework of the Hellinger–Reissner variational principle. We recall that the discretization of such a problem leads to solve the following algebraic system:

$$\begin{bmatrix} \mathbf{A} & \mathbf{B}^{\mathrm{T}} \\ \mathbf{B} & \mathbf{0} \end{bmatrix} \begin{Bmatrix} \hat{\mathbf{q}} \\ \hat{\boldsymbol{\theta}} \end{Bmatrix} = \begin{Bmatrix} \mathbf{0} \\ \mathbf{g} \end{Bmatrix} \tag{221}$$

where

$$\begin{cases} \mathbf{A}|_{ij} = \displaystyle\int_\Omega \left[\mathbf{N}_i^{\mathbf{q}} \cdot \mathbf{D}^{-1}\mathbf{N}_j^{\mathbf{q}}\right] \mathrm{d}\Omega, \quad \hat{\mathbf{q}}|_i = \hat{q}_i \\ \mathbf{B}|_{rj} = -\displaystyle\int_\Omega \left[N_r^\theta \,\mathrm{div}\left(\mathbf{N}_j^{\mathbf{q}}\right)\right] \mathrm{d}\Omega, \quad \hat{\boldsymbol{\theta}}|_r = \hat{\theta}_r \\ \mathbf{g}|_r = \displaystyle\int_\Omega \left[N_r^\theta b\right] \mathrm{d}\Omega \end{cases} \tag{222}$$

Above, $\mathbf{N}_i^{\mathbf{q}}$ and N_r^θ are the interpolation functions for the flux $\mathbf{q}$ and the temperature θ respectively. Moreover, $\hat{\mathbf{q}}$ and $\hat{\boldsymbol{\theta}}$ are the vectors of flux and temperature unknowns, while $i, j = 1, \ldots, n$ and $r = 1, \ldots, m$, where n and m obviously depend on the chosen approximation spaces as well as on the mesh.

Following the notation of the previous section, the norms for which the *inf–sup* and the *elker* conditions should be checked are (cf. (174) and (175))

$$\|\hat{\mathbf{q}}\|_X^2 := \int_\Omega \left[\left(\mathbf{N}_i^{\mathbf{q}}\hat{q}_i\right) \cdot \mathbf{D}^{-1} \left(\mathbf{N}_j^{\mathbf{q}}\hat{q}_j\right) \right] \mathrm{d}\Omega + \frac{\ell^2}{k_*} \int_\Omega |\mathrm{div}\,(\mathbf{N}_i^{\mathbf{q}}\hat{q}_i)|^2 \,\mathrm{d}\Omega \qquad (223)$$

and

$$\|\hat{\boldsymbol{\theta}}\|_Y^2 := \int_\Omega |N_r^\theta \hat{\theta}_r|^2 \,\mathrm{d}\Omega \qquad (224)$$

where ℓ is some characteristic length of the domain Ω and k_* is some characteristic value of the thermal conductivity.

Before proceeding, we remark the following:

- Since no derivative operator acts on the interpolating functions N_r^θ in the matrix $\mathbf{B}$, we are allowed to approximate the temperature θ without requiring any continuity across the elements. On the contrary, the presence of the divergence operator acting on the interpolating functions $\mathbf{N}_i^{\mathbf{q}}$ in the matrix $\mathbf{B}$ suggests that the normal component of the approximated flux should not exhibit jumps between adjacent elements.
- The full ellipticity for $\mathbf{A}$ (i.e. property (125)) typically holds only with a constant $\alpha \simeq h^2$, once the norm (223) has been chosen. However, if a method is designed in such a way that

  $$\hat{\mathbf{q}}_0 = (\hat{q}_i^0)_{i=1}^n \in \mathrm{Ker}(\mathbf{B}) \ \text{ implies } \ \mathrm{div}\,(\mathbf{N}_i^{\mathbf{q}}\hat{q}_i^0) = 0 \qquad (225)$$

  the weaker *elker* condition (159) obviously holds with $\alpha_0 = 1$.
  Condition (225) is verified if, for instance, we insist that

  $$\mathrm{Span}\{\mathrm{div}\,\mathbf{N}_i^{\mathbf{q}};\ i = 1, \ldots, n\} \subseteq \mathrm{Span}\{N_r^\theta;\ r = 1, \ldots, m\} \qquad (226)$$

  that is, the divergences of all the approximated fluxes are contained in the space of the approximated temperatures. Indeed, condition (226) implies that, for every $\hat{\mathbf{q}}_0 \in \mathrm{Ker}(\mathbf{B})$, there exists $\hat{\boldsymbol{\theta}}_0 = (\hat{\theta}_r^0)_{r=1}^m$ such that $\mathrm{div}\,(\mathbf{N}_i^{\mathbf{q}}\hat{q}_i^0) = -N_r^\theta\hat{\theta}_r^0$. It follows that

  $$0 = \hat{\boldsymbol{\theta}}_0^{\mathrm{T}}\mathbf{B}\hat{\mathbf{q}}_0 = -\int_\Omega (N_r^\theta\hat{\theta}_r^0)\mathrm{div}\,(\mathbf{N}_i^{\mathbf{q}}\hat{q}_i^0)\,\mathrm{d}\Omega = \int_\Omega |\mathrm{div}\,(\mathbf{N}_i^{\mathbf{q}}\hat{q}_i^0)|^2 \,\mathrm{d}\Omega \qquad (227)$$

  so that $\mathrm{div}\,(\mathbf{N}_i^{\mathbf{q}}\hat{q}_i^0) = 0$.
  Condition (226) can be always achieved by 'enriching' the temperature approximation, if necessary. However, we remark that a careless enlargement of the approximated temperatures can compromise the fulfillment of the *inf–sup* condition (112), as shown in the following easy result.

  Proposition 3. *Suppose that a given method satisfies condition (226). Then the inf–sup condition (112) implies*

  $$\mathrm{Span}\{\mathrm{div}\,\mathbf{N}_i^{\mathbf{q}};\ i = 1, \ldots, n\} \equiv \mathrm{Span}\{N_r^\theta;\ r = 1, \ldots, m\} \qquad (228)$$

  that is, the divergences of all the approximated fluxes **coincide with** *the space of the approximated temperatures.*

  Proof. By contradiction, suppose that $\mathrm{Span}\{\mathrm{div}\,\mathbf{N}_i^{\mathbf{q}};\ i = 1, \ldots, n\}$ is *strictly* contained in $\mathrm{Span}\{N_r^\theta;\ r = 1, \ldots, m\}$. It follows that there exists $\hat{\boldsymbol{\theta}}_\perp = (\hat{\theta}_r^\perp)_{r=1}^m \in \mathbb{R}^m \backslash \{0\}$ such that

  $$\hat{\mathbf{q}}_*^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\hat{\boldsymbol{\theta}}_\perp = -\int_\Omega (N_r^\theta\hat{\theta}_r^\perp)\mathrm{div}\,(\mathbf{N}_i^{\mathbf{q}}\hat{q}_i^*)\,\mathrm{d}\Omega = 0\ \forall\ \hat{\mathbf{q}}_* \in \mathbb{R}^n \qquad (229)$$

  Therefore,

  $$\sup_{\hat{\mathbf{q}}_* \in \mathbb{R}^n \backslash \{0\}} \frac{\hat{\mathbf{q}}_*^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\hat{\boldsymbol{\theta}}_\perp}{\|\hat{\mathbf{q}}_*\|_X} = 0 \qquad (230)$$

  and the *inf–sup* condition does not hold (cf. (119)).□

  We also remark that the converse of Proposition 3 does not hold, that is, condition (228) is not sufficient for the fulfillment of *inf–sup* (although it does imply *elker*).

From the considerations above, it should be clear that

- degrees of freedom associated with the **normal component** of the approximated flux are needed to guarantee its continuity across adjacent elements;
- the satisfaction of both the *elker* and the *inf–sup* condition requires a careful and well-balanced choice of the interpolating fields.

In the following, we are going to present several elements designed accordingly to the guidelines above, all satisfying property (228).

4.1.1 Triangular elements

Throughout this section, we will always suppose that the domain $\Omega \subset \mathbb{R}^2$, on which the thermal problem is posed, is decomposed by means of a *triangular* mesh $\mathcal{T}_h$ with meshsize h. Moreover, we define $\mathcal{E}_h$ as the set of all the edges of the triangles in $\mathcal{T}_h$.

• *The $RT_0 - P_0$ element.* We now introduce the simplest triangular element proposed for thermal problems. For the discretization of the thermal flux $\mathbf{q}$, we take the so-called *lowest-order Raviart–Thomas element* (RT_0 element), presented in (Raviart and Thomas, 1977); accordingly, the approximated flux $\mathbf{q}^h$ is described as a *piecewise linear* (vectorial) field such that

i. the normal component $\mathbf{q}^h \cdot \mathbf{n}$ is constant on each edge e of $\mathcal{E}_h$;
ii. the normal component $\mathbf{q}^h \cdot \mathbf{n}$ is continuous across each edge e of $\mathcal{E}_h$.

To approximate the temperature, we simply use *piecewise constant functions* in each element (P_0 element).

On the generic triangle $T \in \mathcal{T}_h$, a set of element degrees of freedom for $\mathbf{q}^h$ is given by its 3 normal fluxes on the edges of the triangle, that is,

$$\int_e \mathbf{q}^h \cdot \mathbf{n}\, \mathrm{d}s \qquad \forall\ e \text{ edge of } T \tag{231}$$

Therefore, the space for the element approximation of $\mathbf{q}$ has dimension 3 and a basis is obtained by considering the (vectorial) shape functions

$$\mathbf{N}_k^{\mathbf{q}} = \mathbf{N}_k^{\mathbf{q}}(x, y) = \frac{1}{2\mathrm{Area}(T)} \begin{Bmatrix} x - x_k \\ y - y_k \end{Bmatrix} \qquad k = 1, 2, 3 \tag{232}$$

Above, $\{x_k, y_k\}^{\mathrm{T}}$ denotes the position vector of the kth vertex (local numbering) of the triangle T.

We also remark that, because of (232), $\mathbf{q}^h$ can be locally described by

$$\mathbf{q}^h = \mathbf{p}_0 + p_0 \begin{Bmatrix} x \\ y \end{Bmatrix} = \begin{Bmatrix} a_0 + p_0 x \\ b_0 + p_0 y \end{Bmatrix} \tag{233}$$

where $a_0, b_0, p_0 \in \mathbb{R}$.

As far as the approximated temperature is concerned, an element basis for θ^h is given by the shape function

$$N^{\theta} = N^{\theta}(x, y) = 1 \tag{234}$$

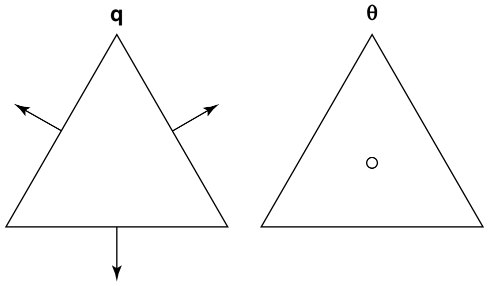

Figure 1. Degrees of freedom for RT_0–P_0 element.

The element degrees of freedom for both $\mathbf{q}^h$ and θ^h are schematically depicted in Figure 1.

• *The $RT_k - P_k$ family.* We now present the extension to higher orders of the $RT_0 - P_0$ method just described (cf. Raviart and Thomas, 1977). Given an integer $k \geq 1$ and using the definition introduced in Nedelec (1980), for the flux $\mathbf{q}^h$, we take a field such that (RT_k element) on each triangle T of $\mathcal{T}_h$, we have

$$\mathbf{q}^h = \mathbf{p}_k(x, y) + p_k(x, y) \begin{Bmatrix} x \\ y \end{Bmatrix} \tag{235}$$

where $\mathbf{p}_k(x, y)$ (resp. $p_k(x, y)$) is a vectorial (resp. scalar) polynomial of degree at most k. Moreover, we require that the *normal component* $\mathbf{q}^h \cdot \mathbf{n}$ is *continuous* across each edge e of $\mathcal{E}_h$. This can be achieved by selecting the following element degrees of freedom:

i. the moments of order up to k of $\mathbf{q}^h \cdot \mathbf{n}$ on the edges of T;
ii. the moments of order up to $k - 1$ of $\mathbf{q}^h$ on T.

For the discretized temperature θ^h, we take piecewise polynomials of degree at most k (P_k element).

The element degrees of freedom for the choice $k = 1$ are shown in Figure 2.

• *The $BDM_1 - P_0$ element.* Another method, widely used to treat the thermal diffusion problem, arises from the approximation of the flux $\mathbf{q}$ by means of the so-called

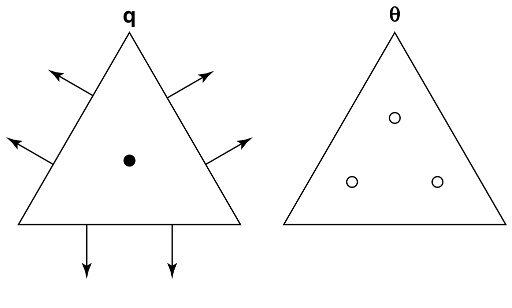

Figure 2. Degrees of freedom for RT_1–P_1 element.

lowest-order *Brezzi–Douglas–Marini* element (BDM_1 element), proposed and analyzed in Brezzi *et al.* (1985). It consists in discretizing $\mathbf{q}$ by vectorial functions $\mathbf{q}^h$ such that

i. $\mathbf{q}^h$ is linear in triangle T of $\mathcal{T}_h$;
ii. the normal component $\mathbf{q}^h \cdot \mathbf{n}$ is continuous across each edge e of $\mathcal{E}_h$.

For the approximated temperature θ^h, we again use piecewise constant functions on each triangle.

Focusing on the generic triangle $T \in \mathcal{T}_h$, we remark that the approximation space for $\mathbf{q}$ has dimension 6, since full linear polynomials are employed. A suitable set of element degrees of freedom is provided by the moments up to order 1 of the normal fluxes $\mathbf{q}^h \cdot \mathbf{n}$ across each edge e of T, explicitly given by the values

$$\left\{ \begin{array}{l} \displaystyle\int_e \mathbf{q}^h \cdot \mathbf{n}\, \mathrm{d}s \\ \displaystyle\int_e s\mathbf{q}^h \cdot \mathbf{n}\, \mathrm{d}s \end{array} \right. \tag{236}$$

where s is a local coordinate on e ranging from -1 to 1.

The element degrees of freedom for the resulting method are shown in Figure 3.

• *The $BDM_{k+1} - P_k$ family.* As for the RT_0–P_0 scheme, also the BDM_1–P_0 finite element method is the lowest order representative of a whole class. Indeed, given an integer $k \geq 1$, we can select the approximations presented in Brezzi *et al.* (1985).

For the discretized flux $\mathbf{q}^h$, the normal component $\mathbf{q}^h \cdot \mathbf{n}$ is *continuous* across each edge e of $\mathcal{E}_h$. Moreover, $\mathbf{q}^h$ is a vectorial polynomial of degree at most $k+1$ on each triangle T of $\mathcal{T}_h$ (BDM_{k+1} element). Also, in this case, the continuity of the normal component can be obtained by a proper choice of the degrees of freedom.

For the approximated temperature θ^h, we use the discontinuous P_k element. Figure 4 shows the element degrees of freedom for the case $k = 1$.

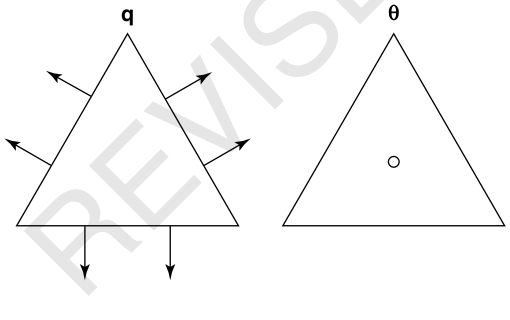

Figure 3. Degrees of freedom for BDM_1–P_0 element.

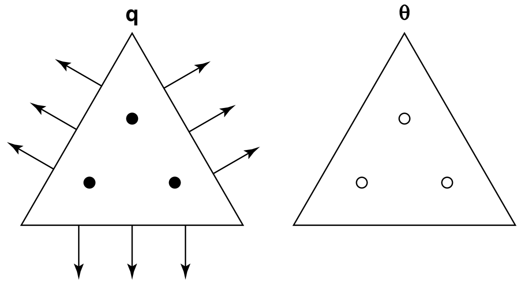

Figure 4. Degrees of freedom for BDM_2–P_1 element.

4.1.2 Quadrilateral elements

We now briefly consider the extension of some of the methods presented in the previous section to quadrilateral meshes. In this case, we define our approximating spaces on the *reference element* $\widetilde{K} = [-1, 1]^2$ equipped with local coordinates (ξ, η). As far as the flux is concerned, the corresponding approximation space on each physical element K must be obtained through the use of a suitable transformation that preserves the normal component of vectorial functions. This is accomplished by the following (contravariant) *Piola's transformation* of vector fields. Suppose that

$$\mathbf{F}\colon \widetilde{K} \longrightarrow K; \qquad (x, y) = \mathbf{F}(\xi, \eta)$$

is an invertible map from $\widetilde{K}$ onto K, with Jacobian matrix $\mathbf{J}(\xi, \eta)$. Given a vector field $\mathbf{q} = \mathbf{q}(\xi, \eta)$ on $\widetilde{K}$, its *Piola's transform* $\mathcal{P}(\mathbf{q}) = \mathcal{P}(\mathbf{q})(x, y)$ is the vector field on K, defined by

$$\mathcal{P}(\mathbf{q})(x, y) := \frac{1}{J(\xi, \eta)} \mathbf{J}(\xi, \eta)\mathbf{q}(\xi, \eta); \quad (x, y) = \mathbf{F}(\xi, \eta)$$

where $J(\xi, \eta) = |\det \mathbf{J}(\xi, \eta)|$. Therefore, if

$$\mathbf{Q}(\widetilde{K}) = \mathrm{Span}\{\mathbf{q}_i^h;\ i = 1, \ldots, n_{el}\}$$

is an n_{el}-dimensional flux approximation space defined on the reference element $\widetilde{K}$, the corresponding space on the physical element K will be

$$\mathbf{Q}(K) = \mathrm{Span}\{\mathcal{P}(\mathbf{q}_i^h);\ i = 1, \ldots, n_{el}\}$$

• *The $RT_{[0]} - P_0$ element.* In the reference element $\widetilde{K}$, we prescribe the approximated flux $\mathbf{q}^h$ as ($RT_{[0]}$ element)

$$\mathbf{q}^h = \left\{ \begin{array}{c} a + b\xi \\ c + d\eta \end{array} \right\}, \qquad a, b, c, d \in \mathbb{R} \tag{237}$$

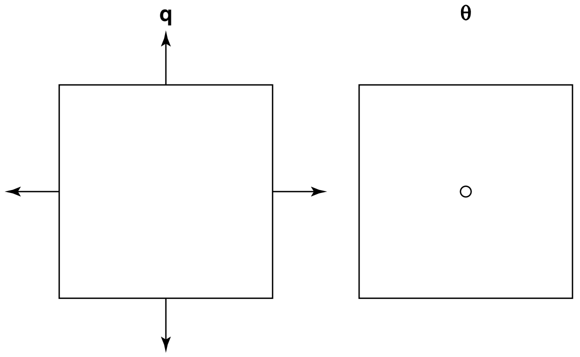

Figure 5. Degrees of freedom for $RT_{[0]}-P_0$ element.

Because of (237), it is easily seen that the four values

$$\int_e \mathbf{q}^h \cdot \mathbf{n}\, \mathrm{d}s \qquad \forall\, e \text{ edge of } \widetilde{K} \tag{238}$$

can be chosen as a set of degrees of freedom. Moreover, $\operatorname{div} \mathbf{q}^h$ is constant in $\widetilde{K}$, suggesting the choice of a constant approximated temperature θ^h in $\widetilde{K}$ (P_0 element). The degrees of freedom for both $\mathbf{q}^h$ and θ^h are shown in Figure 5.

- *The $BDM_{[1]} - P_0$ element*. For the discrete flux $\mathbf{q}^h$ on $\widetilde{K}$, we take a field such that ($BDM_{[1]}$ element)

$$\begin{aligned}\mathbf{q}^h &= \mathbf{p}_1(\xi,\eta) + a\begin{Bmatrix}\xi^2\\ -2\xi\eta\end{Bmatrix} + b\begin{Bmatrix}2\xi\eta\\ -\eta^2\end{Bmatrix}\\ &= \mathbf{p}_1(\xi,\eta) + a(\nabla(\xi^2\eta))^{\perp} + b(\nabla(\xi\eta^2))^{\perp}\end{aligned} \tag{239}$$

Above, $\mathbf{p}_1(\xi,\eta)$ is a vectorial linear polynomial, and a, b are real numbers. This space is carefully designed in order to have

i. $\mathbf{q}^h \cdot \mathbf{n}$ linear on each edge e of $\widetilde{K}$.
ii. $\operatorname{div} \mathbf{q}^h$ constant in $\widetilde{K}$.

Again, for the approximated temperature θ^h, we take constant functions (P_0 element). The element degrees of freedom for both $\mathbf{q}^h$ and θ^h are shown in Figure 6.

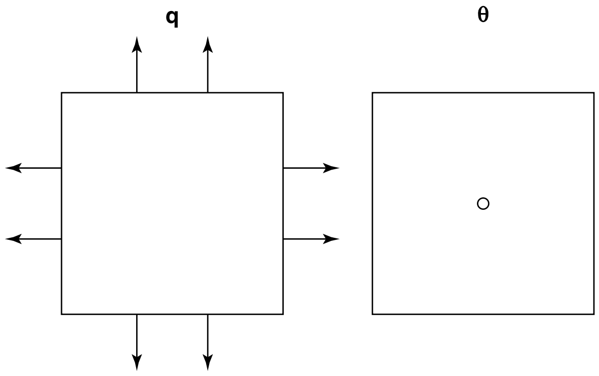

Figure 6. Degrees of freedom for $BDM_{[1]}-P_0$ element.

4.1.3 Three-dimensional elements

All the elements presented above have their three-dimensional counterpart. In this case, the normal component of the approximated flux $\mathbf{q}^h$ should not exhibit jumps across *faces* of adjacent elements.

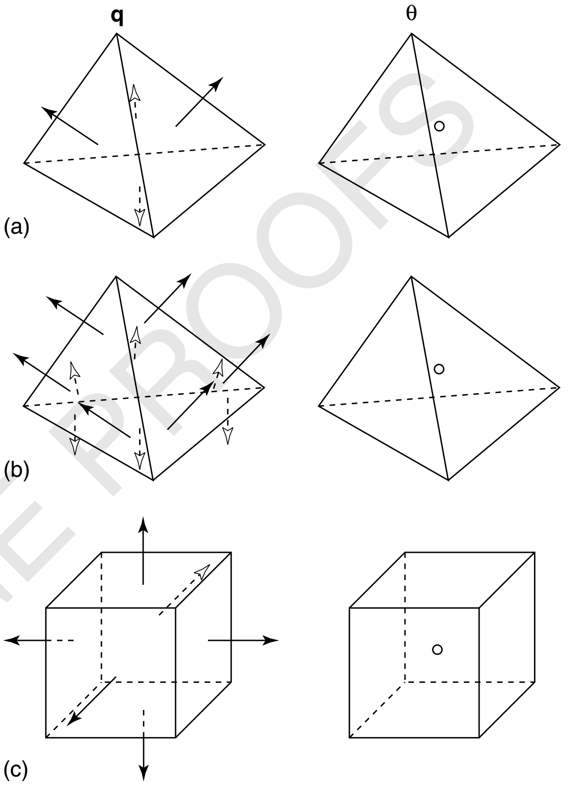

Figure 7. 3-D elements for the thermal problem.

In Figure 7(a), we display the tetrahedral version of the $RT_0 - P_0$ element (cf. Figure 1), consisting of a piecewise constant approximation for the temperature θ and of the following element approximating functions for $\mathbf{q}^h$ (see Nedelec, 1980):

$$\mathbf{q}^h|_T = \mathbf{p}_0 + p_0\begin{Bmatrix}x\\ y\\ z\end{Bmatrix} = \begin{Bmatrix}a_0 + p_0x\\ b_0 + p_0y\\ c_0 + p_0z\end{Bmatrix}$$
$$a_0, b_0, c_0, p_0 \in \mathbb{R} \tag{240}$$

Therefore, in each tetrahedron T, the space for the approximated flux has dimension 4 and the degrees of freedom are precisely the values $\int_f \mathbf{q}^h \cdot \mathbf{n}\, \mathrm{d}\sigma$ on each tetrahedron face f.

The three-dimensional version of the $BDM_1 - P_0$ element (cf. Figure 3) is shown in Figure 7(b). The approximated temperature is still piecewise constant, while the discretized flux $\mathbf{q}^h|_T$ is a *fully linear* vectorial function.

We also present the extension of the $RT_{[0]} - P_0$ to the case of cubic geometry, as depicted in Figure 7(c).

Remark 8. We conclude our discussion on the thermal problem by noticing the obvious fact that the linear system (221) has an indefinite matrix, independent of the chosen approximation spaces. This is a serious source of trouble. For the discretizations considered above, we can however overcome this drawback. Following Fraeijs de Veubeke (1965), one can first work with fluxes that are *totally discontinuous*, forcing back the continuity of the normal components by means of suitable *interelement Lagrange multipliers*, whose physical meaning comes out to be ‘generalized temperatures’ (for instance, approximations of the temperature mean value on each edge). As the fluxes are now discontinuous, it is possible to eliminate them by static condensation at the element level. This will give a system involving only the temperatures and the interelement multipliers. At this point, however, it becomes possible to eliminate the temperatures as well (always at the element level), leaving a final system that involves only the multipliers. This final system has a symmetric and positive definite matrix, a very useful property from the computational point of view. For a detailed discussion about these ideas, we refer to Arnold and Brezzi (1985), Marini (1985), and Brezzi *et al.* (1986, 1987, 1988). For another way to eliminate the flux variables (although with some geometrical restrictions) see also Baranger, Maitre and Oudin (1996). For yet another procedure to reduce the number of unknowns in (221) and getting a symmetric positive definite matrix, see Alotto and Perugia (1999).

4.2 Stokes equation

As detailed in Section 2.2, the discretization of the Stokes problem leads to solving the following algebraic system:

$$\begin{bmatrix} \mathbf{A} & \mathbf{B}^{\mathrm{T}} \\ \mathbf{B} & \mathbf{0} \end{bmatrix} \begin{Bmatrix} \hat{\mathbf{u}} \\ \hat{\mathbf{p}} \end{Bmatrix} = \begin{Bmatrix} \mathbf{f} \\ \mathbf{0} \end{Bmatrix} \tag{241}$$

where

$$\begin{cases} \mathbf{A}|_{ij} = \mu \displaystyle\int_\Omega \left[\nabla \mathbf{N}_i^{\mathbf{u}} : \nabla \mathbf{N}_j^{\mathbf{u}}\right] \mathrm{d}\Omega, \quad \hat{\mathbf{u}}|_i = \hat{u}_i \\ \mathbf{B}|_{rj} = -\displaystyle\int_\Omega \left[N_r^p \operatorname{div}\left(\mathbf{N}_j^{\mathbf{u}}\right)\right] \mathrm{d}\Omega, \quad \hat{\mathbf{p}}|_r = \hat{p}_r \\ \mathbf{f}|_i = \displaystyle\int_\Omega \left[\mathbf{N}_i^{\mathbf{u}} \cdot \mathbf{b}\right] \mathrm{d}\Omega \end{cases} \tag{242}$$

Above, $\mathbf{N}_i^{\mathbf{u}}$ and N_r^p are the interpolation functions for the velocity $\mathbf{u}$ and the pressure p respectively. Also, $\hat{\mathbf{u}}$ and $\hat{\mathbf{p}}$ are the vectors containing the velocity and the pressure unknowns. In the sequel, we will always consider the case of homogeneous boundary conditions for the velocity field along the whole boundary $\partial\Omega$. As a consequence, the pressure field is determined only up to a constant. Uniqueness can, however, be recovered, for instance, by insisting that the pressure has zero mean value over the domain Ω or by fixing its value at a given point.

We also remark that, since there is no derivative of N_r^p in the definition of the matrix $\mathbf{B}$, both continuous and discontinuous pressure approximations can be chosen. On the contrary, the symmetric gradients of $\mathbf{N}_i^{\mathbf{u}}$ entering in the matrix $\mathbf{A}$ suggest that the approximated velocities should be continuous across adjacent elements.

If we introduce the norms

$$\|\hat{\mathbf{u}}\|_X^2 := \mu \int_\Omega \left|\nabla(\mathbf{N}_i^{\mathbf{u}} \hat{u}_i)\right|^2 \mathrm{d}\Omega \tag{243}$$

and

$$\|\hat{\mathbf{p}}\|_Y^2 := \int_\Omega \left|N_r^p \hat{p}_r\right|^2 \mathrm{d}\Omega \tag{244}$$

the continuity conditions (101) and the ellipticity condition (125) of the previous section are clearly satisfied, namely, with $M_a = 1$, $M_b = \sqrt{(d/\mu)}$, and $\alpha = 1$. Therefore, a stable method is achieved provided the only *inf–sup* condition (112) is fulfilled.

4.2.1 *Triangular elements with continuous pressure interpolation*

In this section, we describe some stable triangular element for which the pressure field is interpolated by means of *continuous* functions.

• *The MINI element.* Given a triangular mesh $\mathcal{T}_h$, of Ω, for the approximated velocity $\mathbf{u}^h$ we require that (cf. Arnold, Brezzi and Fortin, 1984)

i. for each $T \in \mathcal{T}_h$, the two components of $\mathbf{u}^h$ are the sum of a linear function plus a standard *cubic bubble* function;
ii. the two components of $\mathbf{u}^h$ are *globally continuous* functions on Ω.

Concerning the discretized pressure p^h, we simply take piecewise linear and globally continuous functions.

For the generic element $T \in \mathcal{T}_h$, the elemental degrees of freedom for $\mathbf{u}^h$ are its (vectorial) values at the triangle vertexes and barycenter. A basis for the element approximation space of each component of $\mathbf{u}^h$ can be obtained by

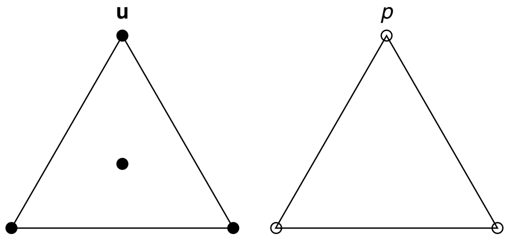

Figure 8. Degrees of freedom for MINI element.

considering the following four shape functions:

$$\begin{cases} N_k = N_k(x,y) = \lambda_k \quad k = 1,2,3 \\ N_b = N_b(x,y) = 27\lambda_1\lambda_2\lambda_3 \end{cases} \tag{245}$$

where $\{\lambda_k = \lambda_k(x,y), k = 1,2,3\}$ denote the usual area coordinates on T.

Furthermore, a set of elemental degrees of freedom for p^h is given by its values at the triangle vertexes, while the three shape functions to be used are obviously

$$N_k = \lambda_k \qquad\qquad k = 1,2,3 \tag{246}$$

The element degrees of freedom for both $\mathbf{u}^h$ and p^h are schematically depicted in Figure 8. We finally remark that the bubble functions for the velocity are internal modes, so that they can be eliminated on the element level by means of the so-called *static condensation* procedure (cf. Hughes (1987), for instance). As a consequence, these additional degrees of freedom do not significantly increase the computational costs.

• *The Hood–Taylor elements.* These elements arise from the experimental evidence that using a velocity approximation of one degree higher than the approximation for pressure gave reliable results (cf. Hood and Taylor, 1973). We are therefore led to consider, for each integer k with $k \geq 1$, the following interpolation fields.

The approximated velocity $\mathbf{u}^h$ is such that

i. for each $T \in \mathcal{T}_h$, the two components of $\mathbf{u}^h$ are polynomials of degree at most $k+1$;
ii. the two components of $\mathbf{u}^h$ are globally continuous functions on Ω.

For the approximated pressure p^h, we ask that

i. for each $T \in \mathcal{T}_h$, p^h is a polynomial of degree at most k;
ii. p^h is a globally continuous function on Ω.

Figure 9 shows the $\mathbf{u}^h$ and p^h element degrees of freedom, for the lowest-order Hood–Taylor method (i.e. $k = 1$).

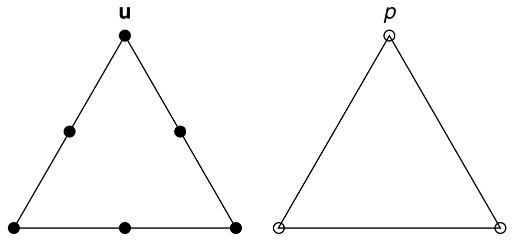

Figure 9. Degrees of freedom for the lowest-order Hood–Taylor element.

Remark 9. A first theoretical analysis of the lowest-order Hood–Taylor method ($k = 1$) was developed in Bercovier and Pironneau (1977), later improved in Verfürth (1984). The case $k = 2$ was treated in Brezzi and Falk (1991), while an analysis covering every choice of k was presented in Boffi (1994). We also remark that the *discontinuous pressure* version of the Hood–Taylor element typically results in an unstable method. However, stability can be recovered by imposing certain restrictions on the mesh for $k \geq 3$ (see Vogelius (1983) and Scott and Vogelius (1985)), or by taking advantage of suitable stabilization procedures for $k \geq 1$; see Mansfield (1982) and Boffi (1995).

• *The* $(P_1$*-iso-*$P_2) - P_1^c$ *element.* This is a ‘composite’ element whose main advantage is the shape function simplicity. We start by considering a triangular mesh $\mathcal{T}_h$ with meshsize h. From $\mathcal{T}_h$, we build another finer mesh $\mathcal{T}_{h/2}$ by splitting each triangle T of $\mathcal{T}_h$ into four triangles using the edge midpoints of T, as sketched in Figure 10.

The approximated velocity $\mathbf{u}^h$ is now defined using the finer mesh $\mathcal{T}_{h/2}$ according to the following prescriptions:

i. for each triangle of $\mathcal{T}_{h/2}$, the two components of $\mathbf{u}^h$ are *linear* functions;
ii. the two components of $\mathbf{u}^h$ are globally continuous functions on Ω.

On the other hand, the interpolated pressure p^h is piecewise linear in the coarser mesh $\mathcal{T}_h$, and globally continuous on Ω.

For every triangle T' of finer mesh $\mathcal{T}_{h/2}$, the degrees of freedom of $\mathbf{u}^h$ are its values at the vertexes, while an element basis is given by taking the shape functions $N_k = \lambda_k$ $(k = 1,2,3)$ relative to T'.

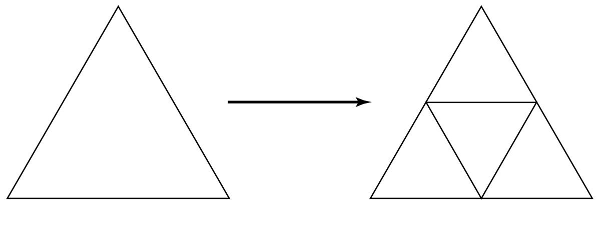

Figure 10. Splitting of a triangle $T \in \mathcal{T}_h$.

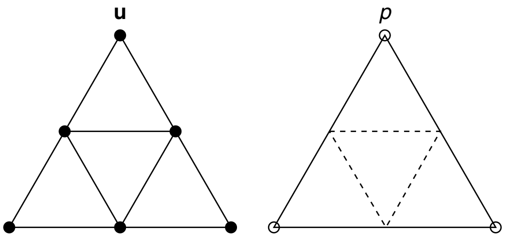

Figure 11. Degrees of freedom for $(P_1\text{-iso-}P_2) - P_1^c$ element.

Instead, by considering the generic triangle T of the coarser mesh $\mathcal{T}_h$, the point values at the three vertexes provide a set of degrees of freedom for p^h. Therefore, the shape functions $N_k = \lambda_k$ $(k = 1, 2, 3)$, relative to T, can be chosen as a basis for the element pressure approximation.

The element degrees of freedom for both $\mathbf{u}^h$ and p^h are schematically depicted in Figure 11.

Remark 10. A popular way to solve system (241) consists in using a penalty method. More precisely, instead of (241), one considers the perturbed system

$$\begin{bmatrix} \mathbf{A} & \mathbf{B}^{\mathrm{T}} \\ \mathbf{B} & -\sigma\mathbf{C} \end{bmatrix} \begin{Bmatrix} \hat{\mathbf{u}} \\ \hat{\mathbf{p}} \end{Bmatrix} = \begin{Bmatrix} \mathbf{f} \\ \mathbf{0} \end{Bmatrix} \tag{247}$$

where the 'mass' matrix $\mathbf{C}$ is defined by

$$\mathbf{C}|_{rs} = \int_\Omega \left[N_r^p N_s^p\right] \mathrm{d}\Omega \tag{248}$$

and $\sigma > 0$ is a 'small' parameter. In the case of discontinuous pressure approximations, the pressure unknowns can be eliminated from (247) on the element level, leading therefore to the following system for $\hat{\mathbf{u}}$:

$$\left(\mathbf{A} + \sigma^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{B}\right)\hat{\mathbf{u}} = \mathbf{f} \tag{249}$$

with $\mathbf{C}$ '*easy-to-invert*' (namely, block diagonal). When continuous pressure approximations are considered, the inverse of $\mathbf{C}$ is in general a full matrix, so that the elimination of the pressure unknowns seems impossible on the element level. We have, however, the right to choose a *different* penalizing term in (247): for instance, we could replace $\mathbf{C}$ by a diagonal matrix $\widetilde{\mathbf{C}}$, obtained from $\mathbf{C}$ by a suitable *mass lumping* procedure (cf. Hughes, 1987). The pressure elimination becomes now easy to perform, leading to (cf. (249))

$$\left(\mathbf{A} + \sigma^{-1}\mathbf{B}^{\mathrm{T}}\widetilde{\mathbf{C}}^{-1}\mathbf{B}\right)\hat{\mathbf{u}} = \mathbf{f} \tag{250}$$

A drawback of this approach, however, not so serious for low-order schemes, stands in a larger bandwidth for the matrix $\left(\mathbf{A} + \sigma^{-1}\mathbf{B}^{\mathrm{T}}\widetilde{\mathbf{C}}^{-1}\mathbf{B}\right)$. For more details about this strategy, we refer to Arnold, Brezzi and Fortin (1984).

4.2.2 *Triangular elements with discontinuous pressure interpolation*

In this section, we describe some stable triangular element for which the pressure field is interpolated by means of discontinuous functions. It is worth noticing that all these elements have velocity degrees of freedom associated with the element edges. This feature is indeed of great help in proving the *inf–sup* condition for elements with discontinuous pressure interpolation (cf. Remark 16).

• *The Crouzeix–Raviart element.* Our first example of discontinuous pressure elements is the one proposed and analyzed in Crouziex and Raviart (1973). It consists in choosing the approximated velocity $\mathbf{u}^h$ such that

i. for each $T \in \mathcal{T}_h$, the two components of $\mathbf{u}^h$ are the sum of a quadratic function plus a standard cubic bubble function;
ii. the two components of $\mathbf{u}^h$ are globally continuous functions on Ω.

Moreover, for the discretized pressure p^h, we simply take the piecewise linear functions, *without requiring any continuity between adjacent elements.*

The elemental approximation of each component of $\mathbf{u}^h$ can be described by means of the following seven shape functions

$$\begin{cases} N_k = \lambda_k \qquad k = 1, 2, 3 \\ N_4 = 4\lambda_2\lambda_3, \quad N_5 = 4\lambda_1\lambda_3, \quad N_6 = 4\lambda_1\lambda_2 \\ N_b = 27\lambda_1\lambda_2\lambda_3 \end{cases} \tag{251}$$

The degrees of freedom are the values at the triangle vertexes and edge midpoints, together with the value at the barycenter.

Concerning the pressure approximation in the generic triangle T, we take the three shape functions

$$\begin{cases} N_1 = 1 \\ N_2 = x \\ N_3 = y \end{cases} \tag{252}$$

and the degrees of freedom can be chosen as the values at three internal and noncollinear points of the triangle.

Figure 12 displays the element degrees of freedom for both $\mathbf{u}^h$ and p^h.

• *The $P_{k+2} - P_k$ family.* We now present a class of mixed methods consisting in choosing, for any integer k with $k \geq 0$, the following interpolation fields.

For the approximated velocity $\mathbf{u}^h$, we require that

i. for each $T \in \mathcal{T}_h$, the two components of $\mathbf{u}^h$ are polynomials of degree at most $k + 2$;

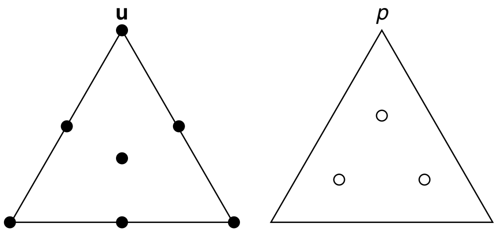

Figure 12. Degrees of freedom for Crouzeix–Raviart element.

ii. the two components of $\mathbf{u}^h$ are globally continuous functions on Ω.

Instead, the approximated pressure p^h is a polynomial of degree at most k for each $T \in \mathcal{T}_h$, with no continuity imposed across the triangles.

Figure 13 shows the local degrees of freedom, for both $\mathbf{u}^h$ and p^h, of the lowest-order method (i.e. $k = 0$), which has been proposed and mathematically analyzed in Fortin (1975).

Remark 11. For the $P_{k+2} - P_k$ family, the discretization error in energy norm is of order h^{k+1} for both the velocity and the pressure, even though the P_{k+2}-approximation should suggest an order h^{k+2} for the velocity field. This 'suboptimality' is indeed a consequence of the poor pressure interpolation (polynomials of degrees at most k). However, taking advantage of a suitable augmented Lagrangian formulation, the $P_{k+2} - P_k$ family can be improved to obtain a convergence rate of order $h^{k+3/2}$ for the velocity, without significantly increasing the computational costs. We refer to Boffi and Lovadina (1997) for details on such an approach.

• *The* $(P_1$*-iso-*$P_2) - P_0$ *element.* Another stable element can be designed by taking the P_1-iso-P_2 element for the approximated velocity, and a piecewise constant approximation for the pressure. More precisely, as for the $(P_1\text{-iso-}P_2) - P_1^c$ element previously described, we consider a triangular mesh $\mathcal{T}_h$ with meshsize h. We then build a finer mesh $\mathcal{T}_{h/2}$ according to the procedure sketched in Figure 10.

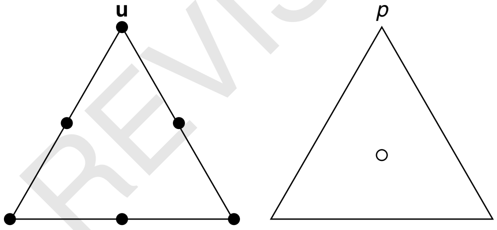

Figure 13. Degrees of freedom for P_2–P_0 element.

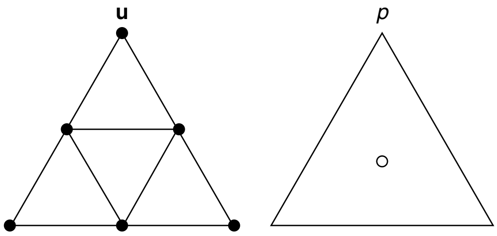

Figure 14. Degrees of freedom for $(P_1\text{-iso-}P_2) - P_0$ element.

We recall that the approximated velocity $\mathbf{u}^h$ is given using the finer mesh $\mathcal{T}_{h/2}$ and requiring that

i. for each triangle of $\mathcal{T}_{h/2}$, the two components of $\mathbf{u}^h$ are *linear* functions;
ii. the two components of $\mathbf{u}^h$ are *globally continuous* functions on Ω.

Instead, the pressure approximation is defined on the coarser mesh $\mathcal{T}_h$ by selecting the piecewise constant functions.

The local degrees of freedom for both $\mathbf{u}^h$ and p^h are shown in Figure 14.

• *The non-conforming* $P_1^{NC} - P_0$ *element.* We present an element, attributable to Crouziex and Raviart (1973), for which the approximated velocity $\mathbf{u}^h$ is obtained by requiring that

i. for each triangle the two components of $\mathbf{u}^h$ are linear functions;
ii. continuity of $\mathbf{u}^h$ across adjacent elements is imposed only at edge midpoints.

For the approximated pressure p^h, we simply take the piecewise constant functions.

Given a triangle $T \in \mathcal{T}_h$, the degrees of freedom for the approximating velocity $\mathbf{u}^h$ are the values at the three edge midpoints. Furthermore, for each component of $\mathbf{u}^h$, an element basis on triangle T is provided by

$$N_k = 1 - 2\lambda_k \qquad k = 1, 2, 3$$

The lack of continuity for the discrete velocity implies that the differential operators (gradient and divergence) acting on $\mathbf{u}^h$ should be taken element-wise. For instance, the matrix $\mathbf{B}$ should be written as

$$\mathbf{B}|_{rj} = -\sum_{T \in \mathcal{T}_h} \int_T \left[N_r^p \text{div}\left(\mathbf{N}_j^{\mathbf{u}}\right)\right] \text{d}\Omega \tag{253}$$

The degrees of freedom are displayed in Figure 15. We remark that applicability of the $P_1^{\text{NC}} - P_0$ element is limited to problems with Dirichlet boundary conditions for the

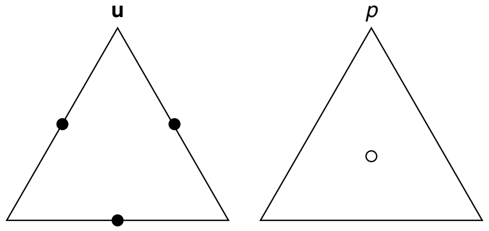

Figure 15. Degrees of freedom for $P_1^{\mathrm{NC}} - P_0$ element.

displacement field imposed on the whole $\partial\Omega$. For other situations (e.g. a pure traction problem), the scheme exhibits spurious mechanisms, because of its inability to control the rigid body rotations (cf. Hughes, 1987). Two stable modifications of the $P_1^{\mathrm{NC}} - P_0$ element have been proposed and analyzed in Falk (1991) and, recently, in Hansbo and Larson (2003).

4.2.3 *Quadrilateral elements*

Many of the triangular elements presented above have their quadrilateral counterpart. As an example, we here show the so-called $Q_2 - Q_1^c$ element, which is the quadrilateral version of the lowest-order Hood–Taylor element (cf. Figure 9). Accordingly, the velocity is approximated by biquadratic and continuous functions, while the pressure is discretized by means of bilinear and continuous functions, as depicted in Figure 16.

Another very popular scheme is the $Q_2 - P_1$ element, based on the same approximated velocities as before. Instead, the interpolating functions for the pressure are piecewise linear, without requiring any continuity across the elements. The local degrees of freedom are displayed in Figure 17.

4.2.4 *Three-dimensional elements*

Several elements previously described extend to the case of three-dimensional problems. In Figure 18(a), we show a continuous pressure tetrahedral element, which is nothing but the 3-D version of the MINI element (cf. Figure 8). Also, the non-conforming $P_1^{\mathrm{NC}} - P_0$ element (cf. Figure 15) has its three-dimensional counterpart, as depicted in Figure 18(b). We remark that the degrees of freedom for the velocity are given by the values at the barycenter of each tetrahedron *face*. Figure 18(c) shows an example of cubic element, which is exactly the 3-D version of the popular $Q_2 - P_1$ element (cf. Figure 17).

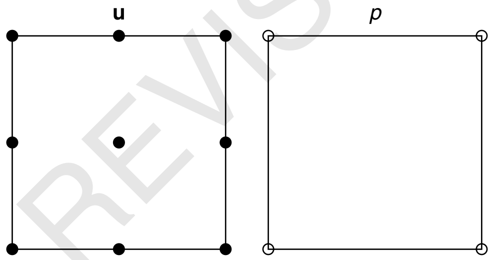

Figure 16. Degrees of freedom for $Q_2 - Q_1^c$ element.

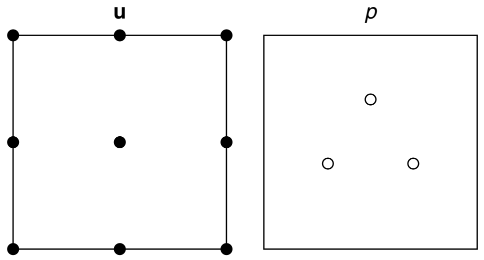

Figure 17. Degrees of freedom for $Q_2 - P_1$ element.

Finally, we refer to Stenberg (1987) for the analysis, based on the so-called *macroelement technique* introduced in Stenberg (1984), of the lowest-order 3-D Hood–Taylor method, and to Boffi (1997) for the higher-order case.

4.2.5 *Stabilized formulations*

From the above discussion, it should be clear that the fulfillment of the *inf–sup* condition requires a careful choice of the discretization spaces for the velocity and the pressure. A first strategy to obtain stability has been to derive the

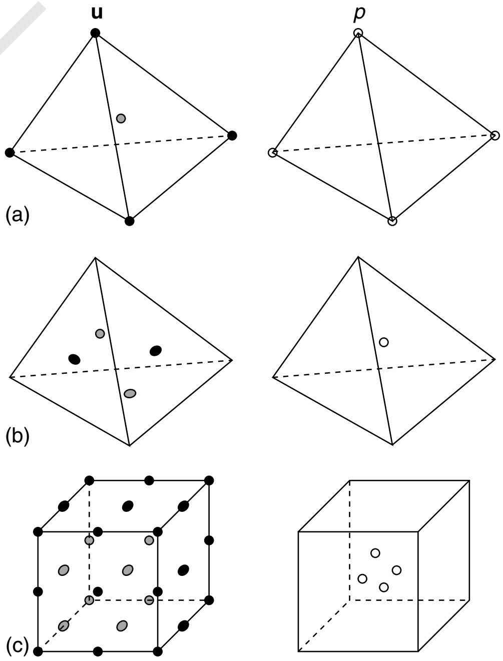

Figure 18. 3-D Stokes elements.

numerical scheme from a perturbation of functional (37), by considering (see Brezzi and Pitkäranta, 1984)

$$\widetilde{L}(\mathbf{u}, p) = \frac{1}{2}\mu \int_\Omega [\nabla\mathbf{u} : \nabla\mathbf{u}]\, \mathrm{d}\Omega - \int_\Omega [\mathbf{b}\cdot\mathbf{u}]\, \mathrm{d}\Omega - \int_\Omega \left[p \operatorname{div}\mathbf{u}\right] \mathrm{d}\Omega - \frac{\alpha}{2}\sum_{K\in\mathcal{T}_h}\int_K h_K^2 \left|\nabla p\right|^2 \mathrm{d}\Omega \quad (254)$$

where α is a positive parameter and h_K is the diameter of the element $K \in \mathcal{T}_h$. The 'perturbation term'

$$\frac{\alpha}{2}\sum_{K\in\mathcal{T}_h}\int_K h_K^2 \left|\nabla p\right|^2 \mathrm{d}\Omega$$

has a stabilizing effect on the discretized Euler–Lagrange equations emanating from (254). It however introduces a consistency error, so that the convergence rate in energy norm cannot be better than $O(h)$, even though higher-order elements are used. Following the ideas in Hughes and Franca (1987) and Hughes *et al.* (1986), this drawback may be overcome by means of a suitable augmented Lagrangian formulation. Instead of considering (37) or (254), one can introduce the augmented functional

$$\begin{aligned} L^{\mathrm{agm}}(\mathbf{u}, p) &= \frac{1}{2}\mu \int_\Omega [\nabla\mathbf{u} : \nabla\mathbf{u}]\, \mathrm{d}\Omega \\ &\quad - \int_\Omega [\mathbf{b}\cdot\mathbf{u}]\, \mathrm{d}\Omega - \int_\Omega \left[p \operatorname{div}\mathbf{u}\right] \mathrm{d}\Omega \\ &\quad - \frac{1}{2}\sum_{K\in\mathcal{T}_h}\int_K \alpha(K)\left|\mu\Delta\mathbf{u} - \nabla p + \mathbf{b}\right|^2 \mathrm{d}\Omega \qquad (255) \end{aligned}$$

where, for each element $K \in \mathcal{T}_h$, $\alpha(K)$ is a positive parameter at our disposal. Because of the structure of the 'additional term' in (255), both the functionals (37) and (255) have the same critical point, that is, the solution of the Stokes problem. Therefore, the discretized Euler–Lagrange equations associated with (255) deliver a consistent method, whenever conforming approximations have been selected. As before, the augmented term may have a stabilizing effect, allowing the choice of a wider class of elements. For instance, if

$$\alpha(K) = \bar{\alpha} h_K^2$$

where $\bar{\alpha}$ is sufficiently 'small', any finite element approximation of velocity and pressure (as far as the pressure is discretized with continuous finite elements) leads to a stable scheme, with respect to an appropriate norm (see Franca and Hughes, 1988).

This approach has several interesting variants. Indeed, considering the Euler–Lagrange equations associated with (255) we have

$$\begin{aligned} &\mu \int_\Omega [\nabla\mathbf{u} : \nabla\mathbf{v}]\, \mathrm{d}\Omega - \int_\Omega [\mathbf{b}\cdot\mathbf{v}]\, \mathrm{d}\Omega - \int_\Omega \left[p \operatorname{div}\mathbf{v}\right] \mathrm{d}\Omega \\ &\quad - \int_\Omega \left[q \operatorname{div}\mathbf{u}\right] \mathrm{d}\Omega - \sum_{K\in\mathcal{T}_h}\int_K \alpha(K)\ \left[\mu\Delta\mathbf{u} - \nabla p + \mathbf{b}\right] \\ &\quad \cdot \left[\mu\Delta\mathbf{v} - \nabla q\right] \mathrm{d}\Omega = 0 \qquad (256) \end{aligned}$$

for all test functions $\mathbf{u}$ and q. The term in second line of (256) represents our *consistent perturbation*. A careful analysis can show that its stabilizing effect still works if we change it into

$$+\sum_{K\in\mathcal{T}_h}\int_K \alpha(K)\ \left[\mu\Delta\mathbf{u} - \nabla p + \mathbf{b}\right]\cdot\left[\mu\Delta\mathbf{v} + \nabla q\right] \mathrm{d}\Omega \qquad (257)$$

(that is, changing the sign of the whole term, but changing also the sign of ∇q in the second factor) or simply into

$$+\sum_{K\in\mathcal{T}_h}\int_K \alpha(K)\ \left[\mu\Delta\mathbf{u} - \nabla p + \mathbf{b}\right]\cdot\nabla q\, \mathrm{d}\Omega \qquad (258)$$

For a general analysis of these possible variants, we refer to Baiocchi and Brezzi (1993). A deeper analysis shows that, in particular, the formulation (257) can be interpreted as changing the space of velocities with the addition of suitable bubble functions and then eliminate them by static condensation. This was pointed out first in Pierre (1989), and then in a fully systematic way in Baiocchi, Brezzi and Franca (1993).

Other possible stabilizations can be obtained by adding penalty terms that penalize the jumps in the pressure variable over suitable macroelements. See, for instance, Silvester and Kechar (1990). This as well can be seen as adding suitable bubbles on the macroelements and eliminating them by static condensation.

For a more general survey of these and other types of stabilizations, see Brezzi and Fortin (2001) and the references therein.

Another approach to get stable elements is based on the so-called *Enhanced Strain Technique*, introduced in Simo and Rifai (1990) in the context of elasticity problems. As already mentioned in Section 2.3, the basic idea consists in enriching the symmetric gradients $\nabla^s\mathbf{u}^h$ with additional local modes. An analysis of this strategy for displacement-based elements has been developed in Reddy and Simo (1995) and Braess (1998). Within the framework of the $(\mathbf{u}, \pi)$ formulation for incompressible elasticity problems (and therefore for the Stokes problem), the enhanced strain technique has been successfully used in Pantuso and Bathe

(1995) (see also Lovadina (1997) for a stability and convergence analysis), and more recently in Lovadina and Auricchio (2003) and Auricchio *et al.* (200x).

4.3 Elasticity

We now briefly consider the elasticity problem in the framework of the Hellinger–Reissner variational principle (59). We recall that after discretization we are led to solve a problem of the following type (cf. (57), (58), and (61)):

$$\begin{bmatrix} \mathbf{A} & \mathbf{B}^{\mathrm{T}} \\ \mathbf{B} & \mathbf{0} \end{bmatrix} \begin{Bmatrix} \hat{\boldsymbol{\sigma}} \\ \hat{\mathbf{u}} \end{Bmatrix} = \begin{Bmatrix} \mathbf{0} \\ \mathbf{g} \end{Bmatrix} \tag{259}$$

where

$$\begin{cases} \mathbf{A}|_{ij} = \displaystyle\int_\Omega \left[\mathbf{N}_i^{\sigma} : \mathbb{D}^{-1}\mathbf{N}_j^{\sigma}\right] \mathrm{d}\Omega, & \hat{\boldsymbol{\sigma}}|_j = \hat{\sigma}_j \\ \mathbf{B}|_{rj} = \displaystyle\int_\Omega \left[\mathbf{N}_r^{\mathbf{u}} \cdot \operatorname{div}\left(\mathbf{N}_j^{\sigma}\right)\right] \mathrm{d}\Omega, & \hat{\mathbf{u}}|_r = \hat{u}_r \\ \mathbf{g}|_r = -\displaystyle\int_\Omega \left[\mathbf{N}_r^{\mathbf{u}} \cdot \mathbf{b}\right] \mathrm{d}\Omega \end{cases} \tag{260}$$

Above, $\mathbf{N}_i^{\sigma}$ and $\mathbf{N}_r^{\mathbf{u}}$ are the interpolation functions for the stress $\boldsymbol{\sigma}$ and the displacement $\mathbf{u}$ respectively. Moreover, $\hat{\boldsymbol{\sigma}}$ and $\hat{\mathbf{u}}$ are the vectors of stress and displacement unknowns. We note that since the divergence operator acts on the shape functions $\mathbf{N}_i^{\sigma}$ (see the $\mathbf{B}$ matrix in (260)), the approximated normal stress $\boldsymbol{\sigma}^h\mathbf{n}$ should be continuous across adjacent elements. On the contrary, no derivative operator acts on the shape functions $\mathbf{N}_r^{\mathbf{u}}$, so that we are allowed to use discontinuous approximation for the displacement field. Analogously to the thermal diffusion problem (see Section 4.1), the proper norms for $\hat{\boldsymbol{\sigma}}$ and $\hat{\mathbf{u}}$ are as follows (cf. (223) and (224)):

$$\begin{aligned} \|\hat{\boldsymbol{\sigma}}\|_X^2 := & \int_\Omega \left[\left(\mathbf{N}_i^{\sigma}\hat{\sigma}_i\right) : \mathbb{D}^{-1}\left(\mathbf{N}_j^{\sigma}\hat{\sigma}_j\right)\right] \mathrm{d}\Omega \\ & + \frac{\ell^2}{D_*}\int_\Omega |\operatorname{div}(\mathbf{N}_i^{\sigma}\hat{\sigma}_i)|^2 \,\mathrm{d}\Omega \end{aligned} \tag{261}$$

and

$$\|\hat{\mathbf{u}}\|_Y^2 := \int_\Omega |\mathbf{N}_r^{\mathbf{u}}\hat{u}_r|^2 \,\mathrm{d}\Omega \tag{262}$$

where ℓ is some characteristic length of the domain Ω and D_* is some characteristic value of the elastic tensor.

Despite the apparent similarity with the corresponding (221) to (222) of the thermal diffusion problem, finding approximation spaces for (259) to (260), which satisfy both the *inf–sup* and the *elker* conditions, is much more difficult (see e.g. Brezzi and Fortin (1991) for a discussion on such a point). Here below, we present two triangular elements proposed and analyzed in Johnson and Mercier (1978) and Arnold and Winther (2002) respectively.

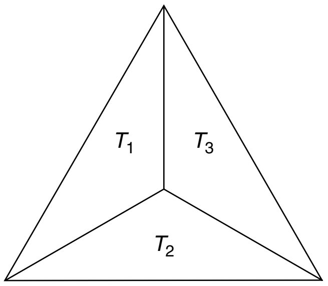

Figure 19. Splitting of a generic triangle for the Johnson–Mercier element.

• *The Johnson–Mercier element.* This method takes advantage of a 'composite' approximation for the stress field. More precisely, we first split every triangle $T \in \mathcal{T}_h$ into three subtriangles T_i $(i = 1, 2, 3)$ using the barycenter of T (see Figure 19).

For the approximated stress $\boldsymbol{\sigma}^h$, we then require that

i. in each subtriangle T_i the components of $\boldsymbol{\sigma}^h$ are linear functions;
ii. the normal stress $\boldsymbol{\sigma}^h\mathbf{n}$ is continuous across adjacent triangles and across adjacent subtriangles.

Accordingly, the discrete stress $\boldsymbol{\sigma}^h$ is not a polynomial on T, but only on the subtriangles T_i. For the generic element $T \in \mathcal{T}_h$, it can be shown (see Johnson and Mercier, 1978) that the elemental degrees of freedom are the following.

i. On the three edges of T: the moments of order 0 and 1 for the vector field $\boldsymbol{\sigma}^h\mathbf{n}$ (12 degrees of freedom);
ii. On T: the moments of order 0 for the symmetric tensor field $\boldsymbol{\sigma}^h$ (3 degrees of freedom).

Moreover, each component of the approximated displacement $\mathbf{u}^h$ is chosen as a piecewise constant function.

Figure 20 displays the element degrees of freedom for both $\boldsymbol{\sigma}^h$ and $\mathbf{u}^h$.

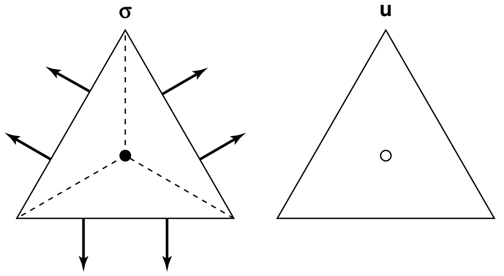

Figure 20. Degrees of freedom for the Johnson–Mercier element.

• *The Arnold–Winther element.* This triangular element has been recently proposed and analyzed in Arnold and Winther (2002), where higher-order schemes are also considered. For the approximated stress $\boldsymbol{\sigma}^h$, we impose that

i. on each $T \in \mathcal{T}_h$, $\boldsymbol{\sigma}^h$ is a symmetric tensor whose components are cubic functions, but $\operatorname{div} \boldsymbol{\sigma}^h$ is a linear vector field;
ii. the normal stress $\boldsymbol{\sigma}^h \mathbf{n}$ is continuous across adjacent triangles.

For each element $T \in \mathcal{T}_h$, the approximation space for the stress field has dimension 24 and the elemental degrees of freedom can be chosen as follows (see Arnold and Winther, 2002):

i. the values of the symmetric tensor field $\boldsymbol{\sigma}^h$ at the vertices of T (9 degrees of freedom);
ii. the moments of order 0 and 1 for the vector field $\boldsymbol{\sigma}^h \mathbf{n}$ on each edge of T (12 degrees of freedom);
iii. the moment of order 0 for $\boldsymbol{\sigma}^h$ on T (3 degrees of freedom).

Furthermore, the components of the approximated displacement $\mathbf{u}^h$ are *piecewise linear* functions, without requiring any continuity across adjacent elements.

In Figure 21, the element degrees of freedom for both $\boldsymbol{\sigma}^h$ and $\mathbf{u}^h$ are schematically depicted.

Remark 12. Other methods exploiting 'composite' approximations as for the Johnson–Mercier element have been proposed and analyzed in Arnold, Douglas and Gupta (1984).

Following the ideas in Fraeijs de Veubeke (1975), a different strategy to obtain reliable schemes for the elasticity problem in the context of the Hellinger–Reissner variational principle consists in the use of unsymmetric approximated stresses. Symmetry is then enforced back in a weak form by the introduction of a suitable Lagrange multiplier. We refer to Amara and Thomas (1979), Arnold, Brezzi and Douglas (1984), Brezzi *et al.* (1986), and Stenberg (1988) for the details on such an approach.

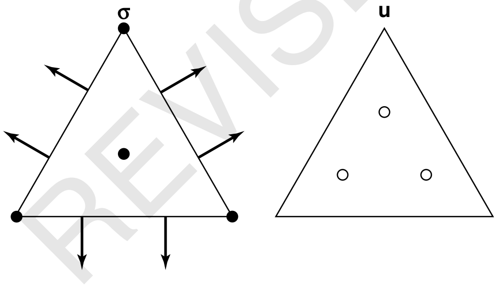

Figure 21. Degrees of freedom for the Arnold–Winther element.

5 TECHNIQUES FOR PROVING THE *INF–SUP* CONDITION

In this section we give some hints on how to prove the *inf–sup* condition (118). We also show how the stability results detailed in Section 3 can be exploited to obtain *error estimates*. We focus on the Stokes problem, as a representative example, but analogous strategies can be applied to analyze most of the methods considered in Section 4.

We begin recalling (cf. (38)) that a weak form of the Stokes problem with homogeneous boundary conditions for the velocity consists in finding $(\mathbf{u}, p)$ such that

$$\begin{cases} \mu \int_\Omega [(\nabla \delta\mathbf{u}) : \nabla \mathbf{u}] \, d\Omega - \int_\Omega \left[\operatorname{div}(\delta\mathbf{u}) \; p\right] d\Omega \\ \qquad\qquad = \int_\Omega [\delta\mathbf{u} \cdot \mathbf{b}] \, d\Omega \\ \int_\Omega \left[\delta p \operatorname{div} \mathbf{u}\right] d\Omega = 0 \end{cases} \tag{263}$$

for any admissible velocity variation $\delta\mathbf{u}$ and any admissible pressure variation δp. On the other hand, as detailed in Section 4.2, the discretized problem consists in solving

$$\begin{bmatrix} \mathbf{A} & \mathbf{B}^{\mathrm{T}} \\ \mathbf{B} & \mathbf{0} \end{bmatrix} \begin{Bmatrix} \hat{\mathbf{u}} \\ \hat{\mathbf{p}} \end{Bmatrix} = \begin{Bmatrix} \mathbf{f} \\ \mathbf{0} \end{Bmatrix} \tag{264}$$

where

$$\begin{cases} \mathbf{A}|_{ij} = \mu \int_\Omega \left[\nabla \mathbf{N}_i^{\mathbf{u}} : \nabla \mathbf{N}_j^{\mathbf{u}}\right] d\Omega, \quad \hat{\mathbf{u}}|_i = \hat{u}_i \\ \mathbf{B}|_{rj} = -\int_\Omega \left[N_r^p \operatorname{div}\left(\mathbf{N}_j^{\mathbf{u}}\right)\right] d\Omega, \quad \hat{\mathbf{p}}|_r = \hat{p}_r \\ \mathbf{f}|_i = \int_\Omega \left[\mathbf{N}_i^{\mathbf{u}} \cdot \mathbf{b}\right] d\Omega \end{cases} \tag{265}$$

with $i, j = 1, \ldots, n$ and $r = 1, \ldots, m$.

With our notation for the Stokes problem, the *inf–sup* condition in its equivalent form (119) consists in requiring the existence of a positive constant β, independent of h, such that

$$\forall \; \hat{\mathbf{q}} \in \mathbf{Y} \quad \sup_{\hat{\mathbf{z}} \in \mathbf{X} \setminus \{\mathbf{0}\}} \frac{\hat{\mathbf{z}}^{\mathrm{T}} \mathbf{B}^{\mathrm{T}} \hat{\mathbf{q}}}{\|\hat{\mathbf{z}}\|_X} \geq \beta \|\hat{\mathbf{q}}\|_Y \tag{266}$$

where $\mathbf{X} \equiv \mathbb{R}^n$ and $\mathbf{Y} \equiv \mathbb{R}^m$.

Moreover, in what follows, we need to introduce the space $\mathcal{X}$ for vectorial functions $\mathbf{v}$, defined by

$$\mathcal{X} = \left\{ \mathbf{v} : \mathbf{v}|_{\partial\Omega} = 0, \qquad \|\mathbf{v}\|_{\mathcal{X}}^2 := \mu \int_\Omega |\nabla \mathbf{v}|^2 \, d\Omega < +\infty \right\} \tag{267}$$

and the space $\mathcal{Y}$ for scalar functions q, defined by

$$\mathcal{Y} = \left\{ q : \|q\|_{\mathcal{Y}}^2 := \int_\Omega |q|^2 \, \mathrm{d}\Omega < +\infty \right\} \tag{268}$$

Remark 13. It is worth noticing that, whenever an approximated velocity $\mathbf{u}^h = \mathbf{N}_i^{\mathbf{u}} \hat{u}_i$ is considered, the following holds (cf. (243))

$$\|\hat{\mathbf{u}}\|_X = \left(\mu \int_\Omega |\nabla \mathbf{N}_i^{\mathbf{u}} \hat{u}_i|^2 \, \mathrm{d}\Omega \right)^{1/2} = \|\mathbf{u}^h\|_{\mathcal{X}} \tag{269}$$

Therefore, the X-norm we have tailored for *vector* $\hat{\mathbf{u}} \in \mathbf{X}$ coincides with the $\mathcal{X}$-norm of the reconstructed function $\mathbf{u}^h$. Similarly (cf. (244)), if $p^h = N_r^p \hat{p}_r$, we have

$$\|\hat{\mathbf{p}}\|_Y = \left(\int_\Omega |N_r^p \hat{p}_r|^2 \, \mathrm{d}\Omega \right)^{1/2} = \|p^h\|_{\mathcal{Y}} \tag{270}$$

5.1 Checking the *inf–sup* condition

As already mentioned, a rigorous proof of the *inf–sup* condition is typically a difficult task, mainly because several technical mathematical problems have to be overcome. In this section, we present two of the most powerful tools for proving the *inf–sup* property. The first technique (*Fortin's trick*) can be used, for instance, to study the stability of the $P_1^{\mathrm{NC}} - P_0$ element (cf. Figure 15) and the Crouzeix–Raviart element (cf. Figure 12), as we are going to detail below. The second one (*Verfürth's trick*) can be applied basically to all the approximations with continuous pressure and it will be exemplified by considering the MINI element (cf. Figure 8).

Although we are aware that the subsequent analysis is not completely satisfactory from the mathematical point of view, it nonetheless highlights some of the basic ideas behind the analysis of mixed finite element methods.

We first need to recall the following important theorem of functional analysis; see Ladyzhenskaya (1969) and Temam (1977), for instance).

Theorem 4. *There exists a constant* $\beta_c > 0$ *such that, for every* $q \in \mathcal{Y}$ *with* $\int_\Omega q \, \mathrm{d}\Omega = 0$, *it holds*

$$\sup_{\mathbf{v} \in \mathcal{X} \setminus \{\mathbf{0}\}} \frac{-\int_\Omega q \operatorname{div} \mathbf{v} \, \mathrm{d}\Omega}{\|\mathbf{v}\|_{\mathcal{X}}} \geq \beta_c \|q\|_{\mathcal{Y}} \tag{271}$$

Remark 14. We remark that estimate (271) is nothing but the *infinite-dimensional* version of the *inf–sup* condition written in its equivalent form (119).

5.1.1 Fortin's trick

The next result provides a criterion for proving the *inf–sup* condition, called *Fortin's trick* (see Fortin, 1977) or, more precisely, Fortin's trick applied to the Stokes problem.

Proposition 4. *Suppose there exists a linear operator* $\widehat{\Pi}_h : \mathcal{X} \longrightarrow \mathbf{X} \equiv \mathbb{R}^n$ *such that*

$$\|\widehat{\Pi}_h \mathbf{v}\|_X \leq C_{\widehat{\Pi}} \|\mathbf{v}\|_{\mathcal{X}} \qquad \forall\, \mathbf{v} \in \mathcal{X} \tag{272}$$

and

$$(\widehat{\Pi}_h \mathbf{v})^{\mathrm{T}} \mathbf{B}^{\mathrm{T}} \hat{\mathbf{q}} = -\int_\Omega \operatorname{div} \mathbf{v} (N_r^p \hat{q}_r) \, \mathrm{d}\Omega \qquad \forall\, \hat{\mathbf{q}} \in \mathbf{Y} \equiv \mathbb{R}^m \tag{273}$$

with $C_{\widehat{\Pi}}$ *independent of h. Then it holds*

$$\forall\, \hat{\mathbf{q}} \in \mathbf{Y} \quad \sup_{\hat{\mathbf{z}} \in \mathbf{X} \setminus \{\mathbf{0}\}} \frac{\hat{\mathbf{z}}^{\mathrm{T}} \mathbf{B}^{\mathrm{T}} \hat{\mathbf{q}}}{\|\hat{\mathbf{z}}\|_X} \geq \frac{\beta_c}{C_{\widehat{\Pi}}} \|\hat{\mathbf{q}}\|_Y \tag{274}$$

that is, the inf–sup condition (266) is fulfilled with $\beta = \beta_c / C_{\widehat{\Pi}}$.

Proof. Take any $\hat{\mathbf{q}} \in \mathbf{Y}$. We notice that from Theorem 4 and Remark 13, we get

$$\sup_{\mathbf{v} \in \mathcal{X} \setminus \{\mathbf{0}\}} \frac{-\int_\Omega \operatorname{div} \mathbf{v} (N_r^p \hat{q}_r) \, \mathrm{d}\Omega}{\|\mathbf{v}\|_{\mathcal{X}}} \geq \beta_c \|N_r^p \hat{q}_r\|_{\mathcal{Y}} = \beta_c \|\hat{\mathbf{q}}\|_Y \tag{275}$$

Therefore, from (273), we have

$$\sup_{\mathbf{v} \in \mathcal{X} \setminus \{\mathbf{0}\}} \frac{(\widehat{\Pi}_h \mathbf{v})^{\mathrm{T}} \mathbf{B}^{\mathrm{T}} \hat{\mathbf{q}}}{\|\mathbf{v}\|_{\mathcal{X}}} \geq \beta_c \|\hat{\mathbf{q}}\|_Y \tag{276}$$

Using (272), from (276) it follows

$$\begin{aligned} \sup_{\mathbf{v} \in \mathcal{X} : \widehat{\Pi}_h \mathbf{v} \neq \mathbf{0}} \frac{(\widehat{\Pi}_h \mathbf{v})^{\mathrm{T}} \mathbf{B}^{\mathrm{T}} \hat{\mathbf{q}}}{\|\widehat{\Pi}_h \mathbf{v}\|_X} &\geq \frac{1}{C_{\widehat{\Pi}}} \sup_{\mathbf{v} \in \mathcal{X} \setminus \{\mathbf{0}\}} \frac{(\widehat{\Pi}_h \mathbf{v})^{\mathrm{T}} \mathbf{B}^{\mathrm{T}} \hat{\mathbf{q}}}{\|\mathbf{v}\|_{\mathcal{X}}} \\ &\geq \frac{\beta_c}{C_{\widehat{\Pi}}} \|\hat{\mathbf{q}}\|_Y \end{aligned} \tag{277}$$

Since, obviously, $\{\widehat{\Pi}_h \mathbf{v};\ \mathbf{v} \in \mathcal{X}\} \subseteq \mathbf{X}$, from (277) we obtain

$$\sup_{\hat{\mathbf{z}} \in \mathbf{X} \setminus \{\mathbf{0}\}} \frac{\hat{\mathbf{z}}^{\mathrm{T}} \mathbf{B}^{\mathrm{T}} \hat{\mathbf{q}}}{\|\hat{\mathbf{z}}\|_X} \geq \sup_{\mathbf{v} \in \mathcal{X} : \widehat{\Pi}_h \mathbf{v} \neq \mathbf{0}} \frac{(\widehat{\Pi}_h \mathbf{v})^{\mathrm{T}} \mathbf{B}^{\mathrm{T}} \hat{\mathbf{q}}}{\|\widehat{\Pi}_h \mathbf{v}\|_X} \geq \frac{\beta_c}{C_{\widehat{\Pi}}} \|\hat{\mathbf{q}}\|_Y \quad \square \tag{278}$$

We now apply Proposition 4 to the $P_1^{\mathrm{NC}} - P_0$ element and the Crouzeix–Raviart element, skipping, however, the proof of (272). In both cases, the strategy for building the operator $\widehat{\Pi}_h$ is the following:

1. On each triangle $T \in \mathcal{T}_h$, we first define a suitable linear operator $\Pi_{h,T}: \mathbf{v} \longmapsto \Pi_{h,T}\mathbf{v}$, valued in the space of velocity approximating functions on T, and satisfying

$$\int_T q^h \operatorname{div}(\Pi_{h,T}\mathbf{v})\,\mathrm{d}\Omega = \int_T q^h \operatorname{div}\mathbf{v}\,\mathrm{d}\Omega \tag{279}$$

for every $\mathbf{v} \in \mathcal{X}$ and every discrete pressure q^h. This will be done by using, in particular, the element degrees of freedom for the velocity approximation.

2. By assembling all the element contributions, we obtain a global linear operator

$$\Pi_h: \begin{cases} \mathcal{X} \longrightarrow \operatorname{Span}\{\mathbf{N}_i^{\mathbf{u}};\, i = 1, \ldots, n\} \\ \mathbf{v} \longmapsto \Pi_h\mathbf{v} = \displaystyle\sum_{T\in\mathcal{T}_h} \Pi_{h,T}\mathbf{v} = \mathbf{N}_i^{\mathbf{u}}\hat{v}_i \end{cases} \tag{280}$$

3. We finally define $\widehat{\Pi}_h: \mathcal{X} \longrightarrow \mathbb{R}^n$ by setting

$$\widehat{\Pi}_h\mathbf{v} = \hat{\mathbf{v}} \quad \text{if} \quad \Pi_h\mathbf{v} = \mathbf{N}_i^{\mathbf{u}}\hat{v}_i \tag{281}$$

that is, $\widehat{\Pi}_h\mathbf{v}$ returns the components of the function $\Pi_h\mathbf{v}$ with respect to the global velocity basis $\{\mathbf{N}_i^{\mathbf{u}};\ i = 1, \ldots, n\}$. From the definition of the matrix $\mathbf{B}$, property (279), (280), and (281), it follows that condition (273) is satisfied.

• *The $P_1^{NC} - P_0$ element.* Fix $T \in \mathcal{T}_h$, and recall that any approximated pressure q^h is a *constant* function on T. We wish to build $\Pi_{h,T}$ in such a way that

$$\int_T q^h \operatorname{div}(\Pi_{h,T}\mathbf{v})\,\mathrm{d}\Omega = \int_T q^h \operatorname{div}\mathbf{v}\,\mathrm{d}\Omega \tag{282}$$

From the divergence theorem, (282) can be alternatively written as

$$\int_{\partial T} q^h (\Pi_{h,T}\mathbf{v}) \cdot \mathbf{n}\,\mathrm{d}s = \int_{\partial T} q^h \mathbf{v} \cdot \mathbf{n}\,\mathrm{d}s \tag{283}$$

Denoting with M_k ($k = 1, 2, 3$) the midpoint of the edge e_k, we define $\Pi_{h,T}\mathbf{v}$ as the unique (vectorial) linear function such that

$$\Pi_{h,T}\mathbf{v}(M_k) = \frac{1}{|e_k|}\int_{e_k} \mathbf{v}\,\mathrm{d}s \qquad k = 1, 2, 3 \tag{284}$$

From the divergence theorem and the Midpoint rule, it follows that

$$\begin{aligned} \int_T q^h \operatorname{div}(\Pi_{1,T}\mathbf{v})\,\mathrm{d}\Omega &= \int_{\partial T} q^h (\Pi_{1,T}\mathbf{v}) \cdot \mathbf{n}\,\mathrm{d}s \\ &= \int_{\partial T} q^h \mathbf{v} \cdot \mathbf{n}\,\mathrm{d}s = \int_T q^h \operatorname{div}\mathbf{v}\,\mathrm{d}\Omega \end{aligned} \tag{285}$$

for every constant function q^h. It is now sufficient to define the global linear operator Π_h as

$$\Pi_h\mathbf{v} = \sum_{T\in\mathcal{T}_h} \Pi_{h,T}\mathbf{v} = \mathbf{N}_i^{\mathbf{u}}\hat{v}_i$$

and the corresponding operator $\widehat{\Pi}_h$ satisfies condition (273) (cf. also (253)).

• *The Crouzeix–Raviart element.* Fix $T \in \mathcal{T}_h$, and recall that any approximated pressure q^h is now a linear function on T. Hence, q^h can be uniquely decomposed as $q^h = q_0 + q_1$, where q_0 is a constant (the mean value of q^h on T), and q_1 is a linear function having zero mean value. We now construct a linear operator $\Pi_{1,T}: \mathbf{v} \longmapsto \Pi_{1,T}\mathbf{v}$, where $\Pi_{1,T}\mathbf{v}$ is a quadratic vectorial polynomial such that

$$\int_T \operatorname{div}(\Pi_{1,T}\mathbf{v})\,\mathrm{d}\Omega = \int_T \operatorname{div}\mathbf{v}\,\mathrm{d}\Omega \tag{286}$$

or, alternatively,

$$\int_{\partial T} (\Pi_{1,T}\mathbf{v}) \cdot \mathbf{n}\,\mathrm{d}s = \int_{\partial T} \mathbf{v} \cdot \mathbf{n}\,\mathrm{d}s \tag{287}$$

Denoting with V_k (resp., M_k) the vertexes of T (resp., the midpoint of the edge e_k), the Cavalieri–Simpson rule shows that condition (287) holds if we set

$$\begin{cases} \Pi_{1,T}\mathbf{v}(V_k) = \mathbf{v}(V_k) \quad k = 1, 2, 3 \\ \Pi_{1,T}\mathbf{v}(M_k) = \dfrac{3}{2|e_k|}\displaystyle\int_{e_k} \mathbf{v}\,\mathrm{d}s - \dfrac{\mathbf{v}(V_{k_1}) + \mathbf{v}(V_{k_2})}{4} \\ \qquad\qquad k = 1, 2, 3 \end{cases} \tag{288}$$

Above, we have denoted with V_{k_1} and V_{k_2} the endpoints of side e_k. So far, we have not used the bubble functions available for the approximated velocity. We now use these two additional degrees of freedom by defining $\mathbf{v}_{b,T}(\mathbf{v})$ as the unique vectorial bubble function such that

$$\int_T q_1 \operatorname{div}\mathbf{v}_{b,T}(\mathbf{v})\,\mathrm{d}\Omega = \int_T q_1 \operatorname{div}(\mathbf{v} - \Pi_{1,T}\mathbf{v})\,\mathrm{d}\Omega \tag{289}$$

for every linear function q_1 having zero mean value on T.

We claim that if $\Pi_{h,T}\mathbf{v} = \mathbf{v}_{b,T}(\mathbf{v}) + \Pi_{1,T}\mathbf{v}$, then

$$\int_T q^h \operatorname{div}(\Pi_{h,T}\mathbf{v})\,\mathrm{d}\Omega = \int_T q^h \operatorname{div}\mathbf{v}\,\mathrm{d}\Omega \tag{290}$$

for every linear polynomial $q^h = q_0 + q_1$. In fact, using (286), (289), and the obvious fact that $\int_T \operatorname{div}\mathbf{v}_{b,T}(\mathbf{v})$

$\mathrm{d}\Omega = 0$, we have

$$\begin{aligned}\int_T q^h \operatorname{div}(\Pi_{h,T}\mathbf{v})\,\mathrm{d}\Omega &= \int_T (q_0+q_1)\operatorname{div}(\mathbf{v}_{b,T}(\mathbf{v})\\ &\quad + \Pi_{1,T}\mathbf{v})\,\mathrm{d}\Omega = \int_T q_0 \operatorname{div}(\mathbf{v}_{b,T}(\mathbf{v}) + \Pi_{1,T}\mathbf{v})\,\mathrm{d}\Omega\\ &\quad + \int_T q_1 \operatorname{div}(\mathbf{v}_{b,T}(\mathbf{v}) + \Pi_{1,T}\mathbf{v})\,\mathrm{d}\Omega\\ &= \int_T q_0 \operatorname{div}(\Pi_{1,T}\mathbf{v})\,\mathrm{d}\Omega + \int_T q_1 \operatorname{div}\mathbf{v}\,\mathrm{d}\Omega\\ &= \int_T q_0 \operatorname{div}\mathbf{v}\,\mathrm{d}\Omega + \int_T q_1 \operatorname{div}\mathbf{v}\,\mathrm{d}\Omega\\ &= \int_T q^h \operatorname{div}\mathbf{v}\,\mathrm{d}\Omega \end{aligned} \tag{291}$$

Hence, the operator $\widehat{\Pi}_h$ arising from the global linear operator

$$\Pi_h\mathbf{v} = \sum_{T\in\mathcal{T}_h} \left(\mathbf{v}_{b,T}(\mathbf{v}) + \Pi_{1,T}\mathbf{v}\right) = \mathbf{v}_b(\mathbf{v}) + \Pi_1\mathbf{v} = \mathbf{N}_i^{\mathbf{u}}\hat{v}_i$$

fulfills condition (273).

Remark 15. Conditions (288) reveal that the operator Π_1 (built by means of the local contributions $\Pi_{1,T}$) exploits, in particular, the point values of $\mathbf{v}$ at all the vertexes of the triangles in $\mathcal{T}_h$. However, this definition makes no sense for an arbitrary $\mathbf{v} \in \mathcal{X}$, since functions in $\mathcal{X}$ are not necessarily continuous. To overcome this problem, one should define a more sophisticated operator Π_1 for instance, taking advantage of an averaging procedure. More precisely, one could define the function $\Pi_1\mathbf{v}$ as the unique piecewise quadratic polynomial such that $\Pi_1\mathbf{v}|_{\partial\Omega} = 0$ and

$$\begin{cases} \Pi_1\mathbf{v}(V) = \dfrac{1}{\mathrm{Area}(D(V))}\displaystyle\int_{D(V)} \mathbf{v}\,\mathrm{d}\Omega \\ \Pi_1\mathbf{v}(M) = \dfrac{3}{2|e_M|}\displaystyle\int_{e_M} \mathbf{v}\,\mathrm{d}s - \dfrac{\Pi_1\mathbf{v}(V_{M_1}) + \Pi_1\mathbf{v}(V_{M_2})}{4} \end{cases} \tag{292}$$

Above, V is any internal vertex of triangles in $\mathcal{T}_h$ and $D(V)$ is the union of the triangles having V as a vertex. Moreover, e_M is any internal edge having M as midpoint and V_{M_1}, V_{M_2} as endpoints. In this case, it is possible to prove that for the resulting $\widehat{\Pi}_h$, the very important property (272) holds with $C_{\widehat{\Pi}}$ independent of h.

Remark 16. It is interesting to observe that condition (283) and (287) suggest the following important fact about the discretization of the Stokes problem. Any reasonable discontinuous pressure approximation contains at least all the piecewise constant functions: relations (283) and (287) show that having some velocity degrees of freedom associated with the triangle edges greatly helps in proving the *inf–sup* condition.

5.1.2 *Verfürth's trick*

We now describe another technique for proving the *inf–sup* condition, which can be profitably used when elements with continuous pressure interpolation are considered: the so-called *Verfürth's trick* (see Verfürth, 1984). We begin by noting that, because of to the pressure continuity, it holds

$$\begin{aligned}\hat{\mathbf{z}}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\hat{\mathbf{q}} &= -\int_\Omega \left(N_r^p\hat{q}_r\right)\ \operatorname{div}\left(\mathbf{N}_j^{\mathbf{u}}\hat{z}_j\right)\ \mathrm{d}\Omega\\ &= \int_\Omega (\nabla N_r^p\hat{q}_r)\cdot\mathbf{N}_j^{\mathbf{u}}\hat{z}_j\,\mathrm{d}\Omega\end{aligned} \tag{293}$$

for every $\hat{\mathbf{z}} \in \mathbf{X}$ and $\hat{\mathbf{q}} \in \mathbf{Y}$. In some cases, it is much easier to use the form (293) and prove a modified version of the *inf–sup* condition (266) with a norm for $\mathbf{Y}$ different from the one defined in (270) and involving the pressure gradients; see Bercovier and Pironneau (1977) and Glowinski and Pironneau (1979). More precisely, given a mesh $\mathcal{T}_h$, we introduce in $\mathbf{Y}$ the norm

$$\|\hat{\mathbf{p}}\|_{Y_*} := \left(\sum_{K\in\mathcal{T}_h} h_K^2\int_K |\nabla N_r^p\hat{p}_r|^2\,\mathrm{d}\Omega\right)^{1/2} \tag{294}$$

where h_K denotes the diameter of the generic element K.

The key point of Verfürth's trick is a smart use of the properties of interpolation operator in order to prove that the *inf–sup* condition with the norm Y_* implies the usual one. Indeed, we have the following result.

Proposition 5. *Suppose that*

(H1) *for every velocity* $\mathbf{v} \in \mathcal{X}$ *there exists a discrete velocity* $\mathbf{v}_I = \mathbf{N}_j^{\mathbf{u}}\hat{v}_j^I$ *such that*

$$\left(\sum_{K\in\mathcal{T}_h} h_K^{-2}\int_K |\mathbf{N}_j^{\mathbf{u}}\hat{v}_j^I - \mathbf{v}|^2\,\mathrm{d}\Omega\right)^{1/2} \le c_0\|\mathbf{v}\|_{\mathcal{X}} \tag{295}$$

$$\|\hat{\mathbf{v}}_I\|_X \le c_1\|\mathbf{v}\|_{\mathcal{X}} \tag{296}$$

with c_0, c_1 *independent of* h *and* $\mathbf{v}$*;*

(H2) *there exists a constant* $\beta_* > 0$ *independent of* h *such that*

$$\forall\ \hat{\mathbf{q}} \in \mathbf{Y} \quad \sup_{\hat{\mathbf{z}}\in\mathbf{X}\setminus\{\mathbf{0}\}} \frac{\hat{\mathbf{z}}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\hat{\mathbf{q}}}{\|\hat{\mathbf{z}}\|_X} \ge \beta_*\|\hat{\mathbf{q}}\|_{Y_*} \tag{297}$$

(i.e. the **inf–sup** *condition holds with the modified* $\mathbf{Y}$*-norm (294)).*

Then the **inf–sup** *condition (with respect to the original norm (270))*

$$\forall\ \hat{\mathbf{q}} \in \mathbf{Y} \quad \sup_{\hat{\mathbf{z}}\in\mathbf{X}\setminus\{\mathbf{0}\}} \frac{\hat{\mathbf{z}}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\hat{\mathbf{q}}}{\|\hat{\mathbf{z}}\|_X} \ge \beta\|\hat{\mathbf{q}}\|_Y \qquad (298)$$

is satisfied with β *independent of* h.

Proof. Given $\hat{\mathbf{q}} \in \mathbf{Y}$, we observe that using (296) and (293) it holds

$$\sup_{\hat{\mathbf{z}}\in\mathbf{X}\setminus\{\mathbf{0}\}} \frac{\hat{\mathbf{z}}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\hat{\mathbf{q}}}{\|\hat{\mathbf{z}}\|_X} \ge \sup_{\mathbf{v}\in\mathcal{X}\setminus\{\mathbf{0}\}} \frac{\hat{\mathbf{v}}_I^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\hat{\mathbf{q}}}{\|\hat{\mathbf{v}}_I\|_X} \ge \sup_{\mathbf{v}\in\mathcal{X}\setminus\{\mathbf{0}\}} \frac{\hat{\mathbf{v}}_I^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\hat{\mathbf{q}}}{c_1\|\mathbf{v}\|_{\mathcal{X}}}$$
$$= \sup_{\mathbf{v}\in\mathcal{X}\setminus\{\mathbf{0}\}} \frac{\int_\Omega (\nabla N_r^p \hat{q}_r)\cdot \mathbf{N}_j^{\mathbf{u}}\hat{v}_j^I\,\mathrm{d}\Omega}{c_1\|\mathbf{v}\|_{\mathcal{X}}} \qquad (299)$$

Furthermore, from Theorem 4, there exists $\mathbf{w} \in \mathcal{X}$ such that

$$-\frac{\int_\Omega \operatorname{div}\mathbf{w}(N_r^p\hat{q}_r)\,\mathrm{d}\Omega}{\|\mathbf{w}\|_{\mathcal{X}}} = \frac{\int_\Omega (\nabla N_r^p\hat{q}_r)\cdot\mathbf{w}\,\mathrm{d}\Omega}{\|\mathbf{w}\|_{\mathcal{X}}} \ge \beta_c\|\hat{\mathbf{q}}\|_Y \qquad (300)$$

For such a velocity $\mathbf{w}$ and the corresponding *discrete* velocity $\mathbf{w}_I = \mathbf{N}_j^{\mathbf{u}}\hat{w}_j^I$, we obviously have

$$\sup_{\mathbf{v}\in\mathcal{X}\setminus\{\mathbf{0}\}} \frac{\int_\Omega (\nabla N_r^p\hat{q}_r)\cdot\mathbf{N}_j^{\mathbf{u}}\hat{v}_j^I\,\mathrm{d}\Omega}{c_1\|\mathbf{v}\|_{\mathcal{X}}} \ge \frac{\int_\Omega (\nabla N_r^p\hat{q}_r)\cdot\mathbf{N}_j^{\mathbf{u}}\hat{w}_j^I\,\mathrm{d}\Omega}{c_1\|\mathbf{w}\|_{\mathcal{X}}} \qquad (301)$$

Subtracting and adding $\mathbf{w}$, we obtain

$$\frac{\int_\Omega (\nabla N_r^p\hat{q}_r)\cdot\mathbf{N}_j^{\mathbf{u}}\hat{w}_j^I\,\mathrm{d}\Omega}{c_1\|\mathbf{w}\|_{\mathcal{X}}} = \frac{\int_\Omega (\nabla N_r^p\hat{q}_r)\cdot(\mathbf{N}_j^{\mathbf{u}}\hat{w}_j^I - \mathbf{w})\,\mathrm{d}\Omega}{c_1\|\mathbf{w}\|_{\mathcal{X}}}$$
$$+ \frac{\int_\Omega (\nabla N_r^p\hat{q}_r)\cdot\mathbf{w}\,\mathrm{d}\Omega}{c_1\|\mathbf{w}\|_{\mathcal{X}}} \qquad (302)$$

To treat the first term in the right-hand side of (302), we observe that using (295) and recalling (294), we have

$$-\int_\Omega (\nabla N_r^p\hat{q}_r)\cdot(\mathbf{N}_j^{\mathbf{u}}\hat{w}_j^I - \mathbf{w})\,\mathrm{d}\Omega$$
$$= -\sum_{K\in\mathcal{T}_h}\int_K (\nabla N_r^p\hat{q}_r)\cdot(\mathbf{N}_j^{\mathbf{u}}\hat{w}_j^I - \mathbf{w})\,\mathrm{d}\Omega$$
$$= -\sum_{K\in\mathcal{T}_h}\int_K h_K(\nabla N_r^p\hat{q}_r)\cdot h_K^{-1}(\mathbf{N}_j^{\mathbf{u}}\hat{w}_j^I - \mathbf{w})\,\mathrm{d}\Omega$$
$$\le \left(\sum_{K\in\mathcal{T}_h} h_K^2\int_K |\nabla N_r^p\hat{q}_r|^2\,\mathrm{d}\Omega\right)^{1/2}$$
$$\times\left(\sum_{K\in\mathcal{T}_h} h_K^{-2}\int_K |\mathbf{N}_j^{\mathbf{u}}\hat{w}_j^I - \mathbf{w}|^2\,\mathrm{d}\Omega\right)^{1/2}$$
$$\le c_0\|\hat{\mathbf{q}}\|_{Y_*}\|\mathbf{w}\|_{\mathcal{X}} \qquad (303)$$

which gives

$$\int_\Omega (\nabla N_r^p\hat{q}_r)\cdot(\mathbf{N}_j^{\mathbf{u}}\hat{w}_j^I - \mathbf{w})\,\mathrm{d}\Omega \ge -c_0\|\hat{\mathbf{q}}\|_{Y_*}\|\mathbf{w}\|_{\mathcal{X}} \qquad (304)$$

Therefore, we get

$$\frac{\int_\Omega (\nabla N_r^p\hat{q}_r)\cdot(\mathbf{N}_j^{\mathbf{u}}\hat{w}_j^I - \mathbf{w})\,\mathrm{d}\Omega}{c_1\|\mathbf{w}\|_{\mathcal{X}}} \ge -\frac{c_0}{c_1}\|\hat{\mathbf{q}}\|_{Y_*} \qquad (305)$$

For the second term in the right-hand side of (302), we notice that (cf. (300))

$$\frac{\int_\Omega (\nabla N_r^p\hat{q}_r)\cdot\mathbf{w}\,\mathrm{d}\Omega}{c_1\|\mathbf{w}\|_{\mathcal{X}}} \ge \frac{\beta_c}{c_1}\|\hat{\mathbf{q}}\|_Y \qquad (306)$$

Therefore, from (299), (301), (302), (305), and (306), we obtain

$$\sup_{\hat{\mathbf{z}}\in\mathbf{X}\setminus\{\mathbf{0}\}} \frac{\hat{\mathbf{z}}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\hat{\mathbf{q}}}{\|\hat{\mathbf{z}}\|_X} \ge \frac{\beta_c}{c_1}\|\hat{\mathbf{q}}\|_Y - \frac{c_0}{c_1}\|\hat{\mathbf{q}}\|_{Y_*} \qquad (307)$$

We now multiply the modified *inf–sup* condition (297) by $c_0/(\beta_* c_1)$ to get

$$\frac{c_0}{\beta_* c_1}\sup_{\hat{\mathbf{z}}\in\mathbf{X}\setminus\{\mathbf{0}\}} \frac{\hat{\mathbf{z}}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\hat{\mathbf{q}}}{\|\hat{\mathbf{z}}\|_X} \ge \frac{c_0}{c_1}\|\hat{\mathbf{q}}\|_{Y_*} \qquad (308)$$

By adding (307) and (308), we finally have

$$\left(1 + \frac{c_0}{\beta_* c_1}\right)\sup_{\hat{\mathbf{z}}\in\mathbf{X}\setminus\{\mathbf{0}\}} \frac{\hat{\mathbf{z}}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\hat{\mathbf{q}}}{\|\hat{\mathbf{z}}\|_X} \ge \frac{\beta_c}{c_1}\|\hat{\mathbf{q}}\|_Y \qquad (309)$$

that is, the *inf–sup* condition (298) holds with $\beta = (\beta_c/c_1)(1 + (c_0/\beta_* c_1))^{-1}$. □

Remark 17. We notice that hypothesis (H1) of Proposition 5 is not very restrictive. Indeed, given a velocity $\mathbf{v} \in \mathcal{X}$, the corresponding $\mathbf{v}_I$ can be chosen as a suitable discrete velocity interpolating $\mathbf{v}$, and (295) and (296) are both satisfied basically for every element of practical interest (see e.g. Brezzi and Fortin (1991) and Ciarlet (1978) for more details).

The Verfürth trick was originally applied to the Hood–Taylor element depicted in Figure 9 (see Verfürth, 1984), but it was soon recognized as a valuable instrument for analyzing all continuous pressure elements. Here we show how to use it for the analysis of the MINI element (whose original proof was given using Fortin's trick in Arnold, Brezzi and Fortin, 1984).

• *The MINI element*. We now give a hint on how to verify hypothesis (H2) of Proposition 5 for the MINI element (cf. Figure 8).

For a generic $\hat{\mathbf{q}} \in \mathbf{Y}$, we take its reconstructed discrete pressure $q^h = N_r^p \hat{q}_r$. Since q^h is a piecewise linear and continuous function, it follows that $\nabla q^h = \nabla N_r^p \hat{q}_r$ is a well-defined piecewise constant vector field. We now construct a discrete (bubble-type) velocity $\mathbf{v}^h = \mathbf{N}_j^{\mathbf{u}} \hat{v}_j$, defined on each triangle $T \in \mathcal{T}_h$ as

$$\mathbf{v}^h = h_T^2 b_T \nabla q^h \tag{310}$$

where b_T is the usual cubic bubble (i.e. in area coordinates, $b_T = 27\lambda_1\lambda_2\lambda_3$). Recalling (293) and using (310), we then obtain

$$\begin{aligned}\hat{\mathbf{v}}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\hat{\mathbf{q}} &= \int_\Omega (\nabla N_r^p \hat{q}_r) \cdot \mathbf{N}_j^{\mathbf{u}} \hat{v}_j \, \mathrm{d}\Omega = \int_\Omega \nabla q^h \cdot \mathbf{v}^h \, \mathrm{d}\Omega \\ &= \sum_{T \in \mathcal{T}_h} h_T^2 \int_T |\nabla N_r^p \hat{q}_r|^2 b_T \, \mathrm{d}\Omega \end{aligned} \tag{311}$$

It is easy to show that for *regular meshes* (roughly: for meshes that do not contain 'too thin' elements, see e.g. Ciarlet (1978) for a precise definition), there exists a constant $C_1 > 0$, independent of h, such that

$$\forall\, T \in \mathcal{T}_h \qquad \int_T |\nabla N_r^p \hat{q}_r|^2 b_T \, \mathrm{d}\Omega \ge C_1 \int_T |\nabla N_r^p \hat{q}_r|^2 \, \mathrm{d}\Omega \tag{312}$$

Therefore, from (311), (312), and (294) we get

$$\hat{\mathbf{v}}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\hat{\mathbf{q}} \ge C_1 \sum_{T \in \mathcal{T}_h} h_T^2 \int_T |\nabla N_r^p \hat{q}_r|^2 \, \mathrm{d}\Omega = C_1 \|\hat{\mathbf{q}}\|_{Y_*}^2 \tag{313}$$

Furthermore, using standard scaling arguments (cf. Brezzi and Fortin, 1991), it is possible to prove that there exists $C_2 > 0$ independent of h such that

$$\begin{aligned}\|\hat{\mathbf{v}}\|_X = \|\mathbf{v}^h\|_{\mathcal{X}} &\le C_2 \left(\sum_{T \in \mathcal{T}_h} h_T^2 \int_T |\nabla N_r^p \hat{q}_r|^2 \, \mathrm{d}\Omega \right)^{1/2} \\ &= C_2 \|\hat{\mathbf{q}}\|_{Y_*}\end{aligned} \tag{314}$$

Hence, estimates (313) and (314) imply

$$\frac{\hat{\mathbf{v}}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\hat{\mathbf{q}}}{\|\hat{\mathbf{v}}\|_X} \ge \frac{C_1}{C_2} \|\hat{\mathbf{q}}\|_{Y_*} \tag{315}$$

and condition (297) then follows with $\beta_* = C_1/C_2$, since

$$\sup_{\hat{\mathbf{z}} \in \mathbf{X} \setminus \{\mathbf{0}\}} \frac{\hat{\mathbf{z}}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\hat{\mathbf{q}}}{\|\hat{\mathbf{z}}\|_X} \ge \frac{\hat{\mathbf{v}}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\hat{\mathbf{q}}}{\|\hat{\mathbf{v}}\|_X} \tag{316}$$

5.2 Appendix — error estimates

In this brief Appendix, we present the guidelines to obtain error estimates, once the stability conditions have been established. We only consider the easiest case of conforming schemes (i.e. when the velocity is approximated by means of continuous functions). We refer to Brezzi (1974), Brezzi and Fortin (1991), and Braess (1997) for more details, as well as for the analysis of more complicated situations involving non-conforming approximations (such as the $P_1^{\mathrm{NC}} - P_0$ element (cf. Figure 15)).

Before proceeding, we recall that for the Stokes problem with our choices of norms, we have $M_a = 1$, $M_b = \sqrt{(d/\mu)}$, and $\alpha = 1$, no matter what the approximations of velocity and pressure are. However, in the subsequent discussion, we will not substitute these values into the estimates, in order to facilitate the extension of the analysis to other problems. We also notice that, on the contrary, the relevant constant β does depend on the choice of the interpolating functions. We have the following result.

Theorem 5. *Let* $(\mathbf{u}, p)$ *be the solution of problem (263) and suppose there exist discrete velocity and pressure*

$$\mathbf{u}_I = \mathbf{N}_i^{\mathbf{u}} \hat{u}_i^I, \quad p_I = N_r^p \hat{p}_r^I \tag{317}$$

such that

$$\|\mathbf{u} - \mathbf{u}_I\|_{\mathcal{X}} \le C h^{k_u}, \quad k_u > 0 \tag{318}$$

$$\|p - p_I\|_{\mathcal{Y}} \le C h^{k_p}, \quad k_p > 0 \tag{319}$$

If $(\hat{\mathbf{u}}, \hat{\mathbf{p}})$ *is the solution of the discrete problem (264), then, setting* $\mathbf{u}^h = \mathbf{N}_i^{\mathbf{u}} \hat{u}_i$ *and* $p^h = N_r^p \hat{p}_r$*, it holds*

$$\|\mathbf{u} - \mathbf{u}^h\|_{\mathcal{X}} + \|p - p^h\|_{\mathcal{Y}} \le C h^k \tag{320}$$

with $k = \min\{k_u, k_p\}$.

Proof. For $\mathbf{u}_I$ and p_I as in (317), we set $\hat{\mathbf{u}}_I = (\hat{u}_i^I)_{i=1}^n \in \mathbf{X}$ and $\hat{\mathbf{p}}_I = (\hat{p}_r^I)_{r=1}^m \in \mathbf{Y}$. Taking into account that

$$\left\{ \begin{matrix} \hat{\mathbf{u}} \\ \hat{\mathbf{p}} \end{matrix} \right\}$$

is the solution of the discretized problem (264), we obtain that

$$\left\{ \begin{matrix} \hat{\mathbf{u}} - \hat{\mathbf{u}}_I \\ \hat{\mathbf{p}} - \hat{\mathbf{p}}_I \end{matrix} \right\}$$

solves

$$\begin{bmatrix} \mathbf{A} & \mathbf{B}^{\mathrm{T}} \\ \mathbf{B} & \mathbf{0} \end{bmatrix} \begin{Bmatrix} \hat{\mathbf{u}} - \hat{\mathbf{u}}_I \\ \hat{\mathbf{p}} - \hat{\mathbf{p}}_I \end{Bmatrix} = \begin{Bmatrix} \mathbf{f} - \mathbf{A}\hat{\mathbf{u}}_I - \mathbf{B}^{\mathrm{T}}\hat{\mathbf{p}}_I \\ -\mathbf{B}\hat{\mathbf{u}}_I \end{Bmatrix} \tag{321}$$

Choosing as (admissible) velocity and pressure variations the interpolating shape functions, from (263), we have

$$\mathbf{f}|_i = \mu \int_\Omega \nabla \mathbf{N}_i^{\mathbf{u}} : \nabla \mathbf{u}\, \mathrm{d}\Omega - \int_\Omega \operatorname{div}(\mathbf{N}_i^{\mathbf{u}}) p \, \mathrm{d}\Omega \qquad i = 1, \ldots, n \tag{322}$$

and

$$\int_\Omega N_s^p \operatorname{div} \mathbf{u} \, \mathrm{d}\Omega = 0 \qquad s = 1, \ldots, m \tag{323}$$

Hence, system (321) may be written as

$$\begin{bmatrix} \mathbf{A} & \mathbf{B}^{\mathrm{T}} \\ \mathbf{B} & \mathbf{0} \end{bmatrix} \begin{Bmatrix} \hat{\mathbf{u}} - \hat{\mathbf{u}}_I \\ \hat{\mathbf{p}} - \hat{\mathbf{p}}_I \end{Bmatrix} = \begin{Bmatrix} \widetilde{\mathbf{f}} \\ \widetilde{\mathbf{g}} \end{Bmatrix} \tag{324}$$

where

$$\widetilde{\mathbf{f}}|_i := \mu \int_\Omega \nabla \mathbf{N}_i^{\mathbf{u}} : \nabla \left(\mathbf{u} - \mathbf{u}_I\right) \mathrm{d}\Omega - \int_\Omega \operatorname{div}(\mathbf{N}_i^{\mathbf{u}}) \left(p - p_I\right) \mathrm{d}\Omega \quad i = 1, \ldots, n \tag{325}$$

and

$$\widetilde{\mathbf{g}}|_r := -\int_\Omega N_r^p \operatorname{div}(\mathbf{u} - \mathbf{u}_I) \, \mathrm{d}\Omega \qquad r = 1, \ldots, m \tag{326}$$

Applying Theorem 1, we thus obtain

$$\|\hat{\mathbf{u}} - \hat{\mathbf{u}}_I\|_X \leq \frac{1}{\alpha}\|\widetilde{\mathbf{f}}\|_F + \frac{M_a^{1/2}}{\alpha^{1/2}\beta}\|\widetilde{\mathbf{g}}\|_G \tag{327}$$

$$\|\hat{\mathbf{p}} - \hat{\mathbf{p}}_I\|_Y \leq \left(\frac{1}{\beta} + \frac{M_a^{1/2}}{\alpha^{1/2}\beta}\right)\|\widetilde{\mathbf{f}}\|_F + \frac{M_a}{\beta^2}\|\widetilde{\mathbf{g}}\|_G \tag{328}$$

We proceed by estimating the dual norms $\|\widetilde{\mathbf{f}}\|_F$ and $\|\widetilde{\mathbf{g}}\|_G$. Since for every $\hat{\mathbf{v}} = (\hat{v}_i)_{i=1}^n$,

$$\hat{\mathbf{v}}^{\mathrm{T}}\widetilde{\mathbf{f}} = \mu \int_\Omega \nabla\left(\mathbf{N}_i^{\mathbf{u}}\hat{v}_i\right) : \nabla\left(\mathbf{u} - \mathbf{u}_I\right) \mathrm{d}\Omega - \int_\Omega \operatorname{div}(\mathbf{N}_i^{\mathbf{u}}\hat{v}_i)\left(p - p_I\right) \mathrm{d}\Omega \leq M_a\|\hat{\mathbf{v}}\|_X\|\mathbf{u} - \mathbf{u}_I\|_{\mathcal{X}} + M_b\|\hat{\mathbf{v}}\|_X\|p - p_I\|_{\mathcal{Y}} \tag{329}$$

we obtain

$$\frac{\hat{\mathbf{v}}^{\mathrm{T}}\widetilde{\mathbf{f}}}{\|\hat{\mathbf{v}}\|_X} \leq M_a\|\mathbf{u} - \mathbf{u}_I\|_{\mathcal{X}} + M_b\|p - p_I\|_{\mathcal{Y}} \tag{330}$$

which gives (cf. the dual norm definition (104))

$$\|\widetilde{\mathbf{f}}\|_F \leq M_a\|\mathbf{u} - \mathbf{u}_I\|_{\mathcal{X}} + M_b\|p - p_I\|_{\mathcal{Y}} \tag{331}$$

Analogously, for every $\hat{\mathbf{q}} = (\hat{q}_r)_{r=1}^m$, we get

$$\hat{\mathbf{q}}^{\mathrm{T}}\widetilde{\mathbf{g}} = -\int_\Omega \left(N_r^p\hat{q}_r\right)\operatorname{div}(\mathbf{u} - \mathbf{u}_I) \, \mathrm{d}\Omega \leq M_b\|\hat{\mathbf{q}}\|_Y\|\mathbf{u} - \mathbf{u}_I\|_{\mathcal{X}} \tag{332}$$

and therefore we have

$$\|\widetilde{\mathbf{g}}\|_G \leq M_b\|\mathbf{u} - \mathbf{u}_I\|_{\mathcal{X}} \tag{333}$$

From (327), (328), (331), and (333) we have

$$\|\hat{\mathbf{u}} - \hat{\mathbf{u}}_I\|_X \leq \left(\frac{M_a}{\alpha} + \frac{M_a^{1/2}M_b}{\alpha^{1/2}\beta}\right)\|\mathbf{u} - \mathbf{u}_I\|_{\mathcal{X}} + \frac{M_b}{\alpha}\|p - p_I\|_{\mathcal{Y}} \tag{334}$$

$$\|\hat{\mathbf{p}} - \hat{\mathbf{p}}_I\|_Y \leq \left(\frac{M_a}{\beta} + \frac{M_a^{3/2}}{\alpha^{1/2}\beta} + \frac{M_aM_b}{\beta^2}\right)\|\mathbf{u} - \mathbf{u}_I\|_{\mathcal{X}} + M_b\left(\frac{1}{\beta} + \frac{M_a^{1/2}}{\alpha^{1/2}\beta}\right)\|p - p_I\|_{\mathcal{Y}} \tag{335}$$

Observing that by triangle inequality and Remark 13, it holds

$$\|\mathbf{u} - \mathbf{u}^h\|_{\mathcal{X}} \leq \|\mathbf{u} - \mathbf{u}_I\|_{\mathcal{X}} + \|\mathbf{u}_I - \mathbf{N}_i^{\mathbf{u}}\hat{u}_i\|_{\mathcal{X}} = \|\mathbf{u} - \mathbf{u}_I\|_{\mathcal{X}} + \|\hat{\mathbf{u}} - \hat{\mathbf{u}}_I\|_X \tag{336}$$

and

$$\|p - p^h\|_{\mathcal{Y}} \leq \|p - p_I\|_{\mathcal{Y}} + \|p_I - N_r^p\hat{p}_r\|_{\mathcal{Y}} = \|p - p_I\|_{\mathcal{Y}} + \|\hat{\mathbf{p}} - \hat{\mathbf{p}}_I\|_Y \tag{337}$$

from (334) and (335), we get the error estimates

$$\|\mathbf{u} - \mathbf{u}^h\|_{\mathcal{X}} \leq \left(1 + \frac{M_a}{\alpha} + \frac{M_a^{1/2}M_b}{\alpha^{1/2}\beta}\right)\|\mathbf{u} - \mathbf{u}_I\|_{\mathcal{X}} + \frac{M_b}{\alpha}\|p - p_I\|_{\mathcal{Y}} \tag{338}$$

$$\|p - p^h\|_{\mathcal{Y}} \leq \left(\frac{M_a}{\beta} + \frac{M_a^{3/2}}{\alpha^{1/2}\beta} + \frac{M_aM_b}{\beta^2}\right)\|\mathbf{u} - \mathbf{u}_I\|_{\mathcal{X}} + \left(1 + \frac{M_b}{\beta} + \frac{M_a^{1/2}M_b}{\alpha^{1/2}\beta}\right)\|p - p_I\|_{\mathcal{Y}} \tag{339}$$

We notice that the constant M_b, which did not appear in the stability estimates, has now come into play. Furthermore, using (318) and (319), from (338) and (339), we infer

$$\|\mathbf{u}-\mathbf{u}^h\|_{\mathcal{X}} + \|p - p^h\|_{\mathcal{Y}} \le Ch^k \qquad (340)$$

with $k = \min\{k_u, k_p\}$ and $C = C(\alpha, \beta, M_a, M_b)$ independent of h. □

Remark 18. A crucial step in obtaining error estimate (340) is to prove the bounds (cf. (331) and (333))

$$\|\widetilde{\mathbf{f}}\|_F \le M_a\|\mathbf{u}-\mathbf{u}_I\|_{\mathcal{X}} + M_b\|p - p_I\|_{\mathcal{Y}} \qquad (341)$$

$$\|\widetilde{\mathbf{g}}\|_G \le M_b\|\mathbf{u}-\mathbf{u}_I\|_{\mathcal{X}} \qquad (342)$$

where $\widetilde{\mathbf{f}}$ and $\widetilde{\mathbf{g}}$ are defined by (325) and (326) respectively. The estimates above result from a suitable choice of the norms for $\mathbf{X} \equiv \mathbb{R}^n$, $\mathbf{Y} \equiv \mathbb{R}^m$, $\mathbf{F} \equiv \mathbb{R}^n$, and $\mathbf{G} \equiv \mathbb{R}^m$. In fact, by choosing for $\mathbf{X}$ the norm (269) and for $\mathbf{F}$ the corresponding dual norm, we can get (341), as highlighted by (329). Similarly, by choosing for $\mathbf{Y}$ the norm (270) and for $\mathbf{G}$ the corresponding dual norm, we can obtain (341) (cf. (332)).

Remark 19. The discrete functions $\mathbf{u}_I$ and p_I in (317) are typically chosen as follows:

• $\mathbf{u}_I$ is the nodal interpolated of $\mathbf{u}$.
Therefore,

$$\mathbf{u}_I = \mathbf{N}_i^{\mathbf{u}}\hat{u}_i^I \qquad (343)$$

where $\hat{\mathbf{u}}_I = (\hat{u}_i^I)_{i=1}^n$ is the vector containing the nodal values of $\mathbf{u}$.

• p_I is the projection of p over the pressure approximation space. Therefore,

$$p_I = N_r^p\,\hat{p}_r^I \qquad (344)$$

where the vector $\hat{\mathbf{p}}_I = (\hat{p}_r^I)_{r=1}^m$ is uniquely determined by the following set of m equations

$$\int_\Omega N_s^p(N_r^p\,\hat{p}_r^I)\,\mathrm{d}\Omega = \int_\Omega N_s^p\,p\,\mathrm{d}\Omega \qquad s = 1, \ldots, m \quad (345)$$

For regular solution ($\mathbf{u}$, p), standard approximation results (see e.g. Ciarlet, 1978) allow to determine the exponents k_u and k_p entering in estimates (318) and (319) in terms of the selected approximation spaces for the velocity and the pressure fields. For instance, when considering the Crouzeix–Raviart element (cf. Figure 12), we have $k_u = k_p = 2$. Hence, Theorem 5 shows that the discretization error is $O(h^2)$.

6 RELATED CHAPTERS

(*See also* **Chapter 4**, **Chapter 15 of this Volume**; **Chapter 2, Volume 3**).

REFERENCES

Alotto P and Perugia I. Mixed finite element methods and tree-cotree implicit condensation. *Calcolo* 1999; **36**:233–248.

Amara M and Thomas JM. Equilibrium finite elements for the linear elastic problem. *Numer. Math.* 1979; **33**:367–383.

Arnold DN. Discretization by finite elements of a model parameter dependent problem. *Numer. Math.* 1981; **37**:405–421.

Arnold DN and Brezzi F. Mixed and non-conforming finite element methods: implementation, post-processing and error estimates. *Math. Modell. Numer. Anal.* 1985; **19**:7–35.

Arnold DN and Winther R. Mixed finite elements for elasticity. *Numer. Math.* 2002; **42**:401–419.

Arnold D, Brezzi F and Douglas J. PEERS: a new mixed finite element for plane elasticity. *Jpn. J. Appl. Math.* 1984; **1**:347–367.

Arnold DN, Brezzi F and Fortin M. A stable finite element for the Stokes equations. *Calcolo* 1984; **21**:337–344.

Arnold DN, Douglas J and Gupta CP. A family of higher order mixed finite element methods for plane elasticity. *Numer. Math.* 1984; **45**:1–22.

Atluri SN, Gallagher RH and Zienkiewicz OC. *Hybrid and Mixed Finite Element Methods.* Wiley: New York, 1983.

Auricchio F, Beirão da Veiga L, Lovadina C and Reali A. Triangular enhanced strain elements for plane linear elasticity. *Comput. Methods Appl. Mech. Eng.* 200x; submitted.

Baiocchi C and Brezzi F. Stabilization of unstable methods. In *Problemi attuali dell'Analisi e della Fisica Matematica*, Ricci PE (ed.). Università La Sapienza: Roma, 1993; 59–63.

Baiocchi C, Brezzi F and Franca LP. Virtual bubbles and Ga.L.S. *Comput. Methods Appl. Mech. Eng.* 1993; **105**:125–141.

Baranger J, Maitre J-F and Oudin F. Connection between finite volume and mixed finite element methods. *RAIRO Modél. Math. Anal. Numér.* 1996; **30**:445–465.

Bathe KJ. *Finite Element Procedures*. Prentice Hall: Englewood Cliffs, NJ, 1996.

Becker EB, Carey GF and Oden JT. *Finite Elements. An Introduction*. Prentice Hall: Englewood Cliffs, NJ, 1981.

Behr MA, Franca LP and Tezduyar TE. Stabilized finite element methods for the velocity-pressure-stress formulation of incompressible flows. *Comput. Methods Appl. Mech. Eng.* 1993; **104**:31–48.

Belytschko T, Liu WK and Moran B. *Non Linear Finite Elements for Continua and Structures*. John Wiley & Sons: New York, 2000.

Bercovier M and Pironneau OA. Error estimates for finite element method solution of the Stokes problem in primitive variables. *Numer. Math.* 1977; **33**:211–224.

Boffi D. Stability of higher order triangular Hood–Taylor methods for the stationary Stokes equations. *Math. Models Methods Appl. Sci.* 1994; **4**(2):223–235.

Boffi D. Minimal stabilizations of the $P_{k+1} - P_k$ approximation of the stationary Stokes equations. *Math. Models Methods Appl. Sci.* 1995; **5**(2):213–224.

Boffi D. Three-dimensional finite element methods for the Stokes problem. *SIAM J. Numer. Anal.* 1997; **34**:664–670.

Boffi D and Lovadina C. Analysis of new augmented Lagrangian formulations for mixed finite element schemes. *Numer. Math.* 1997; **75**:405–419.

Bonet J and Wood RD. *Nonlinear Continuum Mechanics for Finite Element Analysis*. Cambridge University Press: Cambridge, UK, 1997.

Braess D. *Finite Elements. Theory, Fast Solvers and Applications in Solid Mechanics*. Cambridge University Press, 1997.

Braess D. Enhanced assumed strain elements and locking in membrane problems. *Comput. Methods Appl. Mech. Eng.* 1998; **165**:155–174.

Brenner SC and Scott LR. *The Mathematical Theory of Finite Element Methods*. Springer-Verlag: New York, 1994.

Brezzi F. On the existence, uniqueness and approximation of saddle point problems arising from Lagrangian multipliers. *RAIRO Anal. Numer.* 1974; **8**:129–151.

Brezzi F. and Douglas J. Stabilized mixed methods for the Stokes problem. *Numer. Math.* 1988; **53**:225–235.

Brezzi F and Falk RS. Stability of higher-order Hood–Taylor Methods. *SIAM J. Numer. Anal.* 1991; **28**:581–590.

Brezzi F and Fortin M. *Mixed and Hybrid Finite Element Methods*. Springer-Verlag: New York, 1991.

Brezzi F and Fortin M. A minimal stabilisation procedure for mixed finite element methods. *Numer. Math.* 2001; **89**: 457–492.

Brezzi F and Pitkäranta J. On the stabilization of finite element approximations of the Stokes equations. In *Efficient Solutions of Elliptic Systems. Notes on Numerical Fluid Mechanics*, vol. 10, Hackbusch W (ed.). Braunschweig Wiesbaden, 1984; 11–19.

Brezzi F, Douglas J and Marini LD. Two families of mixed finite elements for second order elliptic problems. *Numer. Math.* 1985; **47**:217–235.

Brezzi F, Douglas J and Marini LD. Recent results on mixed finite element methods for second order elliptic problems. In *Vistas in Applied Mathematics, Numerical Analysis, Atmospheric Sciences, Immunology*, Balakrishanan AV, Dorodnitsyn AA and Lions JL (eds). Optimization Software Publications: New York, 1986; 25–43.

Brezzi F, Douglas J, Fortin M and Marini LD. Efficient rectangular mixed finite elements in two and three space variables. *Math. Modell. Numer. Anal.* 1987; **21**:581–604.

Brezzi F, Douglas J, Duran R and Fortin M. Mixed finite elements for second order elliptic problems in three variables. *Numer. Math.* 1988; **51**:237–250.

Carey GF and Oden JT. *Finite Elements: A Second Course*, vol. II. Prentice Hall: Englewood Cliffs, NJ, 1983.

Ciarlet PG. *The Finite Element Method for Elliptic Problems*. North Holland: Amsterdam, 1978.

Clough RW. The finite element method in structural mechanics. In *Stress Analysis. Recent Developments in Numerical and Experimental Methods*, Zienkiewicz OC and Holister GS (eds). John Wiley & Sons: New York, 1965; 85–119.

Crisfield MA. *Finite Elements and Solution Procedures for Structural Analysis*. Pineridge Press: Swansea, UK, 1986.

Crisfield MA. *Non-Linear Finite Element Analysis of Solids and Structures, Vol. 1 – Essentials*. John Wiley & Sons: New York, 1991.

Crisfield MA. *Non-Linear Finite Element Analysis of Solids and Structures, Vol. 2 – Advanced Topics*. John Wiley & Sons: New York, 1997.

Crouziex M and Raviart PA. Conforming and non-conforming finite element methods for the stationary Stokes equations. *RAIRO Anal. Numer.* 1973; **7**:33–76.

Douglas J and Wang J. An absolutely stabilized finite element method for the Stokes problem. *Math. Comput.* 1989; **52**(186):495–508.

Falk RS. Nonconforming finite element methods for the equations of linear elasticity. *Math. Comput.* 1991; **57**:529–550.

Fortin M. Utilisation de la méthode des éléments finis en méchanique des fluides. *Calcolo* 1975; **12**:405–441.

Fortin M. An analysis of the convergence of mixed finite element methods. *RAIRO Anal. Numer.* 1977; **11**:341–354.

Fraeijs de Veubeke B. Displacement and equilibrium models in the finite element method. In *Stress Analysis. Recent Developments in Numerical and Experimental Methods*, Lectures Notes in Math. 606, Zienkiewicz OC and Holister GS (eds). John Wiley & Sons, 1965; 145–197.

Fraeijs de Veubeke B. Stress function approach. In *World Congress on the Finite Element Method in Structural Mechanics*, Bournemouth, 1975; 321–332.

Franca LP and Hughes TJR. Two classes of finite element methods. *Comput. Methods Appl. Mech. Eng.* 1988; **69**:89–129.

Franca LP and Stenberg R. Error analysis of some Galerkin least squares methods for the elasticity equations. *SIAM J. Numer. Anal.* 1991; **28**:1680–1697.

Glowinski R and Pironneau O. Numerical methods for the first biharmonic equation and for the two-dimensional Stokes problem. *SIAM Rev.* 1979; **21**:167–212.

Hansbo P and Larson MG. Discontinuous Galerkin and the Crouzeix–Raviart element: application to elasticity. *Math. Modell. Numer. Anal.* 2003; **37**:63–72.

Hood P and Taylor C. Numerical solution of the Navier–Stokes equations using the finite element technique. *Comput. Fluids* 1973; **1**:1–28.

Hughes TJR. *The Finite Element Method: Linear Static and Dynamic Finite Element Analysis*. Prentice Hall: Englewood Cliffs, NJ, 1987.

Hughes TJR and Franca LP. 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. Eng.* 1987; **65**:85–96.

Hughes TJR, Franca LP and Balestra M. A new finite element formulation for computational fluid dynamics. V: circumventing the Babuška–Brezzi condition: a stable Petrov–Galerkin formulation of the Stokes problem accommodating equal-order interpolations. *Comput. Methods Appl. Mech. Eng.* 1986; **59**:85–99.

Johnson C. *Numerical Solution of Partial Differential Equations by the Finite Element Method*. Cambridge University Press: Cambridge, UK, 1992.

Johnson C and Mercier B. Some equilibrium finite element methods for two-dimensional elasticity problems. *Numer. Math.* 1978; **30**:103–116.

Ladyzhenskaya OA. *The Mathematical Theory of Viscous Incompressible Flow*. Gordon & Breach: New York, 1969.

Lovadina C. Analysis of strain-pressure finite element methods for the Stokes problem. *Numer. Methods Partial Differ. Equations* 1997; **13**:717–730.

Lovadina C and Auricchio F. On the enhanced strain technique for elasticity problems. *Comput. Struct.* 2003; **18**:777–787.

Mansfield L. Finite element subspaces with optimal rates of convergence for the stationary Stokes problem. *RAIRO Anal. Numer.* 1982; **16**:49–66.

Marini LD. An inexpensive method for the evaluation of the solution of the lowest order Raviart-Thomas mixed method. *SIAM J. Numer. Anal.* 1985; **22**:493–496.

Nedelec JC. Mixed finite elements in $\mathbb{R}^3$. *Numer. Math.* 1980; **35**:315–341.

Ottosen NS and Petersson H. *Introduction to the Finite Element Method*. Prentice Hall: New York, 1992.

Pantuso D and Bathe KJ. A four-node quadrilateral mixed-interpolated element for solids and fluids. *Math. Models Methods Appl. Sci.* 1995; **5**:1113–1128.

Pierre R. Regularization procedures of mixed finite element approximations of the Stokes problem. *Numer. Methods Partial Differ. Equations* 1989; **5**:241–258.

Quarteroni A and Valli A. *Numerical Approximation of Partial Differential Equations*. Springer-Verlag: New York, 1994.

Raviart PA and Thomas JM. A mixed finite element method for second order elliptic problems. In *Mathematical Aspects of the Finite Element Method*, Lecture Notes in Mathematics 606, Galligani I and Magenes E (eds). Springer-Verlag: New York, 1977; 292–315.

Reddy JN. *An Introduction to the Finite Element Method*. McGraw-Hill: New York, 1993.

Reddy BD and Simo JC. Stability and convergence of a class of enhanced strain methods. *SIAM J. Numer. Anal.* 1995; **32**:1705–1728.

Scott LR and Vogelius M. Norm estimates for a maximal right inverse of the divergence operator in spaces of piecewise polynomials. *Math. Models Numer. Anal.* 1985; **19**:111–143.

Silvester DJ and Kechar N. Stabilised bilinear-constant velocity-pressure finite elements for the conjugate gradient solution of the Stokes problem. *Comput. Methods Appl. Mech. Eng.* 1990; **79**:71–86.

Simo JC. Topics on the numerical analysis and simulation of plasticity. In *Handbook of numerical analysis*, vol. III, Ciarlet PG and Lions JL (eds). Elsevier Science Publisher B.V., 1999; 193–499.

Simo JC. and Hughes TJR. *Computational Inelasticity*. Springer-Verlag: New York, 1998.

Simo JC and Rifai MS. A class of mixed assumed strain methods and the method of incompatible modes. *Int. J. Numer. Methods Eng.* 1990; **29**:1595–1638.

Stenberg R. Analysis of mixed finite element methods for the Stokes problem: a unified approach. *Math. Comput.* 1984; **42**:9–23.

Stenberg R. On some three-dimensional finite elements for incompressible media. *Comput. Methods Appl. Mech. Eng.* 1987; **63**:261–269.

Stenberg R. A family of mixed finite elements for the elasticity problem. *Numer. Math.* 1988; **53**:513–538.

Strang G and Fix GJ. *An Analysis of the Finite Element Method*. Prentice Hall: Englewood Cliffs, NJ, 1973.

Temam R. *Navier–Stokes Equations*. North Holland: Amsterdam, 1977.

Turner MJ, Clough RW, Martin HC and Topp LJ. Stiffness and deflection analysis of complex structures. *J. Aeronaut. Sci.* 1956; **23**:805–823.

Verfürth R. Error estimates for a mixed finite element approximation of the Stokes equation. *RAIRO Anal. Numer.* 1984; **18**:175–182.

Vogelius M. A right-inverse for the divergence operator in spaces of piecewise polynomials. *Numer. Math.* 1983; **41**:19–37.

Wait R and Mitchell AR. *Finite Element Analysis and Applications*. John Wiley & Sons: Chichester, West Sussex, 1985.

Zienkiewicz OC and Taylor RL. *The Finite Element Method* (5th edn), *Vol. 1 – The Basis*. Butterworth-Heinemann: Oxford, 2000a.

Zienkiewicz OC and Taylor RL. *The Finite Element Method*, (5th edn), *Vol. 2 – Solid Mechanics*. Butterworth-Heinemann: Oxford, 2000b.

Zienkiewicz OC and Taylor RL. *The Finite Element Method* (5th edn), *Vol. 3 – Fluid Dynamics*. Butterworth-Heinemann: Oxford, 2000c.

Zienkiewicz OC, Taylor RL and Baynham JAW. Mixed and irreducible formulations in finite element analysis. In *Hybrid and Mixed Finite Element Methods*, Atluri SN, Gallagher RH and Zienkiewicz OC (eds). John Wiley & Sons: New York, 1983; 405–431, Chap. 21.

Abstract: Within the well-known and highly effective *finite element method* for the computation of approximate solutions of complex boundary value problems, we focus on the often-called *mixed finite element methods*, where in our terminology the word ‘mixed’ indicates the fact that the problem discretization typically results in a linear algebraic system characterized by a null matrix on the main diagonal.

Accordingly, the goals of the present chapter are: (1) to sketch out that several physical problems share such an algebraic structure once a discretization is introduced; (2) to present a simple, algebraic version of the abstract theory that rules most applications of mixed finite element methods; (3) to give several examples of efficient mixed finite element methods; (4) finally, to give some hints on how to perform a stability and error analysis, focusing on a representative problem.