M2AN - Modélisation mathématique et analyse numérique

V. Ruas

Finite element methods for the three-field Stokes system in $\mathbb{R}^3$: Galerkin methods

M2AN - Modélisation mathématique et analyse numérique, tome 30, nº 4 (1996), p. 489-525

<http://www.numdam.org/item?id=M2AN_1996__30_4_489_0>

MATHEMATICAL MODELLING AND NUMERICAL ANALYSIS
MODÉLISATION MATHÉMATIQUE ET ANALYSE NUMÉRIQUE

(Vol 30, n° 4, 1996, p 489 à 525)

FINITE ELEMENT METHODS FOR THE THREE-FIELD STOKES SYSTEM IN $\mathbb{R}^3$: GALERKIN METHODS (*)

by V. RUAS [1]

Résumé — *Dans cet article, on introduit plusieurs méthodes d'éléments finis, pour la résolution du système de Stokes à trois champs, associé à l'écoulement de fluides viscoélastiques dans l'espace à trois dimensions Toutes les méthodes proposées sont fondées sur la formulation de Galerkin standard, et des résultats complets de convergence à l'ordre un ou deux sont donnés pour la plupart d'entre elles*

Abstract — *In this work several new finite element methods for solving the three-field Stokes system associated with viscoelastic flow problems in three-dimension space are introduced The methods presented are based on the standard Galerkin formulation, and complete proofs of first and second order convergence for the corresponding approximations are given for most of them.*

1. PRELIMINARIES

Before starting the study itself we first consider in this section some general aspects of the work, together with its motivation. Besides we give here an outline of the paper and a list of the notation used in the subsequent sections.

1.1. Introduction

In the framework of the convergence study of linear variational problems of non coercive type approximated via finite elements, one is essentially led to a stability analysis of the corresponding discrete problem. Two techniques are basically being employed in order to construct stable solution methods. The first one called here the *technique of stable methods* is based on the use of the same Galerkin formulation as for the continuous problem. This approach requires that the interpolation of the different unknown fields satisfy compatibility conditions, namely the so-called inf-sup conditions or yet in some cases the Ladyzhenskaya-Babuska-Brezzi conditions. In this case it is often neces-

(*) Manuscript received March 3, 1995 , revised May 3, 1995

[1] Faculté des Sciences, Université de Saint-Etienne & Laboratoire de Modélisation en Mécanique/UPMC, URA229-CNRS, Tour 66, 4^e étage, 75252 Paris cedex 05, France Depto de Análise, Universidade Federal Fluminense, Niterói, Rio de Janeiro state, Brazil, until March 1995

sary to choose representations of some of the discrete unknown fields involving locally defined polynomials such as the so-called bubble-functions of the elements These play mainly the role of ensuring stability and in general they add practically nothing as far as accuracy and order of convergence are concerned In the second approach called the *technique of stabilized methods* one attempts to satisfy the very same inf-sup conditions with classical piecewise polynomial representations However in this case the stability of the method is attained through the use of a modified variational form in the discrete case by adding some terms which may depend on the mesh step size In so doing the corresponding methods will be stable if some numerical parameters involved in such formulation satisfy appropriate conditions

This work deals with the approximation of the three-field Stokes system by the first type of technique For some new methods of the Galerkin-least-squares type, applied to the same problem, the author refers to another recent work of his [35] Let us recall that this system is the linearized form of several non linear problems In particular its study is essential for deriving efficient approximation methods of the systems of partial differential equations that govern viscoelastic flow, since in this case the three unknown fields, namely, the velocity, the pressure and the extra-stress tensor are helplessly coupled This study is illustrated in detail through the convergence analysis of a second order method for the three-dimensional case treated with tetrahedral meshes, based on the technique of parametrized degrees of freedom introduced by the author about ten years ago Additionally, some new first order methods are proposed and treated in abbridged form

As it should be stressed, in the present state-of-the-art both the study and the use of three-field finite element methods for solving this class of problems are incipient as far as the three-dimensional case is concerned This is particularly true of methods with a discontinuous pressure Since this approach is generally considered to be the most efficient way to satisfy the mass conservation in the flow, the present work brings about a contribution in this sense, as it deals with methods that fall into this category

Generally speaking, this work attempts to present a number of methods involving reasonable computational costs for three-dimensional systems of this class, in which the number of scalar unknowns is as high as ten

1.2. Motivation

Let us now briefly review the system of partial differential equations that mainly motivates the study carried out in this work, namely the one describing the flow of a viscoelastic liquid in a region Ω of the space considered to be a bounded open set with boundary $\partial\Omega$ These systems are derived on the basis of conservation laws of Continuum Mechanics, complemented with a constitutive law for the fluid assumed here to be of the differential type

We consider as a model problem only the case of isothermal flows, and of stable and incompressible fluids, which means that its physical and mechanical characteristics such as density, viscosity, among other parameters, do not change with the position of its particles. It is important to stress that such circumstances effectively occur in a wide spectrum of applications such as injection molding with melt polymers among other complex materials. The time-dependent, non isothermal and non stable cases may be treated as simple variants of this basic problem and an illustration of this assertion may be found in [1]

Now under the above assumptions denoting by $\vec{u}$ the *velocity field, p* the *hydrostatic pressure* and by τ the *Cauchy stress tensor* of the fluid given by

$$\tau = \sigma - pI \quad \tau = \{\tau_{ij}\}_{i,j=1}^{3}$$

where I is the identity tensor, and σ is the *extra stress tensor,* in a rather general way, the motion of the fluid under the action of body forces $\vec{f}$, is governed by the following system (*cf* [7, 11, 18, 37], where $\vec{x} = (x_1, x_2, x_3)$ represents the cartesian coordinates of the space $\mathbb{R}^3$

$$\left.\begin{array}{l} \rho \displaystyle\sum_{i=1}^{3} u_i \frac{\partial \vec{u}}{\partial x_i} - \overrightarrow{\operatorname{div}}\, \tau = \vec{f} \\ \tau^T \equiv \tau = \sigma - pI \end{array}\right\} \quad (\text{Momentum Equations}) \qquad (1)$$

$$\operatorname{div} \vec{u} = 0 \quad (\text{Mass Conservation Law}) \qquad (2)$$

$$\sigma + \mathcal{D}_0\, \sigma = 2\,\eta[\varepsilon(\vec{u}) + \mathcal{D}_1\, \varepsilon(\vec{u})]$$

(Constitutive Law of the Differential Type)(3)

where

- $\varepsilon(\vec{u}) = [\overrightarrow{\text{grad}}\, \vec{u} + (\overrightarrow{\text{grad}}\, \vec{u})^T]/2$ is the *strain rate tensor*,
- ρ is the density of the fluid ;
- η is the viscosity of the fluid

$\mathcal{D}_i$, $i = 0, 1$, denotes an objective material derivative (nonlinear) operator (*cf.* [4]), which means that its expression is invariant with respect to the frames which the tensor is referred to

In most of the applications the fluid flow is sufficiently slow or equivalently the Reynolds number is low (*cf* [13]), so that the (inertia) term involving ρ

may be neglected in equation (1) In so doing, the system of equations that govern the motion of the viscoelastic liquid in terms of the fields $(\sigma, \vec{u}, p)$ (admitting that law (3) implies that σ is symmetric) reduces to

$$\begin{cases} -\ \overrightarrow{\operatorname{div}}\,\sigma + \overrightarrow{\operatorname{grad}}\,p = \vec{f} \\ \sigma + \mathscr{D}_0\,\sigma = 2\,\eta[\varepsilon(\vec{u}) + \mathscr{D}_1\,\varepsilon(\vec{u})] \\ \operatorname{div}\,\vec{u} = 0 \end{cases} \tag{4}$$

As far as boundary conditions for (4) are concerned, for the sake of simplicity and omitting eventual conditions on σ (see e g [24]), we will consider the case where the velocity field is entirely prescribed on the boundary, that is

$$\vec{u} = \vec{g} \quad \text{on} \quad \partial\Omega \tag{5}$$

where $\vec{g}$ is a field satisfying the global conservation property

$$\int_{\partial\Omega} \vec{g}\,.\,\vec{n}\;dS = 0\,,$$

$\vec{n}$ being the unit outer normal vector with respect to $\partial\Omega$

In any case the three-field Stokes system results from the linearization of (4) in the limiting case where the terms involving the operators $\mathscr{D}_0$ and $\mathscr{D}_1$ may be neglected Our approach then is the study of approximation methods of the three-field linear system

$$\begin{cases} \overrightarrow{\operatorname{div}}\,\sigma - \overrightarrow{\operatorname{grad}}\,p = -\,\vec{f} & \text{in} \quad \Omega \\ \sigma = 2\,\eta\varepsilon(\vec{u}) & \text{in} \quad \Omega \\ \operatorname{div}\,\vec{u} = 0 & \text{in} \quad \Omega \\ \vec{u} = \vec{g} & \text{on} \quad \partial\Omega \end{cases} \tag{6}$$

aiming at applying them to the case of the non linear system (4)-(5) Although we can only conjecture here that any convergent solution method of system (6) is also convergent when applied to system (4)-(5), it is possible to assert on the basis of Baranger & Sandri' pioneer work (see [6] and references therein), that at least for Oldroyd models (i e $\mathscr{D}_1 = \alpha\mathscr{D}_0$, $\alpha \in \mathbb{R}^+$, see e g [11]), under the conditions allowing the convergence of certain type of finite element approximations of system (6), the very same types of approximations converge as well. This is particularly true of second order triangular elements that become actually accurate to the order 3/2 in the nonlinear case [6]. Since we will study more particularly second order tetrahedral methods for system (6) one may legitimately conjecture that their convergence properties are maintained in the case of Oldroyd models, although some loss in order of convergence is to be expected.

1.3. Outline of the paper

As we are going to study finite element methods to solve the three-field Stokes system (6), we first recall in Section 2 some general results about the approximation of linear variational problems. Next in the same section we introduce the variational form under which we will consider system (6) in this work. More specifically we will deal with the standard Galerkin formulation for which we shall search for *stable finite element methods.* Additionally in Section 2 we exhibit the conditions to be satisfied by a finite element method to yield convergent approximations with an appropriate order.

In Section 3 we study in detail a second order stable approximation based on a finite element method for solving the three-dimensional two-field (velocity-pressure) Stokes system introduced by the author in [29]. The method is optimal in terms of the discrete representations of the three fields, at least as far as local stability analysis are concerned.

Following this detailed study we briefly present in Section 4 some first order three-field finite element methods for the Galerkin formulation too.

1.4. Notation

Before starting our study let us specify the notation used in the text that cannot be considered as universal. At the same time we recall some classical definitions related to Sobolev spaces (see e.g. [2]).

Let S be a measurable bounded set of $\mathbb{R}^n$, $n = 1, 2, 3$, $S \subseteq \Omega$, and $\vec{x} = (x_1, x_2, \ldots, x_n)$ be the space variables related to a cartesian coordinate system.

- $f_{|S}$ denotes the restriction to S of a function f defined in Ω or on a subset of Ω that contains S.
- $(f|g)_S$ denotes the standard inner product of $L^2(S)$ given by

$$(f|g)_S = \int_S fg \, dS \quad \forall f, g \in L^2(S)\,,$$

and $\| \cdot \|_{0,S}$ denotes the associated norm, i.e., $\|f\|_{0,S} = (f|f)_S^{1/2}$.

- $|S|$ represents the measure of S, that is,

$$|S| = \int_S dS .$$

- For $m \in \mathbb{N}$, $H^m(S)$ denotes the usual hilbertian Sobolev space equipped with the standard inner product denoted by $(\, . \, | . \,)_{m,S}$ and associated norm denoted by $\| \, . \, \|_{m,S}$.
- For $H^m(S)$ the seminorm involving only the derivatives of order m is denoted by $|v|_{m,S}$.
- S being a sufficiently smooth domain of $\mathbb{R}^n$, with boundary ∂S of piecewise C^1 class, $H_0^1(S)$ is the closed subspace of $H^1(S)$ consisting of those functions whose trace over ∂S vanishes a.e., normed by $| \, . \, |_{1,S}$ (*cf.* [8]).
- Whenever S is Ω itself we shall omit dS in the above integrals, and symbol S in the above defined norms, seminorms and inner products.
- $L_0^2(S)$ is the closed subspace of $L^2(S)$ of those functions f such that

$$\int_S f \, dS = 0 .$$

- V being a function space, $\vec{V}$ denotes the space of fields $\vec{v} = (v_1, v_2, v_3)$ such that $v_i \in V$, $i = 1, 2, 3$, and $\mathbf{V}$ and $\mathbf{V}_S$ denote respectively the space of arbitrary and symmetric tensors $\{\tau_{ij}\}_{i,j=1}^3$ such that $\tau_{ij} \in V$, $\forall i, j \in \{1, 2, 3\}$.
- $\vec{u}$ and $\vec{v}$ being two $\mathbb{R}^3$ valued vector fields, $\vec{u} \, . \, \vec{v}$ denotes their euclidean inner product, that is

$$\vec{u} \, . \, \vec{v} = \sum_{i=1}^{3} u_i \, v_i ,$$

- $|\vec{u}| = (\vec{u} \, . \, \vec{u})^{1/2}$.
- τ and σ being two 3×3 tensors their inner product and associated norm are defined by :

$$\sigma : \tau = \sum_{i=1}^{3} \sum_{j=1}^{3} \sigma_{ij} \, \tau_{ij}$$

- $|\sigma| = (\sigma : \sigma)^{1/2}$.

- The notations $(\,.\,|\,.\,)_S$, $(\,.\,|\,.\,)_{m,S}$ and $\|\,.\,\|_{m,S}$, $\|\,.\,\|_{s,S}$ and $((\,.\,|\,.\,))_{1,S}$ will naturally extended to the spaces $\vec{L}^2(S)$, $\vec{H}^m(S)$, $\vec{H}^s(S)$, $\vec{H}^1_0(S)$ and to $\mathbf{L}^2(S)$, $\mathbf{L}^2_S(S)$, $\mathbf{H}^m(S)$, etc. as well, which means that in the definition of the inner product of these spaces the product appearing in the integrals are to be replaced by vector or tensor inner products, respectively.
- E being a normed vector space with norm $\|\,.\,\|_E$, S_E denotes the unit sphere of E, namely

$$S_E = \{e | e \in E \quad \text{and} \quad \|e\|_E = 1\}\,.$$

- For $\vec{x} \in \Omega$ and $\varepsilon \in \mathbb{R}$, $\varepsilon > 0$, $B(\vec{x}, \varepsilon) = \{\vec{y} \in \mathbb{R}^n / |\vec{x} - \vec{y}| < \varepsilon\}$.

2. VARIATIONAL FORMS

In this Section we will first present the basic and general functional background which the convergence analysis of the methods to be studied in the next two sections relies upon. Next we consider the particular case of the Galerkin formulation used in this work. Without any loss of generality henceforth we take $\eta = 1/2$.

2.1. Functional Background

As we will see later on, system (6) will be written in a variational form of the following type.

Let

(i) Z be a Hilbert space with inner product $(\,.\,|\,.\,)_Z$ and associated norm $\|\,.\,\|_Z$;

(ii) $a : Z \times Z \to \mathbb{R}$ be a continuous bilinear form, which means that $\exists M > 0$ such that

$$a(y,z) \leqslant M \|y\|_Z\, \|z\|_Z \quad \forall y, z \in Z\,;$$

(iii) $L : Z \to \mathbb{R}$ be a continuous linear form.

By definition,

$$\|a\| = \sup_{y,z \in S_Z} a(y,z)\,.$$

The variational problem to be considered is :

$$(\mathscr{P})\begin{cases}\text{Find } y \in Z \text{ such that}\\ a(y,z)=L(z) \quad \forall z \in Z .\end{cases}$$

For problem $(\mathscr{P})$ we have the following well-known result due to Babuška [5] and extended and refined by Dupire [12].

THEOREM 2.1 ([5, 12]) : *Under the assumptions (i)-(ii)-(iii) there exists a unique solution y to problem* $(\mathscr{P})$ *if and only if*

(iv) $\exists \alpha > 0$ *such that* $\forall y \in Z \quad \sup_{z \in S_Z} a(y,z) \geq \alpha \|y\|_Z$

(v) $\forall z \in S_Z, \exists y \in Z,$ *such that* $a(y,z) > 0$. ■

Notice that if a is symmetric, condition (v) is a simple consequence of condition (iv).

Suppose that one wishes to determine approximations y_h of the solution y in a family $\{Z_h\}_h$ of finite dimensional spaces that have suitable approximation properties vis-à-vis Z. The subscript h of the family of spaces is supposed to sweep a non finite set with the same cardinality as $\mathbb{N}$. Assume also that h is strictly positive and that it varies decreasingly tending to zero.

Although *a priori* the converse situation would be desirable, in the cases to be considered in the next section for each h, Z_h will not be a subspace of Z. Otherwise stated we will be dealing with *non conforming* approximations of y. In this way it is not possible to guarantee in general neither that a is defined over $Z_h \times Z_h$ nor that L is defined over Z_h. Moreover the norm $\|\cdot\|_Z$ will not necessarily be defined over Z_h. All this leads to the following additional definitions :

$(\mathrm{i})_h$ For each h, $\|\cdot\|_h : Z_h + Z \to \mathbb{R}$ is a norm that satisfies

$$\|z\|_h = \|z\|_Z \quad \forall z \in Z .$$

In so doing we further introduce :

$(\mathrm{ii})_h$ A bilinear form $a_h : (Z_h + Z) \times (Z_h + Z) \to \mathbb{R}$ uniformly continuous in the sense that $\exists M'$ independent of h such that

$$a(y,z) \leq M' \|y\|_h \, \|z\|_h \quad \forall y, z \in Z_h + Z$$

and

$(\mathrm{iii})_h$ A linear form $L_h : Z_h \to \mathbb{R}$ necessarily continuous.

Analogously we define :

$$\|a_h\| = \sup_{y,z \in S_{Z_h+Z}} a_h(y,z) .$$

Now the family of approximate problems that we wish to solve is $(\mathscr{P}_h)_h$ where

$$(\mathscr{P}_h)\begin{cases}\text{Find } y_h \in Z_h \text{ such that}\\ a_h(y_h, z) = L_h(z) \quad \forall z \in Z_h .\end{cases}$$

The main issue to be addressed is how to estimate the error $y - y_h$ measured in the norm $\| \cdot \|_h$. However in order to do so it is necessary to study beforehand the existence and uniqueness of the solution of $(\mathscr{P}_h)$. The answer to both questions may be obtained by applying the following result slightly adapted from Dupire's [12] (see Remark 2.1).

THEOREM 2.2 ([12]) : *Under assumptions* $(i)_h$, $(ii)_h$ *and* $(iii)_h$ Z_h *being a finite dimensional space* $\forall h$, $(\mathscr{P}_h)$ *has a unique solution* y_h *if and only if*

$(\text{iv})_h$ $\exists \alpha_h > 0$ *such that* $\forall y \in Z_h$ $\sup\limits_{z \in S_{Z_h}} a_h(y, z) \geqslant \alpha_h \| y \|_h$

Furthermore the following estimate holds

$$\| y - y_h \|_h \leqslant \frac{1}{\alpha_h}\left[\| a_h \| \inf_{z \in Z_h} \| y - z \|_h + \sup_{z \in S_{Z_h}} | a_h(y, z) - L_h(z) | \right] . \quad (7)$$

■

Remark 2.1 : As Z_h is a finite dimensional space we may disregard a condition analogous to (v) for problem $(\mathscr{P}_h)$. Indeed in this context $(\text{iv})_h$ ensures that any matrix associated with form a_h and space Z_h is invertible. This clearly suffices to establish both existence and uniqueness of a solution to $(\mathscr{P}_h)$. ■

2.2. The case of the three-field Stokes system

Let us now go back to the main purpose of our study, that is, the approximation of the three-field Stokes system (6).

First let us set it under form $(\mathscr{P})$ and for this purpose we assume that $\vec{f} \in \vec{L}^2(\Omega)$. On the other hand, in order to simplify the notation we shall only consider the case where $\vec{g} = \vec{0}$. The case where $\vec{g}$ is arbitrary may be treated in an entirely analogous way, after performing non essential modifications in the analysis that follows.

The unknown z of our problem is the triple $(\sigma, \vec{u}, p)$ which will be searched for in space

$$Z = \mathbf{L}^2_S(\Omega) \times \vec{H}^1_0(\Omega) \times L^2_0(\Omega) .$$

This space equipped with the natural norm

$$\|(\tau, \vec{v}, q)\|_Z = \left[\|\tau\|_0^2 + |\vec{v}|_1^2 + \|q\|_0^2 \right]^{1/2}$$

is a Hilbert space. It is also so for any other equivalent norm such as the one to be considered in Section 3.

In so doing the problem to solve is :

$$(\mathscr{P}) \begin{cases} \text{Find } (\sigma, \vec{u}, p) \in Z \text{ such that} \\ a((\sigma, \vec{u}, p), (\tau, \vec{v}, q)) = L((\tau, \vec{v}, q)) \quad \forall (\tau, \vec{v}, q) \in Z \end{cases}$$

where

$$a((\sigma, \vec{u}, p), (\tau, \vec{v}, q)) = (\sigma|\tau) + (p|\operatorname{div} \vec{v}) - (\tau|\varepsilon(\vec{u})) + (q|\operatorname{div} \vec{u}) - (\sigma|\varepsilon(\vec{v})) \quad (8)$$

and

$$L((\tau, \vec{v}, q)) = -(\vec{f}|\vec{v}). \quad (9)$$

One can easily prove that every solution of $(\mathscr{P})$ is a solution of (6) with $\vec{g} = \vec{0}$, in the sense of distributions and conversely, under certain regularity assumptions on Ω, every solution of (6) with $\vec{g} = \vec{0}$ is a solution of $(\mathscr{P})$.

On the other hand, the fact that $(\mathscr{P})$ has a unique solution is a consequence of well-known results in connection with Theorem 2.1. Referring to the author's recent work [30] for further details let us just say here that, since form a given by (8) is symmetric, condition (iv) related to $(\mathscr{P})$ is equivalent to the following ones :

(vi) $\exists \beta > 0$ such that $\forall q \in L_0^2(\Omega)$ $\sup_{\vec{v} \in S_{\vec{H}_0^1(\Omega)}} \int_\Omega q \operatorname{div} \vec{v} \geq \beta \|q\|_0$

(vii) $\exists \beta' > 0$ such that $\forall v \in \vec{U}$ $\sup_{\tau \in S_{L_s^2(\Omega)}} \int_\Omega \tau : \varepsilon(\vec{v}) \geq \beta' \|\vec{v}\|_1$

where

$$\vec{U} = \{ \vec{v} \mid \vec{v} \in \vec{H}_0^1(\Omega) \quad \text{and} \quad \operatorname{div} \vec{v} = 0 \text{ a.e. in } \Omega \}.$$

The first condition is nothing but the classical LBB condition (see e.g. [21], [5], [9]) for the Lagrange multiplier associated with the restriction $\operatorname{div} \vec{u} = 0$. It is satisfied according to [19]. The second one was identified by the author in [28] as a necessary condition in a more restrictive form exploited in [16], and was given as such in [30] and [36]. It is actually a consequence of :

$$\sup_{\tau \in S_{L^2_s(\Omega)}} \int_\Omega \tau : \varepsilon(\vec{v}) \geqslant \frac{1}{\|\varepsilon(\vec{v})\|_0} \int_\Omega \varepsilon(\vec{v}) : \varepsilon(\vec{v}) =$$

$$= \left\{ \sum_{i,j=1}^{3} \frac{1}{2} \left[\int_\Omega \left(\frac{\partial v_i}{\partial x_j} \right)^2 + \int_\Omega \frac{\partial v_i}{\partial x_j} \frac{\partial v_j}{\partial x_i} \right] \right\}^{1/2} .$$

By classical density arguments, and by using integration by parts, taking into account that $v_i = 0$ a.e. on $\partial\Omega$ for every i, we derive

$$\sup_{\tau \in S_{L^2_s(\Omega)}} \int_\Omega \tau : \varepsilon(\vec{v}) \geqslant \beta' |\vec{v}|_1 \quad \text{with} \quad \beta' = \frac{\sqrt{2}}{2} .$$

Remark 2.2 *:* The above calculations establishing that

$$|\vec{v}|_1 = \sqrt{2} \|\varepsilon(\vec{v})\|_0 \quad \forall \vec{v} \in \vec{H}^1_0(\Omega) \cap \operatorname{Ker}(\operatorname{div})$$

are well-known. This relation is actually a particular case of Korn's second inequality (*cf.* [14]), stating that $\exists K_2 > 0$ such that $\forall \vec{v} \in \vec{H}^1(\Omega)$ that vanishes a.e. on a portion of $\partial\Omega$ having non zero measure, then

$$\|\varepsilon(\vec{v})\|_0 \leqslant |v|_1 \leqslant K_2 \|\varepsilon(\vec{v})\|_0 .$$

However for fields belonging to a finite element subspace not included in $\vec{H}^1(\Omega)$ these relations do not necessarily hold. This will be precisely the case of a velocity space studied hereafter and in principle it will be necessary to prove equivalence of both norms in the corresponding discrete version. Such results are called discrete Korn's second inequalıty and although this is not strictly necessary a proof of it is given in [35] for this space. ■

Let us now switch to the discrete version of $(\mathscr{P})$ to be considered in this work. For this purpose let us consider that Ω is a domain having a polyhedral boundary.

Let $\{\mathcal{T}_h\}_h$ be a family of partitions of Ω into tetrahedrons respecting the usual conditions required for applying the finite element method (*cf.* [10]). In particular, if T denotes a tetrahedron of $\mathcal{T}_h$ (considered here to be an open set) and defining for every bounded open set S *of* $\mathbb{R}^3$,

$$h_S = \sup_{\vec{x}, \vec{y} \in S} | \vec{x} - \vec{y} | \quad \text{and} \quad \rho_S = \sup_{B(\vec{x}, \varepsilon) \subset S, \vec{x} \in S} \{\varepsilon\}$$

we set as usual

$$h = \max_{T \in \mathcal{T}_h} h_T \quad \text{and} \quad \rho = \min_{T \in \mathcal{T}_h} \rho_T .$$

Next we assume that family $\{\mathcal{T}_h\}_h$ is quasiuniform (*cf.* [10]), i.e. :

$$\exists c > 0 \text{ independent of } h \text{ such that } \rho > ch \quad \forall h .$$

Let us associate with every partition $\mathcal{T}_h$ three finite dimensional spaces $\mathbf{T}_h$, V_h and Q_h in such a way that $\mathbf{T}_h$, $\vec{V}_h$ and Q_h are the respective discrete analogues of $\mathbf{L}^2_s(\Omega)$, $\vec{H}^1_0(\Omega)$ and $L^2_0(\Omega)$, in which we will search for approximations σ_h, $\vec{u}_h$ and p_h of σ, $\vec{u}$ and p. In the cases considered in this work we have $\mathbf{T}_h \subset \mathbf{L}^2_s(\Omega)$ and $Q_h \subset L^2_0(\Omega)\ \forall h$, but not necessarily $\vec{V}_h \subset \vec{H}^1_0(\Omega)$. In this way among other possibilities the norm $\| \cdot \|_h$ that we have selected here for $Z_h = \mathbf{T}_h \times \vec{V}_h \times Q_h$ is the one given by :

$$\| (\tau, \vec{v}, q) \|_h = [\, \|\tau\|_0^2 + \|\varepsilon(\vec{v})\|_{0,h}^2 + \|q\|_0^2]^{1/2} \tag{10}$$

where

$$\|R\|_{0,h} = (R|R)_h^{1/2} \quad \text{with} \quad (R|S)_h = \sum_{T \in \mathcal{T}_h} (R|S)_T , \tag{11}$$

whereby R and S are a pair of functions, vector fields or tensors defined in each element of $\mathcal{T}_h$, whose components belong to $L^2(T)\ \forall T \in \mathcal{T}_h$. Notice that this will be the case of $\vec{v}_{|T}$ or of $(\overrightarrow{\text{grad}}\, \vec{v})_{|T}$ and $\varepsilon(\vec{v})_{|T}\ \forall \vec{v} \in \vec{V}_h$.

Remark 2.3 : As we assume that $\forall \vec{v} \in \vec{V}_h$, $\vec{v} \in \vec{C}^\infty(\bar{T})\ \forall T \in \mathcal{T}_h$, we will abusively denote by $\overrightarrow{\text{grad}}\, \vec{v}$ the tensor of $L^2(\Omega)$ defined by

$$(\overrightarrow{\text{grad}}\, \vec{v})_{|T} = \overrightarrow{\text{grad}}\, (\vec{v}_{|T}) \quad \forall T \in \mathcal{T}_h$$

while $\varepsilon(\vec{v})$ will represent

$$\frac{1}{2}[\overrightarrow{\text{grad}}\,\vec{v} + (\overrightarrow{\text{grad}}\,\vec{v})^T],$$

with $\text{div}\,\vec{v} = \text{Tr}\,[\varepsilon(\vec{v})]$. ■

Notice that our choice implies that we will not be working with the norm of $\vec{V}_h$ that appears to be the most natural in the present framework, that is, norm $|\cdot|_{1,h}$ given by :

$$|R|_{1,h} = ((R|R))_{1,h}^{1/2} \quad \text{and} \quad ((R|S))_{1,h} = (\overrightarrow{\text{grad}}\,R|\overrightarrow{\text{grad}}\,S)_h$$

where R and S play the same role as in (11).

Now let us introduce the variational form $(\mathcal{P}_h)$ to be considered in Section 3, namely, the standard Galerkin formulation, where

$$\left.\begin{aligned} a_h((\sigma,\vec{u},p),(\tau,\vec{v},q)) &= (\sigma|\tau) - (\sigma|\varepsilon(\vec{v}))_h - (\tau|\varepsilon(\vec{u}))_h \\ &\quad + (p|\text{div}\,\vec{v})_h + (q|\text{div}\,\vec{u})_h \end{aligned}\right\} \tag{12}$$

and

$$L_h = L\,. \tag{13}$$

With the above definitions problem $(\mathcal{P}_h)$ will take the form

$$(\mathcal{P}_h^1)\begin{cases} \text{Find } (\sigma_h, \vec{u}_h, p_h) \in Z_h \text{ such that} \\ a_h((\sigma_h,\vec{u}_h,p_h),(\tau,\vec{v},q)) = L_h((\tau,\vec{v},q)) \quad \forall(\tau,\vec{v},q) \in Z_h\,. \end{cases}$$

The analysis related to problem $(\mathcal{P}_h)$ will be carried out in the light of an adaption of the analysis given in [34] for the conforming case. Although the essential modifications are aimed at using discrete norms or inner products, we will recall below the main arguments of this analysis in order to clarify the steps to follow.

First of all we observe that problem $(\mathcal{P}_h^1)$ may be set in the following « mixed form ».

Let Ψ and Ξ be two Hilbert spaces with norms $\| \cdot \|_{\Psi}$ and $\| \cdot \|_{\Xi}$ respectively. We wish to solve :

$$(\mathscr{M}) \begin{cases} \text{Find } (\chi, \pi) \in \Psi \times \Xi \text{ such that} \\ c(\chi, \psi) + b(\psi, \pi) = F(\psi) \quad \forall \psi \in \Psi \\ b(\chi, \xi) = G(\xi) \quad \forall \xi \in \Xi \end{cases}$$

where $F \in \Psi'$, $G \in \Xi'$ and $c: \Psi \times \Psi \to \mathbb{R}$ with

$$c(\chi, \psi) = c(\psi, \chi) \quad \forall \chi, \psi \in \Psi$$

and $b: \Psi \times \Xi \to \mathbb{R}$ are continuous bilinear forms over the respective pair of spaces which they are applied to.

According to Theorem 2.1 problem $(\mathscr{M})$ has a unique solution if and only if $\exists \alpha > 0$ such that $\forall (\chi, \pi) \in \Psi \times \Xi$.

$$\sup_{(\psi, \xi) \in S_{\Psi \times \Xi}} [c(\chi, \psi) + b(\psi, \pi) + b(\chi, \xi)] \geqslant \alpha \| (\chi, \pi) \| \tag{14}$$

where the above norm is the natural one for the product space $\Psi \times \Xi$.

On the other hand, according to well-known results, the above condition is satisfied if and only if (*cf.* [9])

- $\exists \beta > 0$ such that $\forall \xi \in \Xi$, $\sup\limits_{\psi \in S_{\Psi}} b(\psi, \xi) \geqslant \beta \| \xi \|_{\Xi}$

and

- $\exists \gamma > 0$ such that $\forall \chi \in X$, $\sup\limits_{\psi \in S_X} c(\psi, \chi) \geqslant \gamma \| \chi \|_{\Psi}$ where

$$X = \{ \psi | \psi \in \Psi \text{ and } b(\psi, \xi) = 0 \quad \forall \xi \in \Xi \}.$$

Besides this, according to [12] constant α in (14) may be chosen to be

$$\alpha = \frac{\gamma \beta^2}{\| c \|^2 + \beta^2} \quad \text{where} \quad \| c \| = \sup_{\chi, \psi \in S_{\Psi}} c(\chi, \psi). \tag{15}$$

In the case of problem $(\mathscr{P}_h^1)$ we may take

- $\Psi = \mathbf{T}_h \times \vec{V}_h$ and $\Xi = Q_h$
- $c((\sigma, \vec{u}), (\tau, \vec{v})) = (\sigma | \tau) - (\sigma | \varepsilon(\vec{v}))_h - (\tau | \varepsilon(\vec{u}))_h$
- $b((\tau, \vec{v}), q) = (\operatorname{div} \vec{v} | q)_h$
- $F((\tau, \vec{v})) = -(\vec{f} | \vec{v})$ and $G \equiv 0$,

and we trivially have

$$\| c \| \leqslant 2. \tag{16}$$

In this way, since X is space $\mathbf{T}_h \times \vec{U}_h$ where

$$\vec{U}_h = \{ \vec{v} \mid \vec{v} \in \vec{V}_h \quad \text{and} \quad (\operatorname{div} \vec{v} \mid q)_h = 0 \quad \forall q \in Q_h \}$$

$(\mathscr{P}_h^1)$ will have a unique solution if and only if

$(\mathrm{vi})_h$ $\exists \beta_h > 0$ such that $\forall q \in Q_h \sup\limits_{\vec{v} \in S_{\vec{V}_h}} (\operatorname{div} \vec{v} \mid q)_h \geqslant \beta_h \| q \|_0$

$(\mathrm{vii})_h$ $\exists \gamma_h > 0$ such that $\forall (\sigma, \vec{u}) \in \mathbf{T}_h \times \vec{U}_h$,

$$\sup_{(\tau, \vec{v}) \in S_{\mathbf{T}_h \times O_h}} c((\sigma, \vec{u}), (\tau, \vec{v})) \geqslant \gamma_h \| (\tau, \vec{v}) \| =$$

$$= \gamma_h [\| \tau \|_0^2 + \| \varepsilon(\vec{v}) \|_{0,h}^2]^{1/2} . \quad (17)$$

Notice that condition (17) is nothing but the necessary and sufficient condition for existence and uniqueness of a solution to a problem $(\mathscr{P})$, which is again of the form $(\mathscr{M})$, with :

- $\Psi = \mathbf{T}_h$ and $\Xi = \vec{U}_h$
- $c(\sigma, \tau) = (\sigma \mid \tau)$
- $b(\tau, \vec{v}) = (\tau \mid \varepsilon(\vec{v}))_h$
- $F \equiv 0$ and $G(\vec{v}) = -(\vec{f} \mid \vec{v})$.

Hence by virtue of the same arguments as above we conclude that (17) is satisfied if and only if

$(\mathrm{viii})_h$ $\exists \beta_h' > 0$ such that

$$\forall \vec{v} \in \vec{U}_h, \sup_{\tau \in S_{\mathbf{T}_h}} (\tau \mid \varepsilon(\vec{v}))_h \geqslant \beta_h' \| \varepsilon(\vec{v}) \|_{0,h}$$

since for every subspace $\mathbf{S}_h$ of $\mathbf{T}_h$, $\forall \tau \in \mathbf{S}_h$,

$$\sup_{\sigma \in S_{\mathbf{S}_h}} (\sigma \mid \tau) = \| \tau \|_0 \quad \text{i.e. } \gamma = \| c \| = 1 .$$

Finally repeating the argument applied in [12] to the mixed problem under consideration, assumption $(\mathrm{iv})_h$ of Theorem 2.2 applies as well to the new mixed problem with $\alpha_h = \gamma_h$, where

$$\gamma_h = \frac{\beta_h'^2}{1 + \beta_h'^2} . \quad (18)$$

Summarizing we have :

THEOREM 2.3 ([34]) : *Problem* ($\mathcal{P}_h^1$) *satisfies condition* (iv)$_h$ *if and only if* (vi)$_h$ *and* (vii)$_h$ *are satisfied, and in this case according to* (15), (16), (18) :

$$\alpha_h = \frac{(\beta_h \beta'_h)^2}{(4 + \beta_h^2)(1 + \beta_h'^2)} . \tag{19}$$

■

Now recalling Theorem 2.2, since a_h is clearly uniformly continuous (condition (ii)$_h$), under the assumption that $\vec{u} \in H^{k+1}(\Omega)$ and $p \in H^k(\Omega)$ with k integer, $k \geqslant 1$, we have :

(2a) If $\| \cdot \|_h$ is a norm over $\mathbf{T}_h \times \vec{V}_h \times Q_h$ (condition $((i)_h)$;

(2b) If β_h and β'_h in conditions (vi)$_h$ and (viii)$_h$ are independent of h ;

(2c) If the non conformity term can be bounded by $C_N(\sigma, \vec{u}, p)\, h^k$ where C_N is independent of h :

(2d) If both $\mathbf{T}_h$ and $\bar{Q}_h = Q_h \oplus \{1\}$ contain the space consisting of tensors or functions whose components restricted to each element of $\mathcal{T}_h$ is a polynomial of degree less than or equal to $k - 1$, and if $\vec{V}_h$ contains the space of fields whose components (with suitable vanishing properties of the boundary of Ω) restricted to each element of $\mathcal{T}_h$ is a polynomial of degree less than or equal to k,

then there will exist a constant $C(\vec{u}, p)$ independent of h such that

$$\| (\sigma, \vec{u}, p) - (\sigma_h, \vec{u}_h, p_h) \|_h \leqslant C(\vec{u}, p)\, h^k .$$

Since condition (2d) is classical in finite element theory, in the next section all the analysis will be devoted to proving condition (2b) besides the bound given by (2c) for the non conformity term, after having treated appropriately issue (2a).

In the remainder of this work a capital C in different forms will represent constants independent of h.

3. A STABLE SECOND ORDER METHOD

In this Section we equip $\vec{H}_0^1(\Omega)$ with the norm $\| \varepsilon(\cdot) \|_0$. According to Remark 2.2, $\vec{H}_0^1(\Omega)$ is also a Hilbert space with the corresponding inner product. Let us then study the particular form of approximate problem ($\mathcal{P}_h^1$), defined as follows :

Let $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ be the barycentric coordinates of a generic tetrahedron T, related to the respective vertices S_i, $i = 1, 2, 3, 4$ and $\varphi_T = \lambda_1 \lambda_2 \lambda_3 \lambda_4$ be the bubble function of T. We define :

- $T_h = T_h^0 \oplus T_h^C$ where

$$T_h^0 = \{v \mid v \in C^0(\overline{\Omega}) \quad \text{and} \quad v|_T \in P_1(T) \quad \forall T \in \mathcal{T}_h\}$$

  $P_k(X)$ being the space of polynomials of degree less than or equal to k defined in an open set X of $\mathbb{R}^n$, and

$$T_h^C = \{v \mid v_{|T} \in \{\lambda_{i\varphi_T}\}_{i=1}^4 \quad \forall T \in \mathcal{T}_h\}.$$

- $V_h = V_h^0 \oplus V_h^B$, where

$$V_h^B = \{v \mid v_{|T} \in \{\varphi_T\} \quad \forall T \in \mathcal{T}_h\}$$

  and V_h^0 (*cf.* [29]) is the space of those functions v whose restriction to each $T \in \mathcal{T}_h$ is a polynomial of degree less than or equal to two, and that satisfy the following continuity and nullity properties :

  1) v is continuous at the centroid of every face common to two distinct tetrahedrons of $\mathcal{T}_h$;
  2) $v = 0$ at the centroid of every face of an element of $\mathcal{T}_h$ contained in $\partial\Omega$;
  3) Over every edge l of $\mathcal{T}_h$ not contained in $\partial\Omega$ the values

$$f_l(v) = \frac{\nu}{|l|} \int_l v \, dl + (1 - \nu)\, v(M),$$

     related to all the tetrahedrons containing l coincide, M being the mid-point of l and $\nu = 9/5$;
  4) $f_l(v) = 0$ if $l \subset \partial\Omega$.

- Q_h is the space $\tilde{Q}_h \cap L_0^2(\Omega)$ where

$$\tilde{Q}_h = \{v \mid v_{|T} \in P_1(T) \quad \forall T \in \mathcal{T}_h\}.$$

Notice that a function of T_h is uniquely defined in each tetrahedron by its values at the four vertices and in four arbitrarily chosen distinct inner points.

On the other hand the choice of degree of freedom that appears to be natural in order to uniquely define a function $v \in V_h$ in a tetrahedron T, consists of the values $f_l(v)$ [2] given in condition 3) of the definition of V_h^0 related to the six edges of T, the values of v at the four centroids of faces of T, and at the barycenter of T itself (see also [29]).

[2] Called parametrized degrees of freedom.

Finally the set of degrees of freedom used to uniquely define a function of $\bar{Q}_h$ in a tetrahedron T is somewhat arbitrary, since in general such a function is discontinuous at interelement boundaries. One may take for instance the three components of the gradient of the function in T and its own value at the barycenter of T.

Now let us address issue (2a) for the so-defined problem $(\mathscr{P}_h^1)$.

LEMMA 3 1 : *The expression* $\| \varepsilon(\vec{v}) \|_{0\,h}$ *defines a norm* $\forall \vec{v} \in \vec{V}_h + \vec{H}_0^1(\Omega)$.

Proof: Clearly the lemma will be proved if we verify that

$$\| \varepsilon(\vec{v}) \|_{0\,h} = 0 \Rightarrow \vec{v} = \vec{0} .$$

In this case we have

$$\varepsilon(\vec{v})|_T = O \quad \forall T \in \mathscr{T}_h .$$

Now, according to [13] this implies that in each tetrahedron T, $\vec{v}$ is of the form $\vec{a} + \vec{b} \wedge \vec{x}$ where $\vec{a}$ and $\vec{b}$ are two constant vectors, and $\wedge$ denotes the vector product. Let us take a tetrahedron T_0 having at least one face F contained in $\partial\Omega$. We then note that for every linear function v such as the components of $\vec{v}|_{T_0}$, the degree of freedom $f_l(v)$ associated with an edge l reduces to $v(M)$ where M is the mid-point of l. Moreover $v(M) = 0$ for $M \in F$ if $\vec{v} \in \vec{H}_0^1(\Omega) + \vec{V}_h$ Hence, letting M_1, M_2 and M_3 be the mid-points of the three edges of F with position vectors $\vec{x}_1$, $\vec{x}_2$ and $\vec{x}_3$, and letting $\vec{\tilde{x}}_i = \vec{x}_i - \vec{x}_3$, $i = 1, 2$, the condition $\vec{a} + \vec{b} \wedge \vec{x} = \vec{0}$ over F implies that

$$\vec{b} \wedge \vec{\tilde{x}}_1 = \vec{0} \quad \text{and} \quad \vec{b} \wedge \vec{\tilde{x}}_2 = \vec{0} .$$

Since the tetrahedrons of $\mathscr{T}_h$ are not degenerated by assumption, $\vec{\tilde{x}}_1$ and $\vec{\tilde{x}}_2$ are not parallel Thus $\vec{b} = \vec{0}$, which also implies that $\vec{a} = \vec{0}$. vanishes at M_1.

Now the same argument applies to every tetrahedron having a face common to T_0, since the three edge degrees of freedom of the components of $\vec{v}$ related to such a face necessarily coincide for both tetrahedrons whenever $\vec{v} \in \vec{V}_h + \vec{H}_0^1(\Omega)$ is piecewise linear. In this way we can sweep the whole partition $\mathcal{T}_h$ successively and we thus conclude that $\vec{v}\,|_T = \vec{0}$, $\forall T \in \mathcal{T}_h$. q.e.d. ■

Remark 3.1 : Although this is not necessary to carry out the analysis that will follow, it is possible to prove that L_h is uniformly continuous, or equivalently that the norm of L_h can be bounded above independently of h. This is a consequence of the validity of discrete Korn's third inequality, namely : There exists a constant K_3 independent of h such that

$$\forall \vec{v} \in \vec{V}_h, \quad \| \vec{v} \|_0 \leqslant K_3 \| \varepsilon(\vec{v}) \|_{0,h} .$$

The essential tool needed to prove this inequality is a regularity result for the hookean elasticity system in a polyhedron due to Grisvard [20]. ■

Let us now turn to issue (2b). In this respect, we first recall that, according to [29] the pair of spaces $\vec{V}_h$ and Q_h satisfies the following condition : $\exists \beta > 0$ independent of h such that

$$\forall q \in Q_h, \quad \sup_{\vec{v} \in \vec{V}_h} \frac{(q|\operatorname{div} \vec{v})_h}{|\vec{v}|_{1,h}} \geqslant \beta \| q \|_0 .$$

Therefore, since we trivially have

$$\| \varepsilon(\vec{v}) \|_{0,h} \leqslant |\vec{v}|_{1,h}, \quad \forall \vec{v} \in \vec{V}_h ,$$

condition $(\mathrm{vi})_h$ is satisfied in the present case with $\beta_h = \beta$.

Let us then prove that condition $(\mathrm{viii})_h$ also holds for the element studied in this Section, with a constant β' independent of h.

In order to do so let us first characterize $\vec{U}_h$ for the method under consideration.

LEMMA 3.2 : *Let $\vec{v} \in \vec{U}_h$ and $T \in \mathcal{T}_h$. Then the restriction of $\varepsilon(\vec{v})$ over T is a linear combination of the twenty-four tensors $\bar{\varepsilon}_{ij}$, $1 \leqslant i \leqslant 4$, $1 \leqslant j \leqslant 6$, given below :*

$$\left. \begin{cases} \bar{\varepsilon}_{ij} = \varepsilon_{ij} + 42 \displaystyle\sum_{k=1}^{3} \bar{x}_{i,k}\, \Delta_k , & \text{for} \quad 1 \leqslant j \leqslant 3 \\ \bar{\varepsilon}_{ij} = \varepsilon_{ij} , & \text{for} \quad 4 \leqslant j \leqslant 6 \end{cases} \right\} \; 1 \leqslant i \leqslant 4 . \tag{20}$$

with

$$\left\{\begin{array}{ll} \varepsilon_{ij} = \lambda_i \vec{e}_j \otimes \vec{e}_j , & for \quad 1 \leqslant j \leqslant 3 \\ & for \quad 4 \leqslant j \leqslant 6 \\ \varepsilon_{ij} = \lambda_i \displaystyle\sum_{\substack{k+l=j-1 \\ k \neq l \\ k\ l \in \{1,2,3\}}} \vec{e}_k \otimes \vec{e}_l , & \end{array}\right. \tag{21}$$

where $\vec{e}_j$ *represents the j-th vector of an orthonormal reference frame associated with coordinates* x_1, x_2 *and* x_3, *and* $\otimes$ *denotes the tensor product,*

$$\Delta_k = \frac{1}{2} \sum_{j=1}^{3} \frac{\partial \varphi_T}{\partial x_j} (\vec{e}_k \otimes \vec{e}_j + \vec{e}_j \otimes \vec{e}_k) ,$$

and $\bar{x}_{i,k} = x_{i,k} - x_{G_k}$, *where* $x_{i,k}$ *and* x_{G_k} *are respectively the k-th cartesian coordinate of i-th vertex and of the barycenter of T.*

Proof : First notice that the restriction of $\varepsilon(\vec{v})$ to a tetrahedron $T \in \mathcal{T}_h$ for $\vec{v} \in V_h$ is of the form

$$\varepsilon(\vec{v})|_T = \sum_{i=1}^{4} \sum_{j=1}^{6} c_{ij}\, \varepsilon_{ij} + \sum_{k=1}^{3} c_k\, \Delta_k \tag{22}$$

where the $c_{ij}'s$ and the $c_k's$ are real coefficients.

On the other hand, if $\vec{v} \in \vec{U}_h$ then

$$\int_T p Tr[\varepsilon(\vec{v})]\, dT = 0 , \quad \forall p \in P_1(T) .$$

Now take $p = x_l - x_{G_l}$, $l = 1, 2, 3$. Since $\forall i, Tr[\varepsilon_{ij}] = \lambda_i$ for $1 \leqslant j \leqslant 3$ and $Tr[\varepsilon_{ij}] = 0$ for $4 \leqslant j \leqslant 6$. from (22) we have

$$\sum_{i=1}^{4} \sum_{j=1}^{3} c_{ij} \int_T (x_l - x_{G_l})\, \lambda_i\, dT + \sum_{k=1}^{3} c_k \int_T (x_l - x_{G_l}) \frac{\partial \varphi_T}{\partial x_k} dT = 0, \quad l = 1, 2, 3 .$$

On the other hand,

$$\int_T (x_l - x_{G_l})\,\lambda_i\,dT = \int_T \left(\sum_{j=1}^{4} x_{j,l}\,\lambda_j\right)\lambda_i\,dT - \frac{x_{G_l}}{4}\,|T|$$

$$= \frac{1}{20}\left[\sum_{j=1}^{4} x_{j,l} + x_{i,l} - 5\,x_{G_l}\right]|T| = \frac{|T|}{20}\,(x_{i,l} - x_{G_l})\,.$$

Moreover, since φ_T vanishes on ∂T we have

$$-\sum_{k=1}^{3} c_k \int_T (x_l - x_{G_l})\,\frac{\partial \varphi_T}{\partial x_k}\,dT = \sum_{k=1}^{3} c_k \int_T \varphi_T\,\frac{\partial x_l}{\partial x_k}\,dT = c_l|T|/840\,.$$

Therefore we have

$$c_k = 42\sum_{i=1}^{4} \bar{x}_{i,k} \sum_{j=1}^{3} c_{ij}\,.$$

Recalling (22), it follows that $\forall \vec{v} \in \vec{V}_h$ we have

$$\varepsilon(\vec{v})|_T = \sum_{i=1}^{4}\left\{\sum_{j=1}^{3} c_{ij}\left[\varepsilon_{ij} + 42\sum_{k=1}^{3} \bar{x}_{i,k}\,\Delta_k\right] + \sum_{j=4}^{6} c_{ij}\,\varepsilon_{ij}\right\},$$

which leads to (20). q.e.d. ■

In the same way as in Section 2 the following result holds :

$$\forall \vec{v} \in \vec{U}_h, \quad \sup_{\substack{\tau \in \mathbf{L}^2_s(\Omega)\\ \neq O}} \frac{(\tau|\varepsilon(\vec{v}))_h}{\|\tau\|_0} = \frac{(\tau_0|\varepsilon(\vec{v}))_h}{\|\tau_0\|_0} = \|\varepsilon(\vec{v})\|_{0,h}$$

with $\tau_0|_T = \varepsilon(\vec{v})|_T$, $\forall T \in \mathcal{T}_h$. Hence, similarly to [16] and [34] we may assert that the result we are searching for in connection with condition (viii)$_h$ will hold, if we are able to construct $\tau_h \in T_h^C$ such that

$$(\tau_h|\varepsilon(\vec{v}))_h = (\tau_0|\varepsilon(\vec{v}))_h, \quad \forall \vec{v} \in \vec{U}_h \tag{23}$$

and

$$\|\tau_h\|_0 \leqslant \bar{C}\|\tau_0\|_0, \quad \text{with } \bar{C} \text{ independent of } h. \tag{24}$$

Clearly (23) will hold if $\forall T \in \mathcal{T}_h$ we have

$$(\tau_h|\varepsilon)_T = (\tau_0|\varepsilon)_T, \quad \forall \varepsilon \in E_s^T \tag{25}$$

where E_s^T is the space of symmetric tensors generated by the $\bar{\varepsilon}'_{ij}s$, $1 \leqslant i \leqslant 4$, $1 \leqslant j \leqslant 6$ given by (20).

Taking into account the fact that T_h^C is locally generated by the tensors $\tilde{\varepsilon}_{ij}$, $1 \leqslant i \leqslant 4$, $1 \leqslant j \leqslant 6$ defined like the $\varepsilon'_{ij}s$ by replacing in (21) λ_i with $\lambda_i \varphi_T$, assuming that (25) holds for a given τ_h, $\exists! \vec{t} \in \mathbb{R}^{24}$, $\vec{t} = (t_{11}, \dots, t_{16}, \dots, t_{41}, \dots, t_{46})$ such that

$$\tau_h|_T = \sum_{i=1}^{4} \sum_{j=1}^{6} t_{ij} \tilde{\varepsilon}_{ij}.$$

In so doing, setting in (25) ε successively equal to $\bar{\varepsilon}_{ij}$, $1 \leqslant i \leqslant 4$, $1 \leqslant j \leqslant 6$, and dividing both sides of the resulting relation by $|T|$, it is readily seen that (25) is equivalent to the system

$$A\vec{t} = \vec{t}^0$$

where $\vec{t}^0 = (t_{11}^0; \dots, t_{16}^0, \dots, t_{41}^0, \dots, t_{46}^0) \in \mathbb{R}^{24}$ is the vector given by

$$t_{ij}^0 = (\tau_0|\bar{\varepsilon}_{ij})_T/|T|.$$

Therefore, since every entry of A is an $O(1)$, if we prove that $\det A = O(1) \neq 0$, every entry of A^{-1} will also be an $O(1)$. This will imply in turn the existence of a constant independent of h that is an upper bound of $\|A^{-1}\|$. Such bound is essential to establish (24), as seen in Theorem 3.1 below (see also [34]).

Let us then study the invertibility of matrix $A = \{a_{mn,ij}\}$ given by

$$a_{mn,\,ij} = \frac{(\bar{\varepsilon}_{mn}|\tilde{\varepsilon}_{ij})_T}{|T|}\,. \tag{26}$$

For the sake of clearness it is convenient to consider A as a 12×12 block matrix, i.e., $A = \{A_{ij}\}_{i,j=1}^{2}$, where

- A_{11} is the matrix whose coefficients correspond to values $1 \leqslant j$, $n \leqslant 3$;
- A_{12} is the matrix whose coefficients correspond to values $4 \leqslant j \leqslant 6$ and $1 \leqslant n \leqslant 3$;
- A_{21} is the matrix whose coefficients correspond to values $1 \leqslant j \leqslant 3$ and $4 \leqslant n \leqslant 6$;
- A_{22} is the matrix whose coefficients correspond to values $4 \leqslant j$, $n \leqslant 6$.

As a consequence, we consider the following ordering of the unknowns : 11, 12, 13 ; 21, 22, 23 ; 31, 32, 33 ; 41, 42, 43/14, 15, 16 ; 24, 25, 26 ; 34, 35, 36 ; 44, 45, 46.

In so doing we have the following lemmas :

LEMMA 3.3 : A_{21} *is a null matrix.*

Proof : The lemma is a simple consequence of the fact that $\bar{\varepsilon}_{mn} = \varepsilon_{mn}$ for $4 \leqslant n \leqslant 6$. q.e.d. ■

LEMMA 3.4 : A_{22} *is the positive definite matrix whose entries are given by*

$$a_{mn,\,ij} = \frac{24}{9!}\,(2 + \delta_{im})\,\delta_{jn}, \quad 1 \leqslant i, m \leqslant 4, 4 \leqslant j, n \leqslant 6\,.$$

Proof : The values given above for the entries of A_{22} are obtained through a straightforward calculation by applying the following formula (*cf.* [38]) :

$$\int_T \prod_{i=1}^{4} \lambda_i^{n_i} = 6|T| \prod_{i=1}^{4} n_i!/(n_1 + n_2 + n_3 + n_4 + 3)!\,. \tag{27}$$

The fact that such a matrix is positive definite is a simple consequence of the following argument : $\forall \vec{y} \in \mathbb{R}^{12}$, $\vec{y} = (y_{14}, y_{15}, y_{16}, \dots, y_{44}, y_{45}, y_{46})$

$$\begin{aligned}
\frac{9!}{24} A_{22}\vec{y}\cdot\vec{y} &= \sum_{j,n=4}^{6}\ \sum_{i,m=1}^{4} a_{mn,\,ij}\, y_{mn}\, y_{ij} \\
&= \sum_{j=4}^{6}\left[\sum_{i=1}^{4}\sum_{m=1}^{4} 2\, y_{ij}\, y_{mj} + \sum_{i=1}^{4} y_{ij}^{2}\right] \\
&= \sum_{j=4}^{6}\left[2\left(\sum_{i=1}^{4} y_{ij}\right)^{2} + \sum_{i=1}^{4} y_{ij}^{2}\right] \geqslant |y|^{2}\,.
\end{aligned}$$

q.e.d. ■

We are primarily concerned about knowing whether det A vanishes or not. By virtue of Lemmas 3.3 and 3.4 the question may be shifted to det A_{11}, since $\det A = \det A_{11} \det A_{22}$. In order to prove that $\det A_{11} = O(1) \neq 0$, let us subdivide in turn A_{11} into sixteen blocks B^{mi}, $1 \leqslant i, m \leqslant 4$, where B^{mi} is the 3×3 matrix whose entry $[B^{mi}]_{jn}$ is $(\tilde{\varepsilon}_{mn}|\tilde{\varepsilon}_{ij})_T/|T|$ for $1 \leqslant j, n \leqslant 3$. We actually have :

LEMMA 3.5 : *With the above notations the entries of A_{11} are given by*

$$[B^{mi}]_{jn} = \frac{24}{11!}\,[55\,\delta_{nj}(2+\delta_{im}) - 84\, b_j^{mi}] \tag{28}$$

where $b_j^{mi} = \dfrac{\partial \lambda_i}{\partial x_j}\,\bar{x}_{m,j}$.

Proof : The lemma is a direct consequence of a straightforward calculation from (26) by applying again formula (27). q.e.d. ■

According to Lemma 3.5 matrix A does depend on T because of the expressions of b_j^{mi} given by (28). However we have :

LEMMA 3.6 : *The sum $\sum_{j=1}^{3} b_j^{mi}$ does not depend on T.*

Proof : We have

$$\sum_{j=1}^{3} b_j^{mi} = \overrightarrow{\text{grad}}\,\lambda_i \,.\, \vec{\bar{x}}_m$$

where $\vec{\bar{x}}_m$ is the vector leading from the barycenter of T to vertex S_m.
Moreover

$$\overrightarrow{\text{grad}}\,\lambda_i \,.\, \vec{\bar{x}}_m = \frac{3}{4}\,\frac{|l_m|}{|h_i|}\cos\theta_{im}\,,$$

where l_m is the median of T passing through vertex S_m, h_i is the height of T passing through vertex S_i, and θ_{im} is the angle between $\overrightarrow{\text{grad}}\,\lambda_i$ and $\vec{\bar{x}}_m$, that is, the angle formed by segments l_m and h_i duly oriented in the sense of the respective vertices. Since l_i is the hypotenusa of a right triangle whose catheti are h_i and the segments joining the intersection of h_i and l_i with the face opposite to S_i we have

$$\cos\theta_{ii} = \frac{|h_i|}{|l_i|}\,.$$

On the other hand, for $i \neq m$, by drawing a parallel to h_i passing through the centroid of the face opposite to S_m we construct another right triangle whose catheti are the lower fourth of h_i on this line, and the segment joining the intersection of the latter with the face opposite to S_i, to vertex S_m. Since l_m is the hypotenusa of the so-constructed triangle, taking into account the orientation of h_i and l_m, we have

$$\cos \theta_{im} = -\frac{|h_i|}{3|l_m|}, \quad \text{if } i \neq m .$$

It follows that

$$\sum_{j=1}^{3} b_j^{mi} = (-1)^{\delta_{im}+1} \frac{1+2\delta_{im}}{4}, \tag{29}$$

q.e.d. ■

As a final preparatory result we have.

LEMMA 3.7 : *The determinant of* A_{11} *is a constant independent of h.*

Proof: Let

$$d = \left(\frac{11!}{24}\right)^{12} \det A_{11} .$$

According to Lemmas 3.5 and 3.6 d is the determinant of the block matrix D_0 consisting of sixteen 3×3 blocks, where block D_0^{mi}, $1 \leqslant i$, $m \leqslant 4$ is of the form $D_0^{mi} = 110\, I - 84\, D_b^{mi}$ if $i \neq m$, and $D_0^{ii} = 165\, I - 84\, D_b^{ii}$, with

$$D_b^{mi} = \begin{bmatrix} b_1^{mi} & b_2^{mi} & b_3^{mi} \\ b_1^{mi} & b_2^{mi} & b_3^{mi} \\ b_1^{mi} & b_2^{mi} & b_3^{mi} \end{bmatrix} .$$

Similarly to [34], we first subtract the $(3k)$-th row from the $(3k-1)$-th and the $(3k-2)$-th rows of D_0, for $k = 1, 2, 3, 4$, threreby obtaining matrix D_1. Next we add the $(3k-2)$-th and the $(3k-1)$-th columns of D_1 to its $3k$-th column. Taking into account (29) we thus obtain a block matrix D_2 containing sixteen 3×3 blocks, where block D_2^{mi} is given by

$$D_2^{mi} = \begin{bmatrix} 110 & 0 & 0 \\ 0 & 110 & 0 \\ -84\, b_1^{mi} & -84\, b_2^{mi} & 131 \end{bmatrix}, \quad \text{if } i \neq m$$

and

$$D_2^{ll} = \begin{bmatrix} 165 & 0 & 0 \\ 0 & 165 & 0 \\ -84\,b_1^{ll} & -84\,b_2^{ll} & 102 \end{bmatrix}.$$

Next we notice that, by means of a suitable permutation of rows and columns in which the 3 k-th (resp. column) becomes the k-th row (resp. column), $k = 1, 2, 3, 4$, the determinant of D_2 — hence the one of D_0 — is the determinant of a matrix D_3 that is in fact the block matrix $\{D_3^{ij}\}_{i,j=1}^{2}$ of the following form :

- D_3^{11} is the 4×4 matrix whose diagonal entries are 102 and the off-diagonal entries are 131 ;
- D_3^{12} is a 4×8 matrix involving coefficients b_j^{mi}, $j = 1, 2$, $1 \leqslant m, i \leqslant 4$;
- $D_3^{21} = O$:
- D_3^{22} is the 8×8 block matrix whose 2×2 diagonal blocks are $165\,I$ and whose 2×2 off-diagonal blocks are $110\,I$.

As a consequence we have $\det D_0 = \det D_3^{11} \det D_3^{22}$.

Moreover in the same way as $\det D_3^{11}$, the determinant of D_3^{22} is the one of a 4×4 matrix whose diagonal entries are all equal to $\bar{r}$ and whose off-diagonal entries are all equal to $\bar{s}$, whereby for D_3^{22}, $\bar{r} = 165^2$ and $\bar{s} = 110^2$.

By a straightforward calculation we show that the value of such determinant is $(\bar{r} - \bar{s})^3(\bar{r} + 3\,\bar{s})$. It immediately follows that $\det D_0$, and hence $\det A_{11}$ is a non zero constant which proves the lemma. q.e.d. ■

Summarizing the arguments given above we have the following results.

THEOREM 3.1 : *$\forall \tau_0 \in \mathbf{L}_s^2(\Omega)$ there exists a unique $\tau_h \in \mathbf{T}_h$ that satisfies* (23) *and* (24).

Proof: This result is a consequence of Lemmas 3.3, 3.4 and 3.7. Indeed according to them the spectral norm of A^{-1} is bounded above by a constant independent of h. In this way from classical results there exist C_1 and C_2 such that

$$C_1 |T|^{-\frac{1}{2}} \|\bar{\tau}\|_{0,T} \leqslant |\vec{t}\,| \leqslant \|A^{-1}\| \, |\vec{t}_0| \leqslant C_2 \|\tau_0\|_{0,T} |T|^{-\frac{1}{2}}.$$

This implies the existence of $\bar{C}$ for which (24) holds. Finally recalling that (23) is fulfilled by construction, the result follows. q.e.d. ■

THEOREM 3.2 : *Condition* (iv)$_h$ *is satisfied with a constant α_h independent of h.*

Proof : According to Theorem 3.1 $(\text{viii})_h$ is satisfied with β'_h independent of h. Hence, recalling (19) and the fact that $(\text{vi})_h$ also holds with β_h independent of h, the result follows, q.e.d. ■

Now all that is left to do is treating condition (2c). For this purpose we have :

THEOREM 3.3 : *If Ω is a polyhedron, $\vec{u} \in \vec{H}^3(\Omega)$ and $p \in H^2(\Omega)$, and if the family $\{\mathcal{T}_h\}_h$ is quasiuniform, there exists $\mathscr{C}$ such that*

$$\sup_{(\tau,\, \vec{v},\, q) \in S_{\mathcal{T}_h}} |a_h((\sigma, \vec{u}, p), (\tau, \vec{v}, q)) - L_h((\tau, \vec{v}, q))| \leqslant$$

$$\leqslant \mathscr{C}[|\vec{u}_3| + |p|_2]\, h^2 .$$

Proof : First we note that $\sigma \in \mathbf{H}^2(\Omega)$. Thus since

$$-\overrightarrow{\operatorname{div}}\, \sigma + \overrightarrow{\operatorname{grad}}\, p = \vec{f} \text{ a.e. in } \Omega ,$$

we have

$$|a_h((\sigma, \vec{u}, p), (\tau, \vec{v}, q)) - L_h((\tau, \vec{v}, q))| =$$

$$\left| \sum_{T \in \mathcal{T}_h} \times \int_{\partial T} [\sigma - pI]\, \vec{n} \,.\, \vec{v}\, dS \right| .$$

Now standard arguments for non conforming elements (*cf.* [29]) lead to

$$\sum_{T \in \mathcal{T}_h} \int_{\partial T} [\sigma - pI]\, \vec{n} \,.\, \vec{v}\, dS = \sum_{T \in \mathcal{T}_h} \sum_{F \subset \partial T} \int_F \vec{w} \,.\, [\vec{v} - \pi_1^F(\vec{v})]\, dF ,$$

where $\pi_1^F(\vec{v})$ is the L^2-orthogonal projection of $\vec{v}$ onto $\vec{P}_1(F)$, and

$$\vec{w} = (\sigma - pI)\, \vec{n} \tag{30}$$

defined over T for $\vec{n}$ constant equal to the unit outer normal with respect to F.

Notice that by the Trace Theorem [23], for $F \subset \partial\Omega$ the sum of the restrictions of $\vec{w}$ to F from both sides of this face is zero. On the other hand, according to [29], for each face F, both orthogonal projections onto $P_1(F)$ in the sense of $L^2(F)$ of $\vec{v}\in\vec{V}_h$ restricted to F from the tetrahedrons containing this face coincide if $F\subset\partial\Omega$ The same projection vanishes if $F\subset\partial\Omega$ Thus

$$\sum_{T\in\mathcal{T}_l}\int_{\partial T}[\sigma-pI]\,\vec{n}\,.\,\vec{v}\,dS=\sum_{T\in\mathcal{T}_l}\sum_{F\subset\partial T}\times$$

$$\times\int_F[\vec{w}-\pi_1^F(\vec{w})]\,.\,[\vec{v}-\pi_1^F(\vec{v})]\,dF$$

Now setting ω'_T $\vec{H}^2(T)\times\vec{P}_2(T)\to\mathbb{R}$, where

$$\omega'_T(\vec{w},\vec{v})=\int_F[\vec{w}-\pi_1^F(\vec{w})]\,.\,[\vec{v}-\pi_1^F(\vec{v})]\,dF,$$

and going to the reference tetrahedron $\hat{T}$, for a face $\hat{F}$ of $\hat{T}$ corresponding to F we have

$$\frac{|\hat{F}|}{|F|}\omega'_T(\vec{w},\vec{v})=\hat{\omega}'_T(\hat{\vec{w}}\,.\,\hat{\vec{v}})\overset{\text{def}}{=}\int_{\hat{F}}[\hat{\vec{w}}-\pi_1^F(\hat{\vec{w}})]\,.\,[\hat{\vec{v}}-\pi_1^F(\hat{\vec{v}})]\,d\hat{F}\tag{31}$$

Equipping $P_2(\hat{T})$ with the norm $\|\,.\,\|_{2\,T}$, we notice that the form $\hat{\omega}'_T$ $\vec{H}^2(\hat{T})\times\vec{P}_2(\hat{T})\to\mathbb{R}$ is continuous

On the other hand, according to [10], and taking into account that

$$\hat{\omega}'_T(\hat{w},\hat{v}\ )=0\begin{cases}\forall\hat{\vec{w}}\in\vec{H}^2(\hat{T})\ \text{and}\ \ \forall\hat{\vec{v}}\in\vec{P}_1(\hat{T})\\ \forall\hat{\vec{w}}\in P_1(\hat{T})\ \ \text{and}\ \ \forall\hat{\vec{v}}\in P_2(\hat{T})\end{cases}$$

$\exists\hat{C}_1(\hat{T})$ such that $\hat{\omega}'_T(\hat{\vec{w}},\hat{\vec{v}})\leqslant\hat{C}_1(\hat{T})|\hat{\vec{w}}|_{2\,T}\,|\hat{\vec{v}}|_{2\,T}$ $\forall\hat{\vec{w}}\in\vec{H}^2(\hat{T})$ and $\forall\hat{\vec{v}}\in\vec{P}_2(\hat{T})$

Going back to the generic element T and recalling (30) and (31), following classical results there exists C_1 such that

$$\omega'_F(\vec{w},\vec{v})\leqslant C_1\,h^3[\,|\sigma|_{2\,T}^2+|p|_{2\,T}^2]^{1/2}\,|\vec{v}|_{2\,T},$$

which yields

$$\sum_{T\in\mathcal{T}_l}\int_{\partial T}(\sigma-p\mathbf{I})\,\vec{n}\,.\,\vec{v}\,dS\leqslant 4\,C_1\,h^3[\,|\vec{u}|_3+|p|_2]\,|\overrightarrow{\text{grad}}\,\vec{v}|_{1\,h}\tag{32}$$

On the other hand $\forall i, j, k \in \{1, 2, 3\}$ we have :

$$\frac{\partial^2 v_i}{\partial x_j \partial x_k} = \frac{1}{2}\frac{\partial}{\partial x_k}\left(\frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i}\right) + \frac{1}{2}\frac{\partial}{\partial x_j}\left(\frac{\partial v_i}{\partial x_k} + \frac{\partial v_k}{\partial x_i}\right) - \frac{1}{2}\frac{\partial}{\partial x_i}\left(\frac{\partial v_j}{\partial x_k} + \frac{\partial v_k}{\partial x_j}\right). \tag{33}$$

Thus according to classical inverse inequalities (*cf.* [10]) for families of quasiuniform partitions (33), implies the existence of constants C_2 and C_3 such that :

$$|\overrightarrow{\text{grad}}\ \vec{v}\,|_{1,h} \leqslant C_2 |\varepsilon(\vec{v}\,)|_{1,h} \leqslant \frac{C_3}{h}\, \|\varepsilon(\vec{v}\,)\|_{0,h}, \quad \forall \vec{v} \in \vec{V}_h . \tag{34}$$

Finally, recalling (32) we may assert that there exists $\mathcal{C}$, such that

$$\sup_{\substack{\vec{v} \in \vec{V}_h \\ \neq 0}} \frac{\left| \sum_{T \in \mathcal{T}} \int_{\partial T} [\sigma - p\mathbf{I}]\ \vec{n}\ .\ \vec{v}\ dS \right|}{\|\varepsilon(\vec{v}\,)\|_{0,h}} \leqslant \mathcal{C} h^2 [\,|\vec{u}\,|_3 + |p|_2] ,$$

which proves the Theorem. q.e.d. ■

Finally, recalling Lemma 3.1, Corollary 3.1 and (2a), (2b), (2c), and noticing that (2d) holds with $k = 2$, we readily have :

THEOREM 3.4 : *Under the assumptions of Theorem 3.3, there exists a constant C independent of h such that the solution* $(\sigma_h, \vec{u}_h, p_h)$ *of problem* (P_h^1) *satisfies :*

$$\|(\sigma, \vec{u}, p) - (\sigma_h, \vec{u}_h, p_h)\|_h \leqslant C h^2 [\,|\vec{u}\,|_3 + |p|_2] . \quad ■$$

Remark 3.2 : The regularity assumptions $\vec{u} \in \vec{H}^3(\Omega)$ and $p \in H^2(\Omega)$ are not fulfilled in general for the class of domains considered here, even if $\vec{f} \in \vec{H}^1(\Omega)$. Therefore, strictly speaking, the kind of analysis based on stronger regularity assumptions that led to Theorem 3.4, though currently employed in the litterature, should undergo some modifications in order to accommodate smoother domains. Here we mean to consider appropriate curved elements. ■

4. SOME FIRST ORDER METHODS

The method studied in the previous section is surely attractive from several standpoints. In this respect it should be stressed more particularly that second order approximations with a discontinuous pressure are obtained while the number of involved degrees of freedom is reduced to a minimum, at least as far as local stability analysis are concerned (see also [34]). However still that method has a high implementation cost for the class of problems under consideration, in terms of the computers available nowadays. For this reason we shall add in this section a brief study of some first order methods, which are significantly less costly for solving the three-field systems of the type considered in this work.

Remark 4.1 : All the methods to be presented hereafter are based on the Galerkin formulation. A Galerkin-least-squares method having no connection with those already considered in [17] is introduced in [35]. Actually it seems to be very promising not only due to the use of discontinuous pressures, but mainly because a relatively simple formulation involving only one parameter is employed. ■

Let us turn to the study of five new Galerkin type methods. Three of them stem from a particular two-field (velocity-pressure) method for tetrahedrons with a continuous pressure. First order convergence results were proven to hold for all of them, but here for the sake of conciseness we skip the details of those proofs. The other two methods are based on two-field methods with a discontinuous pressure for parallelotopes and tetrahedrons respectively. Although the author strongly conjectures that first order convergence results apply to both methods, it has not yet been possible to conclude them up to now, because the corresponding analyses involve a number of intrincate technicalities. The option to include them in this work is explained by the fact that both look very promising in terms of computational efficiency. Such *a priori* evaluation applies in particular to the parallelotopic element and actually for this reason it would be interesting to implement it in the near future.

All the Galerkin methods considered below are conforming. Thus, recalling the expression of the bilinear form for problem (6) given by (8), together with the expression of the right hand side given by (9), the approximate problem in this case reduces to

$$(\tilde{P}_h^1)\begin{cases}\text{Find } (\sigma_h, \vec{u}_h, p_h) \in Z_h \text{ such that} \\ a((\sigma_h, \vec{u}_h, p_h),(\tau, \vec{v}, q)) = L((\tau, \vec{v}, q)) \quad \forall (\tau, \vec{v}, q) \in Z_h\end{cases}$$

where

$$Z_h = \mathbf{T}_h \times \vec{V}_h \times Q_h \subset \mathbf{L}_s^2(\Omega) \times \vec{H}_0^1(\Omega) \times L_0^2(\Omega) .$$

Error bound (7) then simply becomes

$$\|(\sigma, \vec{u}, p) - (\sigma_h, \vec{u}_h, p_h)\|_Z \leqslant \frac{\|a\|}{\alpha_h} \inf_{z \in Z_h} \|(\sigma, \vec{u}, p) - z\|_Z \qquad (35)$$

where α_h is a strictly positive constant which supposedly satisfies

$$\sup_{z \in S_{Zh}} a(y, z) \geqslant \alpha_h \| y \|_Z \quad \forall y \in Z_h .$$

We further recall that, according to (22) we have

$$\alpha_h = \frac{(\beta_h \beta'_h)^2}{(4 + \beta_h^2)(1 + \beta_h'^2)}$$

where β_h and β'_hsatisfy respectively :

$$\beta_h > 0 \quad \text{and} \quad \sup_{\vec{v} \in S_{\vec{V}_h}} (q | \text{div } \vec{v}) \geqslant \beta_h \| q \|_0 \quad \forall q \in Q_h \qquad (36)$$

$$\beta'_h > 0 \quad \text{and} \quad \sup_{\tau \in S_{\mathbf{T}_h}} (\tau | \varepsilon(\vec{v})) \geqslant \beta'_h \| \vec{v} \|_1 \quad \forall \vec{v} \in \vec{V}_h . \qquad (37)$$

In the following we specify the spaces $\vec{V}_h$, Q_h and $\mathbf{T}_h$ defining the different methods that are introduced. We also state what has been possible to prove for each one of them in terms of inequalities (35), (36) and (37). Here again we assume that Ω is a polyhedron not necessarily convex and naturally enough, that $\mathcal{T}_h$ belongs to a quasiuniform family of partitions $\{\mathcal{T}_h\}_h$.

(I) Let $\mathcal{T}_h$ be a partition of Ω into tetrahedrons. For the first three methods we have :

- $V_h = \{ v \in C^0(\bar{\Omega}) \cap H_0^1(\Omega) | v_{|T} \in P_1(T) \oplus \{\varphi_T\} \quad \forall T \in \mathcal{T}_h \}$
- $Q_h = \tilde{Q}_h \cap L_0^2(\Omega)$ where

$$\tilde{Q}_h = \{ q \in C^0(\bar{\Omega}) | q_{|T} \in P_1(T) \quad \forall T \in \mathcal{T}_h \} .$$

The construction of this pair of spaces follows an author's proposal for treating incompressible media [26]. It fulfills condition (36) with β_h independent of h (*cf.* [3]). As for $\mathbf{T}_h$ we propose the following structures :

(Ia) $\mathbf{T}_h = \mathbf{T}_h^a \oplus \mathbf{T}_h^c$ where

$$\mathbf{T}_h^c = \{ \xi | \xi_{|T} = \text{constant} \quad \forall T \in \mathcal{T}_h \}$$

$$\mathbf{T}_h^a = \{ \xi | \xi_{|T} \in \{\tau_i\}_{i=1}^3 \quad \forall T \in \mathcal{T}_h \}$$

where $\tau_i = x_i \vec{e}_i \otimes \vec{e}_i$, $i = 1, 2, 3$.

(Ib) $\mathbf{T}_h = \mathbf{T}_h^b \oplus \mathbf{T}_h^c$ where

$$\mathbf{T}_h^b = \{ \xi | \xi_{|T} \in \{\varphi_T \tau_i\}_{i=1}^3 \quad \forall T \in \mathcal{T}_h \}$$

(Ic) $\mathbf{T}_h = \mathbf{T}_h^B \oplus \mathbf{T}_h^b \oplus \tilde{\mathbf{Q}}_H$ where

$$T_h^B = \{\xi \mid \xi_{|T} \in \{\varphi_T\} \quad \forall T \in \mathcal{T}_h\}$$

In all the above three cases condition (37) holds with a constant β_h' independent of h. As a consequence the right hand side of (35) may be estimated as follows, assuming that $\vec{u} \in \vec{H}^2(\Omega)$ and $p \in H^1(\Omega)$

$$\|(\sigma, \vec{u}, p) - (\sigma_h, \vec{u}_h, p_h)\|_Z \leqslant Ch(|\vec{u}|_2 + |p|_1)$$

with C independent of h.

Notice that among those three methods, only (Ic) allows the solution of viscoelastic flow systems with any approach, since the extra stress tensor is continuous. Indeed, this renders the extrastress transport term $\sum_{i=1}^{3} u_i \frac{\partial \sigma}{\partial x_i}$ computable in the discrete case, whatever algorithm is used to solve the nonlinear problem. Nevertheless, if a numerical technique based on Lesaint & Raviart' scheme [22] is employed, both methods (Ia) and (Ib) become feasible. In this context method (Ib) appears to be more efficient than (Ia) since only the stress jumps at inter-element boundaries of the constant components of T_h^c have to be taken into account (*cf.* [15]), while the role of the $\mathbf{T}_h^b$ components is to ensure the stability of the method In this respect we conjecture that the latter may even be omitted in this transport term, without any loss of accuracy.

(II) Let $\mathcal{T}_h$ be a partition of Ω into non degenerated convex parallelotopes, whose definition we recall : Letting $\hat{H}$ be the unit reference cube $[-1, 1] \times [-1, 1] \times [-1, 1]$ of $\mathbb{R}^3$, referred to a coordinate system $\vec{\hat{x}} = (\hat{x}_1, \hat{x}_2, \hat{x}_3)$, we consider $Q_1(\hat{H})$ to be the space of polynomials in $\vec{\hat{x}}$, of degree lesss than or equal to one in each variable $\hat{x}_i$. An element $H \in \mathcal{T}_h$ is the image of $\hat{H}$ through a given invertible mapping $\Phi_H : \mathbb{R}^3 \to \mathbb{R}^3$ such that $(\Phi_H)_i \in Q_1(\hat{H})$, $i = 1, 2, 3$

For this method we have :

- $V_h = W_h \cap H_0^1(\Omega)$ where

$$W_h = \{v \in C^0(\bar{\Omega}) \mid v_{|H} = \hat{v} \; o \, \Phi_H^{-1}, \hat{v} \in Q_1(\hat{H}) \quad \forall H \in \mathcal{T}_h\}$$

- $Q_h = \{q \mid q \in L_0^2(\Omega)$ and q is constant in $H \quad \forall H \in \mathcal{T}_h\}$
- $\mathbf{T}_h = \mathbf{W}_h \oplus \mathbf{T}_h^d$ where

$$\mathbf{T}_h^d = \{\xi \mid \xi_{|H} \in \{\eta_i\}_{i=1}^{18} \quad \forall H \in \mathcal{T}_h\}$$

with $\eta_i = \hat{\eta}_i o \Phi_H^{-1}$, $i = 1, 2, ..., 18$, $\hat{\eta}_i$ being defined over $\hat{H}$ and referred to the basis $\{\vec{\hat{e}}\}_{i=1}^{3}$, associated with $\vec{\hat{x}}$, where $\hat{\eta}_i = \hat{\varphi}\hat{\zeta}_i$, with $\hat{\varphi} = (1 - \hat{x}_1^2)(1 - \hat{x}_2^2)(1 - \hat{x}_3^2)$ and

$$\begin{array}{ll}
\hat{\zeta}_i = \vec{\hat{e}}_i \oplus \vec{\hat{e}}_i & \text{for } i = 1, 2, 3\,; \\
\hat{\zeta}_i = \vec{\hat{e}}_k \oplus \vec{\hat{e}}_j + \vec{\hat{e}}_j \oplus \vec{\hat{e}}_k & \text{with } j \neq k, \text{ for } i = j + k + 1\,; \\
\hat{\zeta}_i = \hat{x}_j\, \vec{\hat{e}}_k \otimes \vec{\hat{e}}_k & \text{with } j \neq k, \text{ for } \begin{cases} i = 2j + 3k \leqslant 9 \text{ and} \\ i = 2j + 3k - 1 \geqslant 10\,; \end{cases} \\
\hat{\zeta}_i = \hat{x}_l(\vec{\hat{e}}_k \otimes \vec{\hat{e}}_j + \vec{\hat{e}}_j \otimes \vec{\hat{e}}_k) & \text{with } j, k \text{ and } l \text{ distinct, for } i = 3l + 2(j + k)\,; \\
\hat{\zeta}_i = \hat{x}_k \hat{x}_j\, \vec{\hat{e}}_l \otimes \vec{\hat{e}}_l & \text{with } j, k \text{ and } l \text{ distinct, for } i = 2l + 3(j + k) + 1\,.
\end{array}$$

Remark 4.2 : From the computational point of view space T_h corresponds to an extra- stress finite element with three inner nodes, besides the vertices of the parallelotopes. ■

For such an element, in the case where every $H \in \mathcal{T}_h$ is rectangular, we are able to prove that (37) holds with a constant β_h' independent of h similarly to [33]. However it is well-known that to the best condition (36) holds with $\beta_h = O(h)$ (*cf.* [25]). The way out seems to be the attempt to prove like in [25] that, at least in the particular case where there is an even number of elements in each direction (Ω being also rectangular) the velocity converges with $|\vec{u} - \vec{u}_h|_1 = O(h)$, $\vec{u}$ and p being smoother than for methods (I). In this case we would also have $\|\sigma - \sigma_h\|_0 = O(h)$ under the same assumptions on the mesh and on $\vec{u}$ and p. As for the pressure, only a post-processing using as input the computed values of $\vec{u}_h$ and σ_h would yield reasonable accuracy. A description of such a procedure will be given in a forthcoming author's publication, as soon as computer tests involving this three-field finite element will be concluded.

Remark 4.3 : Theoretically in the case of non rectangular parallelotopes it is necessary to refer the η_i's to a particular frame for each element of $\mathcal{T}_h$. In this respect the author reports to the final remarks in [32]. However it seems that in practice the use of a fixed frame for all the elements causes no harm to stability. ■

(III) Let $\mathcal{T}_h$ be a partition of Ω into tetrahedrons. For this method we have :

- $\vec{V}_h = \vec{\bar{Q}}_h \cap \mathrm{H}_0^1(\Omega) \oplus \vec{V}_h^F$ where $\bar{Q}_h$ is the space related to the pressure space for elements (I) and

$$\vec{V}_h^F = \left\{ \vec{v} \mid \vec{v} \in \vec{H}_0^1(\Omega) \quad \text{and} \quad \vec{v}_{|T} \in \left\{ \frac{\varphi_T}{\lambda_i} \vec{n}_i \right\}_{i=1}^4 \quad \forall T \in \mathcal{T}_h \right\}$$

where $\vec{n}_i$, $i = 1, 2, 3, 4$ is the outer normal to the face opposite to vertex S_i of T.

- $Q_h = \{q \mid q \in L_0^2(\Omega)$ and $q_{|T} = \text{constant} \quad \forall T \in \mathcal{T}_h\}$
- $T_h = \bar{Q}_h \oplus T_h^B \oplus T_h^e$ where T_h^B is the space used in the definition of element (Ic) and

$$T_h^e = \{\xi \mid \xi_{|T} \in \{\zeta_i\}_{i=1}^4 \quad \forall T \in \mathcal{T}_h\}$$

where

$$\zeta_i = \lambda_i \varphi_T \vec{n}_i \otimes \vec{n}_i, \quad i = 1, 2, 3, 4 .$$

While on the one hand the pair $(\vec{V}_h, Q_h)$ fulfills condition (36) with β_h independent of h (*cf.* [19]), for the moment we can only conjecture that (37) holds with β_h' independent of h. In the present case the proof of such result involves intrincate and fastidious calculations, but in any case if (37) holds, then we have a three-field method with a discontinuous pressure for which (35) yields :

$$\|(\sigma, \vec{u}, p) - (\sigma_h, \vec{u}_h, p_h)\|_Z \leq Ch(|\vec{u}|_2 + |p|_1)$$

provided $\vec{u} \in \vec{H}^2(\Omega)$ and $p \in H^1(\Omega)$.

Remark 4.4 : This element has a variant with a nonconforming velocity in which in the definition of $\vec{V}_h^F$, φ_T/λ_i is replaced by

$$\sum_{\substack{j=1 \\ j \neq i}}^{3} \sum_{\substack{k=j+1 \\ k \neq i}}^{4} \lambda_j \lambda_k .$$

The so defined pair $(\vec{V}_h, Q_h)$ should satisfy condition (36) adapted with $(\,.\,|\,.\,)_h$ instead of $(\,.\,|\,.\,)$, according to [27] and [31]. The nonconformity term can be treated in a similar way to the case considered in [27]. Here again we conjecture that the right choice of T_h for (36) to hold with β_h' independent of h is the same as above, or yet a modification of this space in which φ_T is replaced with

$$\sum_{j,k=1, j \neq k}^{4} \lambda_j \lambda_k ,$$

in order to render the element computationally simpler. ■

Final Remark : A summary on the new elements for the three-field Stokes system in three-dimension space introduced by the author is given at the end of [35], together with some miscellaneous remarks. Among those, some comments are made on the computational efficiency of the Galerkin approach vs. the Galerkin least-squares one, in the context of viscoelastic flow.

REFERENCES

[1] G. ACQUADRO QUACCHIA, 1987, Resolução Computacional de um Problema de Viscoelasticidade Plana Incompressível via Elementos Finitos Mistos, Master's Dissertation. Pontifícia Universidade Católica do Rio de Janeiro.

[2] R. A. ADAMS, 1970, *Sobolev Spaces,* Academic Press, New York.

[3] D. N. ARNOLD, F. BREZZI and M. FORTIN, 1984, A stable finite element method for the Stokes equation, *Calcolo,* 21-4, pp. 337-344.

[4] G. ASTARITA and G. MARRUCCI, 1974, *Principles of Non-Newtonian Fluid Mechanics,* McGraw-Hill, New York.

[5] I. BABUSKA, 1973, The finite element method with lagrange multipliers, *Numer. Math.,* **20,** pp. 179-192.

[6] J. BARANGER and D. SANDRI, 1991, Approximation par éléments finis d'écoulements de fluides viscoélastiques. Existence de solutions approchées et majorations d'erreur. *C. R. Acad. Sci. Paris,* **312,** Série I, pp. 541-544.

[7] R. B. BIRD, R. C. ARMSTRONG and O. HASSAGER, 1987, *Dynamics of Polymeric Liquids,* Vol. 1, Fluid Mechanics, Second edition, John Wiley & Sons, New York.

[8] H. BRÉZIS, 1983, *Analyse Fonctionnelle, Théorie et Applications,* Masson, Paris.

[9] F. BREZZI, 1974, On the existence, uniqueness and approximation of saddle-point problems arising from lagrange multipliers, *RAIRO, Série rouge, Analyse Numérique R-2,* pp. 129-151.

[10] P. G. CIARLET, 1986, *The Finite Element Method for Elliptic Problems,* North-Holland, Amsterdam.

[11] M. J. CROCHET, A. R. DAVIES and K. WALTERS, 1984, *Numerical Simulation of Non-Newtonian Flow,* Elsevier, Amsterdam.

[12] B. DUPIRE, 1985, Problemas Variacionais Lineares, sua Aproximação e Formulações Mistas, Doctoral Thesis, Pontifícia Universidade Católica do Rio de Janeiro.

[13] G. DUVAUT, 1990, *Mécanique des Milieux Continus,* Masson, Paris.

[14] G. DUVAUT and J.-L. LIONS, 1972, *Les Inéquations en Mécanique et en Physique,* Masson, Paris.

[15] A. FORTIN and M. FORTIN, 1990, A preconditioned generalized minimal residual algorithm for the numerical solution of viscoelastic flows, *Journal of Non-Newtonian Fluid Mechanics,* **36,** pp. 277-288.

[16] M. FORTIN and R. PIERRE, 1988, On the convergence of the Mixed Method of Crochet & Marchal for Viscoelastic Flows, *Comp. Meth. Appl. Mech. Engin.,* **73,** pp. 341-350.

[17] L. FRANCA and R. STENBERG, 1991, Error analysis of some Galerkin-least-squares methods for the elasticity equations, *SIAM Journal of Numerical Analysis,* **28-6,** pp. 1680-1697.

[18] P. GERMAIN, 1973, *Cours de Mécanique des Milieux Continus,* Masson & Cie, Paris.

[19] V. GIRAULT and P. A. RAVIART, 1986, *Finite Element Methods for Navier-Stokes Equations,* Springer Series in Computational Mathematics 5, Springer-Verlag, Berlin

[20] P. GRISVARD, 1992, Singularities in Boundary Values Problems, in Research *Notes in Applied Mathematics,* P. G. Ciarlet & J.-L. Lions eds., Masson & Springer Verlag, Paris.

[21] O. LADYZHENSKAYA, 1963, *The Mathematical Theory of Viscous Incompressible Flow,* Gordon & Breach, Reading, Berkshire.

[22] P. LESAINT and P. A. RAVIART, 1976. *On a finite element method for solving the neutron transport equations,* in : *Mathematical Aspects of Finite Element Methods in Partial Differential Equations,* C. de Boor ed., Academic Press, New York.

[23] J.-L. LIONS and E. MAGENES, 1968, *Problèmes aux Limites Non-homogènes et Applications,* Dunod, Paris.

[24] M. A. MONTEIRO SILVA RAMOS, 1993, Um Modelo Numérico para a Simulação do Escoamento de Fluidos Viscoelásticos via Elementos Finitos, Doctoral Thesis, Pontifícia Universidade Católica do Rio de Janeiro.

[25] J. PITKÄRANTA, 1982, On a mixed finite element method for the Stokes Problem in $\mathbb{R}^3$, *RAIRO, Analyse Numérique,* **16-3,** pp. 275-291.

[26] V. RUAS, 1980, Sur l'application de quelques méthodes d'éléments finis à la résolution d'un problème d'élasticité incompressible non linéaire, *Rapport de Recherche 24,* INRIA, Rocquencourt, France.

[27] V. RUAS, 1982, Une méthode d'éléments finis non conformes en vitesse pour le problème de Stokes tridimensionnel. *Matemática Aplicada e Computacional,* **1-1,** pp. 53-74.

[28] V. RUAS, 1985, Une méthode mixte contrainte-déplacement-pression pour la résolution de problèmes de viscoélasticité incompressible en déformations planes, *C. R. Acad. Sc., Paris,* **301,** série II, 16, pp. 1171-1174.

[29] V. RUAS, 1985, Finite element solution of 3D viscous flow problems using non standard degrees of freedom, *Japan Journal of Applied Mathematics,* **2-2,** pp. 415-431.

[30] V. RUAS and J. H. CARNEIRO DE ARAUJO, 1992, Un método de elementos finitos quadrilatelares mejorado para el sistema de Stokes asociado al flujo de fluidos viscoelásticos, *Revista Internacional sobre Métodos Numéricos para Cálculo y Diseño en Ingeniería,* **8-1,** pp. 77-85.

[31] V. RUAS, 1992, Finite Element Methods for Three-Dimensional Incompressible Flow, in : *Finite Element in Fluids,* Vol. 8, T. J. Chung ed., Hemisphere Publishing Corporation, Washington, Chapter X, pp. 211-235.

[32] V. RUAS, 1992, A convergent three field quadrilateral finite element method for simulating viscoelastic flow on irregular meshes. *Revue Européenne des Éléments Finis,* **4-1,** pp. 391-406.

[33] V. RUAS, J. H. CARNEIRO DE ARAUJO, M. A. SILVA RAMOS, 1993, Approximation of the three-field Stokes system via optimized quadrilateral finite elements, *Modélisation Mathématiques et Analyse Numérique,* **27-1,** pp. 107-127.

[34] V. RUAS, 1994, An optimal three-field finite element approximation of the Stokes system with continuous extra stresses, *Japan Journal of Industrial and Applied Mathematics,* **11-1,** pp. 103-130.

[35] V. RUAS, Galerkin-least-squares finite element methods for the three-field Stokes system in three-dimension space, to appear.

[36] D. SANDRI, 1993, Analyse d'une formulation à trois champs du problème de Stokes, *Modélisation Mathématique et Analyse Numérique,* **27-7,** pp. 817-841.

[37] R. I. TANNER, 1985, *Engineering Rheology,* Claredon Press, Oxford.

[38] O. C. ZIENKIEWICZ, 1971, *The Finite Element Method in Engineering Science.* McGraw Hill, Maidenhead.