Available online at www.sciencedirect.com
ScienceDirect
Comput. Methods Appl. Mech. Engrg. 196 (2007) 3682–3692
Computer methods in applied mechanics and engineering
www.elsevier.com/locate/cma

A new discretization methodology for diffusion problems on generalized polyhedral meshes

Franco Brezzi [a,d], Konstantin Lipnikov [b,*], Mikhail Shashkov [b], Valeria Simoncini [c,d]

[a] CeSNA, Istituto Universitario di Studi Superiori, Pavia, Italy
[b] Los Alamos National Laboratory, Theoretical Division, MS B284, Los Alamos, NM 87545, United States
[c] Università di Bologna, Dipartimento di Matematica, Bologna, CIRSA, Ravenna, Italy
[d] IMATI-CNR, Pavia, Italy

Received 1 November 2005; accepted 30 October 2006
Available online 16 March 2007

Abstract

We develop a family of inexpensive discretization schemes for diffusion problems on generalized polyhedral meshes with elements having non-planar faces. The material properties are described by a full tensor. We also prove superconvergence for the scalar (pressure) variable under very general assumptions. The theoretical results are confirmed with numerical experiments. In the practically important case of logically cubic meshes with randomly perturbed nodes, the mixed finite element with the lowest order Raviart–Thomas elements does not converge while the proposed mimetic method has the optimal convergence rate.

Keywords: Finite difference; Compatible discretizations; Polyhedral meshes

1. Introduction

Tetrahedral and structured hexahedral meshes have been used for decades in the majority of engineering simulations; they are relatively easy to generate and there exists an enormous repository of numerical methods designed for these meshes. Nowadays, a growing number of complex simulations show advantage of using polyhedral meshes. For example, in the simulation of flow through a water jacket of an engine [14], the results obtained on a polyhedral mesh are more accurate than the results obtained on a tetrahedral mesh with a comparable number of elements. In oil reservoir simulations, the polyhedral mesh topology offers unlimited possibilities: elements can be automatically joined, split, or modified by introducing additional points, edges and faces to model complex geological features. Unfortunately, most of the existing numerical methods cannot be extended to polyhedral meshes, especially to meshes with elements having non-planar faces. This includes the practically important case of logically cubic meshes with randomly perturbed nodes.

In this paper we consider a diffusion problem, which appears in computational fluid dynamics, heat conduction, radiation transport, etc., and develop a *family* of simple inexpensive numerical schemes. This paper continues our analysis of the new discretization methodology that we began in [5]. The methodology follows the general principle of the mimetic finite difference (MFD) method – to mimic the essential underlying properties of the original continuum differential operators such as the conservation laws, solution symmetries, and the fundamental identities and theorems of vector and tensor calculus [7,11,12,4,6] (see also the book [15] and the references therein).

The mixed form of our diffusion problem is

$$\vec{F} = -\mathbb{K}\,\mathrm{grad}\,p, \quad \mathrm{div}\,\vec{F} = b, \tag{1.1}$$

where the first equation is the constitutive equation relating the scalar function p (pressure in flow simulations) to the

* Corresponding author.
E-mail addresses: brezzi@imati.cnr.it (F. Brezzi), lipnikov@lanl.gov (K. Lipnikov), shashkov@lanl.gov (M. Shashkov), valeria@dm.unibo.it (V. Simoncini).

0045-7825/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.cma.2006.10.028

flow field $\vec{F}$ and the second one is the mass conservation law. The material properties are described by the full symmetric tensor $\mathbb{K}$, and b is the source function. For this problem, the MFD method mimics the Gauss divergence theorem, the symmetry between the continuous gradient and divergence operators, and the null spaces of these operators. Therefore, it produces a discretization scheme which is symmetric and locally conservative.

In some sense, the MFD method lies between the standard mixed finite element (MFE) and finite volume (FV) methods. In the FV method, the fluxes are defined only at interfaces between mesh elements and a finite difference formula is used to discretize the constitutive equation. On the contrary, in the MFE method, a polynomial representation of the vector field inside each mesh element is used to define the inner product between vectors and then to write the constitutive equation by duality. This, however, can be done only for simple geometrical shapes. In the MFD method, there is notion of the inner product between vectors but the vector field inside a mesh element is never introduced explicitly. It is like a "guardian angel" who helps us prove convergence results but cannot be seen in the method formulation. Since the inner product is derived without any reconstruction, the practical implementation of the method is quite simple.

The MFD method developed in [4] (the *old* method) uses one degree of freedom per element to approximate the pressure and one degree of freedom per mesh face (the average normal component of the flow) to approximate the flow field. The same degrees of freedom are used in the mixed finite element method on tetrahedral and hexahedral meshes. We demonstrate with numerical experiments that both methods lack convergence on generalized polyhedral meshes.

The MFD method developed in [5] (the *new* method) uses three degrees of freedom (three average flow components) to approximate the flow field on non-planar faces. We proved that this recovers the optimal convergence rate on generalized polyhedral meshes, thus making our discretization methodology appealing in practical applications. When mesh elements are regular polyhedra, the new MFD method is reduced to the old one. When the element faces are strongly curved, the extra degrees of freedom allow the new method to succeed and perform much better than other methods.

The efficient implementation of the old MFD method has been analyzed in [6] where we derived a family of MFD methods with optimal convergence properties. In this article, we develop a family of MFD methods for generalized polyhedral meshes. Moreover, we prove superconvergence for the pressure (with an $O(h^2)$ rate) under very general assumptions. This result was already observed experimentally in the case of flat faces, but it was not proved. Here we prove it for both flat and curved faces. The key to this proof is to show existence of a reconstruction field inside a mesh element. Then, the superconvergence result follows from Theorem 5.3 in [5].

The outline of the paper is as follows. In Section 2, we present the mimetic finite difference method on generalized polyhedral meshes. In Section 3, we develop a family of efficient (inexpensive and easy-to-code) numerical schemes. In Section 4, we prove the superconvergence result for the scalar variable. In Section 5, the theoretical results are confirmed with numerical experiments on logically cubic and generalized polyhedral meshes.

2. A mimetic finite difference method

To simplify the presentation, we consider the homogeneous Dirichlet boundary value problem. Other types of boundary conditions are naturally embedded in the mimetic methodology [10].

Let $\Omega \in \Re^3$ be a domain with a Lipschitz continuous boundary. Furthermore, let Ω_h be a non-overlapping conformal partition of Ω into simply-connected *generalized polyhedral elements*. The generalized polyhedral element is, roughly speaking, the image of a polyhedral element under a bi-Lipschitz mapping, and can be thought as a "polyhedron" with possibly non-planar faces. Some basic assumptions of shape regularity are necessary to prove convergence estimates [5]; however most of these assumptions are not required until Section 3 and will be discussed there. To simplify the presentation, we assume that the tensor $\mathbb{K}$ is constant inside each mesh element but may strongly vary across mesh faces. We also assume that $\mathbb{K}$ is strongly elliptic, that is there exist two positive constants κ_* and κ^* such that

$$\kappa_* \|\mathbf{v}\|^2 \leqslant \|\mathbb{K}^{1/2}\mathbf{v}\|^2 \leqslant \kappa^* \|\mathbf{v}\|^2 \quad \forall \mathbf{v} \in \Re^3, \tag{2.1}$$

where $\|\cdot\|$ denotes the Euclidean norm.

The *first* step of the MFD method is to specify the degrees of freedom for the primary variables p and $\vec{F}$ which we shall refer to as the pressure and the flow, respectively. With a common abuse of language we shall often refer to $\vec{F}$ as the *velocity field* as well.

We consider the space Q^h of discrete pressures that are constant on each element E. For $\mathbf{q} \in Q^h$, we denote by q_E its value on E. The number, N_Q, of discrete pressure unknowns is equal to the number of mesh elements.

In order to introduce the space X^h of discrete velocities we have first to define, on each face of the decomposition, a reference system. For that, for every element E and for each face e of E we consider the unit outward normal $\mathbf{n}_E^e$, which varies continuously on e. Thus, we can define the *average normal vector* $\tilde{\mathbf{n}}_E^e$ as

$$\tilde{\mathbf{n}}_E^e = \frac{1}{|e|} \int_e \mathbf{n}_E^e \, d\Sigma, \tag{2.2}$$

where $|e|$ denotes the area of e. Later, we shall need the unit vector

$$\mathbf{a}_E^{e,3} = \frac{\tilde{\mathbf{n}}_E^e}{\|\tilde{\mathbf{n}}_E^e\|}.$$

It is not difficult to see that $\|\tilde{\mathbf{n}}_E^e\| \leqslant 1$ and equality is reached if and only if e is planar. It is also clear that if E_1 and E_2 are

two elements having the face e in common then $\tilde{\mathbf{n}}^e_{E_1} = -\tilde{\mathbf{n}}^e_{E_2}$. The same is obviously true for $\mathbf{a}^{e,3}_{E_1} = -\mathbf{a}^{e,3}_{E_2}$.

Then we associate to each face e two additional unit vectors $\mathbf{a}^{e,1}$ and $\mathbf{a}^{e,2}$ that are orthogonal to each other and to the vector $\tilde{\mathbf{n}}^e_E$. Note that (in contrast to $\mathbf{a}^{e,3}_E$ that points in the *outward* direction to E) the two vectors $\mathbf{a}^{e,1}$ and $\mathbf{a}^{e,2}$ depend on the face e but not on the element E.

The space X^h of discrete velocities is then defined as follows. To every element E and to every face e of E, we associate a constant *vector* $\mathbf{F}^e_E$. We will now make precise the continuity assumptions on our discrete velocity field. For this, we need to distinguish between moderately curved faces and strongly curved ones.

(M1) (*Moderately and strongly curved faces*). Let σ_* be a constant independent of the partition. Then, we say that the face e of the element E is *moderately curved* if at every point of e we have

$$\|\mathbf{n}^e_E - \tilde{\mathbf{n}}^e_E\| \leqslant \sigma_* |e|^{1/2}. \tag{2.3}$$

Otherwise, we say that the face e is *strongly curved*.

We impose the following *continuity* of the face-based velocity unknowns: we assume that for each face e, shared by two generalized polyhedrons E_1 and E_2, we have

$$\mathbf{F}^e_{E_1} \cdot \tilde{\mathbf{n}}^e_{E_1} = -\mathbf{F}^e_{E_2} \cdot \tilde{\mathbf{n}}^e_{E_2}. \tag{2.4}$$

Moreover, we assume that on *strongly curved faces* we have the full continuity of the discrete velocity vector. This means that together with (2.4) we also have

$$\mathbf{F}^e_{E_1} \cdot \mathbf{a}^{e,i} = \mathbf{F}^e_{E_2} \cdot \mathbf{a}^{e,i}, \quad i = 1, 2, \tag{2.5}$$

where the unit vectors $\mathbf{a}^{e,1}$ and $\mathbf{a}^{e,2}$ are the ones chosen above.

We denote the vector space of face-based velocity unknowns by X^h. The number, N_X, of our discrete velocity unknowns is equal to three times the number of boundary faces plus *six times* the number of internal faces. In our theoretical discussion, we shall consider X^h as the subspace of $\Re^{N_X}$ which verifies (2.4) on all faces and (2.5) on strongly curved faces.

On moderately curved faces, only the normal component of $\mathbf{F}^e_E$ is subject to the continuity requirements, and the other two components are treated as *internal degrees of freedom* and are eliminated during the assembly process by *static condensation*. Hence, in the final matrix, after static condensation, the total number of velocity unknowns equals the total number of moderately curved faces, plus three times the number of strongly curved faces.

The *second* step of the MFD method is to define suitable inner products in the discrete spaces. In the space Q^h, the inner product is almost straightforward:

$$[\mathbf{p}, \mathbf{q}]_{Q^h} = \sum_{E \in \Omega_h} p_E q_E |E|, \tag{2.6}$$

where $|E|$ is the volume of E. In the space X^h, the inner product is a sum of elemental inner products $[\mathbf{F}, \mathbf{G}]_E$ defined for every element E in Ω_h. Let $\mathbf{F}_E$ be the restriction of $\mathbf{F} \in X^h$ to element E. Furthermore, let k_E be the total number of faces in E, so that the total number of scalar components of $\mathbf{F}_E$ and $\mathbf{G}_E$ is $\ell_E = 3k_E$. We denote them by $\{\mathbf{F}_E\}_1, \ldots, \{\mathbf{F}_E\}_{\ell_E}$ and $\{\mathbf{G}_E\}_1, \ldots, \{\mathbf{G}_E\}_{\ell_E}$, respectively. For every positive integer number r, we define two unique integer numbers $\alpha(r)$ and $\beta(r)$ such that

$$r = 3(\alpha(r) - 1) + \beta(r), \quad \alpha(r) \geqslant 1, \ 1 \leqslant \beta(r) \leqslant 3.$$

Then, we say that $\{\mathbf{F}_E\}_r$ is associated with a face $e^{\alpha(r)}_E$ and a unit vector $\mathbf{a}^{e_{\alpha(r)},\beta(r)}_E$ (hereafter, we shall write $\mathbf{a}^{(r)}_E$ to simplify the notation).

Let us assume that we are given (for each E) a symmetric positive definite $\ell_E \times \ell_E$ matrix $\mathbb{M}_E$. Then, we set

$$[\mathbf{F}, \mathbf{G}]_E = \sum_{r,s=1}^{\ell_E} \mathbb{M}_{E,s,r} \{\mathbf{F}_E\}_s \{\mathbf{G}_E\}_r. \tag{2.7}$$

Here and in the sequel, $\mathbb{M}_{E,s,r}$ indicates the (s,r) entry of the given matrix $\mathbb{M}_E$. From (2.7), we can easily construct the inner product in X^h by setting

$$[\mathbf{F}, \mathbf{G}]_{X^h} = \sum_{E \in \Omega_h} [\mathbf{F}, \mathbf{G}]_E \quad \forall \mathbf{F}, \mathbf{G} \in X^h. \tag{2.8}$$

Some minimal approximation properties for the scalar product (2.7) are required, that make the construction of the matrix $\mathbb{M}_E$ a non-trivial task. We formulate and analyze these conditions in the next section.

The *third* step of the MFD method is to discretize the divergence operator. For each $\mathbf{G}$ in X^h, we define $\mathscr{DIV}^h\mathbf{G}$ as the element of Q^h such that

$$(\mathscr{DIV}^h\mathbf{G})_E := \frac{1}{|E|} \sum_{s=1}^{k_E} \mathbf{G}^{e_s}_E \cdot \tilde{\mathbf{n}}^{e_s}_E |e_s|. \tag{2.9}$$

Note that (2.9) is a discrete version of the Gauss divergence theorem.

The *fourth* step of the MFD method is to define the discrete flux operator, $\mathscr{G}^h$, as the adjoint to the discrete divergence operator, $\mathscr{DIV}^h$, with respect to the inner product (2.8), i.e.

$$[\mathbf{F}, \mathscr{G}^h\mathbf{p}]_{X^h} = [\mathbf{p}, \mathscr{DIV}^h\mathbf{F}]_{Q^h}, \quad \forall \mathbf{p} \in Q^h \ \forall \mathbf{F} \in X^h. \tag{2.10}$$

Using the discrete flux and divergence operators, the continuum problem (1.1) is discretized as follows:

$$\mathscr{DIV}^h\mathbf{F}_h = \mathbf{b}, \quad \mathbf{F}_h = \mathscr{G}^h\mathbf{p}_h, \tag{2.11}$$

where $\mathbf{b} \in Q^h$ is the vector of mean values of the source function b. This completes the derivation of the MFD method.

3. A family of accurate scalar products

The choice of the matrix $\mathbb{M}_E$ in the inner product (2.7) is not unique and every choice one makes will result in a new

MFD method. In this section, we describe a family of acceptable matrices $\mathbb{M}_E$. Recall our assumption that the tensor $\mathbb{K}$ has a constant value inside each mesh element E, which we denote by $\mathbb{K}_E$. To simplify the notation, we omit the subscript E unless it becomes necessary to avoid confusion.

3.1. Matrix algebraic equation

For every vector-valued function $\vec{G} \in (H^1(\Omega))^3$, we define $\mathbf{G}^I \in X^h$ as follows. To define the components of $(\mathbf{G}^I)_E^e$ in the three orthogonal directions, we set

$$(\mathbf{G}^I)_E^e \cdot \mathbf{a}_E^{e,3} := \frac{1}{|e|\|\tilde{\mathbf{n}}_E^e\|} \int_e \vec{G} \cdot \mathbf{n}_E^e \, d\Sigma \quad \text{and}$$

$$(\mathbf{G}^I)_E^e \cdot \mathbf{a}^{e,i} := \frac{1}{|e|} \int_e \vec{G} \cdot \mathbf{a}^{e,i} \, d\Sigma, \tag{3.1}$$

where $i = 1, 2$. If $\vec{G}$ is continuous across the interior mesh faces, it is easy to see that the resulting vector $\mathbf{G}^I$ will satisfy the continuity conditions (2.4) and (2.5). Hence $\mathbf{G}^I \in X^h$.

We begin our analysis with two conditions on the inner product (2.7) that are sufficient for getting a convergent MFD method [4].

(**S1**) There exist two positive constants s_* and S^* such that, for every element E, we have

$$s_* |E| \sum_{s=1}^{k_E} |\mathbf{G}_E^{e_s}|^2 \leqslant [\mathbf{G}, \mathbf{G}]_E \leqslant S^* |E| \sum_{s=1}^{k_E} |\mathbf{G}_E^{e_s}|^2 \quad \forall \mathbf{G} \in X^h. \tag{3.2}$$

(**S2**) For every element E, every linear function q^1 on E, and every $\mathbf{G} \in X^h$, we have

$$[(\mathbb{K}\nabla q^1)^I, \mathbf{G}]_E + \int_E q^1 (\mathscr{D}\mathscr{I}\mathscr{V}^h \mathbf{G})_E \, dV = \int_{\partial E} q^1 \mathbf{G}_E \cdot \mathbf{n}_E \, d\Sigma. \tag{3.3}$$

Assumption (**S1**) states that the matrix $\mathbb{M}_E$ is spectrally equivalent to the scalar matrix $|E|\mathbb{I}_{\ell_E}$ where $\mathbb{I}_{\ell_E}$ is the $\ell_E \times \ell_E$ identity matrix. In practice, the constants s_* and S^* are expected to depend only on the skewness of the mesh elements and on the tensor $\mathbb{K}$.

Assumption (**S2**) is the discrete Gauss–Green formula with the constant velocity $\mathbb{K}\nabla q^1$. Since $\mathscr{D}\mathscr{I}\mathscr{V}^h\mathbf{G}$ is a constant, the second term in (3.3) can be easily computed. Also, note that all terms in (3.3) are linear functionals of q^1. For each q^1, this assumption results in a system of linear equations where the unknowns are the coefficients of the matrix $\mathbb{M}_E$.

Taking $q^1 = 1$ in (3.3), we get the formula for the discrete divergence operator. As we obviously expect frame invariance, we use this freedom and, for every element E, we set the origin in center of mass of E, which simplifies the construction of the matrix $\mathbb{M}_E$. Thus, Assumption (**S2**) can be replaced by the following one.

(**S2′**) For every element E with center of mass at the origin, for each $i = 1, 2, 3$, and for each $s = 1, \ldots, \ell_E$, the $\ell_E \times \ell_E$ matrix $\mathbb{M}_E$ satisfies,

$$\sum_{r=1}^{\ell_E} \mathbb{M}_{E,s,r} \{(\mathbb{K}\nabla x_i)^I\}_r = \int_{e^{\alpha(s)}} x_i \mathbf{a}^{(s)} \cdot \mathbf{n}_E \, d\Sigma, \tag{3.4}$$

where (x_1, x_2, x_3) are the Cartesian coordinates.

We continue by pointing out the Gauss–Green formula for linear functions x_i and x_j:

$$\int_{\partial E} (\mathbb{K}\nabla x_i) \cdot \mathbf{n}_E x_j \, d\Sigma = \int_E \mathbb{K}\nabla x_i \cdot \nabla x_j \, dV = |E| \mathbb{K}_{i,j}. \tag{3.5}$$

If we further introduce the $\ell_E \times 3$ matrices $\mathbb{R}$ and $\mathbb{D}$ by

$$\mathbb{R}_{s,i} = \int_{e^{\alpha(s)}} \mathbf{a}^{(s)} \cdot \mathbf{n}_E x_i \, d\Sigma \quad \text{and} \quad \mathbb{D}_{s,i} = \{(\mathbb{K}\nabla x_i)^I\}_s, \tag{3.6}$$

where $s = 1, 2, \ldots, \ell_E$ and $i = 1, 2, 3$, then the identity (3.5) becomes

$$\mathbb{R}^{\mathrm{T}} \mathbb{D} = |E| \mathbb{K}, \tag{3.7}$$

implying, among other things, that both matrices $\mathbb{D}$ and $\mathbb{R}$ have full rank 3. Using (3.6), Assumption (**S2′**) thus becomes

$$\mathbb{M}_E \mathbb{D} = \mathbb{R}. \tag{3.8}$$

Next, we shall construct $\mathbb{M}_E$ as the sum of two positive symmetric semi-definite matrices, $\mathbb{M}_E = \mathbb{M}_0 + \mathbb{M}_1$, where $\mathbb{M}_0$ satisfies (3.8) and $\mathbb{M}_1 \mathbb{D} = 0$.

Lemma 3.1. *Let $\mathbb{R}$ be given by (3.6). Then, the symmetric and positive semi-definite matrix*

$$\mathbb{M}_0 \equiv \frac{1}{|E|} \mathbb{R}\mathbb{K}^{-1}\mathbb{R}^{\mathrm{T}} \tag{3.9}$$

satisfies (3.8).

Proof. From (3.7) and (3.8) we have $\mathbb{M}_0\mathbb{D} = \frac{1}{|E|}\mathbb{R}\mathbb{K}^{-1}\mathbb{R}^{\mathrm{T}}\mathbb{D} = \mathbb{R}$. □

Since the matrix $\mathbb{M}_0$ is only positive semi-definite, assumption (**S1**) does not hold. The following result shows how $\mathbb{M}_0$ can be completed to meet the positive definiteness requirement.

Theorem 3.2. *Let $\mathbb{C}$ be an $\ell_E \times (\ell_E - 3)$ matrix whose $\ell_E - 3$ columns span the null space of the full rank matrix $\mathbb{D}^{\mathrm{T}}$, so that $\mathbb{D}^{\mathrm{T}}\mathbb{C} = 0$. Then, for every $(\ell_E - 3) \times (\ell_E - 3)$ symmetric positive definite matrix $\mathbb{U}$, the following symmetric matrix:*

$$\mathbb{M}_E = \mathbb{M}_0 + \mathbb{C}\mathbb{U}\mathbb{C}^{\mathrm{T}} \tag{3.10}$$

satisfies (3.8) and is positive definite.

Proof. By construction, $\mathbb{M}_E\mathbb{D} = \mathbb{M}_0\mathbb{D}$, and therefore by Lemma 3.1, the matrix $\mathbb{M}_E$ satisfies (3.8). Moreover, again

by construction, $\mathbb{M}_E$ is symmetric and positive semi-definite. We show that it is non-singular. Let us assume that there exists a non-zero vector $\mathbf{v}$ such that $\mathbb{M}_E\mathbf{v} = 0$. Then we must have

$$\left\| \frac{1}{|E|^{1/2}} \mathbb{K}^{-1/2}\mathbb{R}^{\mathrm{T}}\mathbf{v} \right\|^2 + \|\mathbb{U}^{1/2}\mathbb{C}^{\mathrm{T}}\mathbf{v}\|^2 = 0, \tag{3.11}$$

which in turn implies that $\mathbb{R}^{\mathrm{T}}\mathbf{v} = 0$ and $\mathbb{C}^{\mathrm{T}}\mathbf{v} = 0$. Hence $(\mathbf{v}, \mathbb{C}\mathbf{u}) = 0$ for any vector $\mathbf{u}$ in $\mathfrak{R}^{\ell_E}$, and therefore we get

$$\mathbf{v} \in \{\mathrm{im}(\mathbb{C})\}^{\perp} = \{\ker(\mathbb{D}^{\mathrm{T}})\}^{\perp} = \mathrm{im}(\mathbb{D}),$$

so that $\mathbb{R}^{\mathrm{T}}\mathbf{v} = \mathbb{R}^{\mathrm{T}}\mathbb{D}\mathbf{w} = 0$ for some $\mathbf{w} \in \mathfrak{R}^3$. Now the identity (3.7) implies that $\mathbf{w} = 0$, so that $\mathbf{v} = 0$, and the non-singularity of $\mathbb{M}_E$ follows. □

Since $\mathbb{U}$ has size $\ell_E - 3$, a general symmetric positive definite matrix of this size has $(\ell_E - 2)(\ell_E - 3)/2$ free parameters, yielding a family of matrices with the required properties. The liberty of choosing $\mathbb{U}$ within this family could be used to tackle other computational problems, e.g., enforcement the discrete maximum principle.

One of the efficient ways for solving the discrete problem (2.11) is based on the KKT theory of constrained minimization (see e.g. [13, Chapter 16]) where the constraints are given by (2.4) and (2.5). The solution of the KKT system is reduced to the solution of a sparse system for Lagrange multipliers with a symmetric positive definite matrix. This is what in the Finite Element context is often called *hybridization* and is usually attributed to Fraeijs de Veubeke [8] (see also [1], or [3] pp. 178–181). The procedure requires the inversion of matrix $\mathbb{M}_E$. More precisely, during the whole procedure we *only* need the matrix $\mathbb{M}_E^{-1}$, while the explicit knowledge of the matrix $\mathbb{M}_E$ is not required. We show that we can directly compute a matrix $\mathbb{W}_E$, the inverse of an inner product matrix, with the required properties.

Theorem 3.3. *Let $\mathbb{Q}$ be a $\ell_E \times (\ell_E - 3)$ matrix whose $\ell_E - 3$ columns span the null space of the full rank matrix $\mathbb{R}^{\mathrm{T}}$, so that $\mathbb{R}^{\mathrm{T}}\mathbb{Q} = 0$. Then, for every $(\ell_E - 3) \times (\ell_E - 3)$ symmetric positive definite matrix $\widetilde{\mathbb{U}}$, the following symmetric matrix*

$$\mathbb{W}_E := \frac{1}{|E|}\mathbb{D}\mathbb{K}_E^{-1}\mathbb{D}^{\mathrm{T}} + \mathbb{Q}\widetilde{\mathbb{U}}\mathbb{Q}^{\mathrm{T}} \tag{3.12}$$

satisfies $\mathbb{W}_E\mathbb{R} = \mathbb{D}$ and is positive definite.

The proof of this result follows the proofs of Lemma 3.1 and Theorem 3.2; therefore, it is omitted. Note that the matrix $\mathbb{D}$ contains the material properties and thus the first term in (3.12) is scaled properly.

Since, in practice, we are interested *only* in the matrix $\mathbb{M}_E^{-1}$, we could define $\mathbb{M}_E^{-1} := \mathbb{W}_E$. Indeed, the matrix $\mathbb{M}_E$ defined in this way will be symmetric positive definite, and will satisfy (3.8). Moreover, it is not difficult to see that the matrix $\mathbb{M}_E := \mathbb{W}_E^{-1}$ can still be written in the form (3.10), where the choice of the matrices $\mathbb{U}$ and $\mathbb{C}$ obviously depends on the choice of $\widetilde{\mathbb{U}}$ and $\mathbb{Q}$. In Section 3.3 we explicitly derive a matrix $\mathbb{Q}$ that satisfies the hypotheses of Theorem 3.3, and we provide the computational costs associated with the use of $\mathbb{W}_E$.

3.2. Spectral analysis

Assumption (**S1**) imposes some restrictions on the choice of the *parameter* matrix $\mathbb{U}$ in Theorem 3.2 (or on $\widetilde{\mathbb{U}}$ in Theorem 3.3), and requires fixing some further hypotheses on the shape-regularity of the mesh elements formulated in [4,5]. They hold for basically all meshes which are not totally unreasonable, thus making our discretization methodology appealing in practical applications. For instance, they allow degenerate and non-convex elements. Let h_E denote a diameter of E and let the following assumptions hold:

(**M2**) There exist a positive integer N_e such that every element E has at most N_e faces.

(**M3**) There exist a positive number γ_* such that, for every generalized polyhedron E, there exist a polyhedron E_0 (with planar faces $e_{0,s}$) and a radial map Φ with center at a point c_E and such that $\Phi(E_0) = E$, $\Phi(e_{0,s}) = e_s$,

$$\|\mathscr{J}(\Phi)\| \leqslant \gamma_*, \quad \text{and} \quad \|\mathscr{J}(\Phi^{-1})\| \leqslant \gamma_*, \tag{3.13}$$

where $\mathscr{J}$ denotes the Jacobi matrix.

(**M4**) There exists a positive number τ_* such that every element E and the corresponding polyhedron E_0 are star-shaped with respect to every point of a common sphere of radius $\tau_* h_E$ centered at the point c_E.

Before entering the discussion on Assumption (**S1**), we re-scale the matrices $\mathbb{D}$ and $\mathbb{R}$ and prove a technical lemma. Let us define

$$\widetilde{\mathbb{D}} := \mathbb{D}\mathbb{K}^{-1} \quad \text{and} \quad \widetilde{\mathbb{R}} := \frac{1}{|E|}\mathbb{R}, \tag{3.14}$$

so that

$$\widetilde{\mathbb{R}}^{\mathrm{T}}\widetilde{\mathbb{D}} = \widetilde{\mathbb{D}}^{\mathrm{T}}\widetilde{\mathbb{R}} = \mathbb{I}_3. \tag{3.15}$$

It is not difficult to see that the rth row of the matrix $\widetilde{\mathbb{D}}$ is $(\mathbf{a}_E^{(r)})^{\mathrm{T}}$. Let D_s be the 3×3 matrix whose rows are the orthonormal vectors $\mathbf{a}^{e_s,1}$, $\mathbf{a}^{e_s,2}$, and $\mathbf{a}_E^{e_s,3}$ of the face e_s and $\mathscr{D} = \mathrm{diag}\{D_1, \ldots, D_{k_E}\}$.

Then,

$$D_s D_s^{\mathrm{T}} = D_s^{\mathrm{T}} D_s = \mathbb{I}_3 \quad \text{and} \quad \mathscr{D}\mathscr{D}^{\mathrm{T}} = \mathscr{D}^{\mathrm{T}}\mathscr{D} = \mathbb{I}_{\ell_E}. \tag{3.16}$$

If we further introduce the $\ell_E \times 3$ matrix $\mathbb{N}$ by

$$\mathbb{N}_{s,i} = \int_{e^{x(s)}} \nabla x_{\beta(s)} \cdot \mathbf{n}_E x_i \, \mathrm{d}\Sigma,$$

where $s = 1, \ldots, \ell_E$ and $i = 1, 2, 3$, then

$$\mathbb{R} = \mathscr{D}\mathbb{N} \quad \text{and} \quad \widetilde{\mathbb{D}} = \begin{pmatrix} D_1 \\ D_2 \\ \vdots \\ D_{k_E} \end{pmatrix}. \tag{3.17}$$

With the notation above, the following bounds hold.

Lemma 3.4. *Assume that* (**M3**) *and* (**M4**) *hold. Then for every element E we have the following bounds*:

$$\|\widetilde{\mathbb{D}}\mathbf{w}\| = \sqrt{k_E}\|\mathbf{w}\| \quad \text{and} \quad \frac{1}{\sqrt{k_E}} \leqslant \frac{\|\widetilde{\mathbb{R}}\mathbf{w}\|}{\|\mathbf{w}\|} \leqslant \frac{3\gamma_*^2}{\tau_*}, \quad \forall \mathbf{w} \neq 0. \tag{3.18}$$

Proof. Using (3.17), for every $\mathbf{w} \in \mathfrak{R}^3$ we have $\|\widetilde{\mathbb{D}}\mathbf{w}\|^2 = \mathbf{w}^{\mathrm{T}}\widetilde{\mathbb{D}}^{\mathrm{T}}\widetilde{\mathbb{D}}\mathbf{w} = k_E\mathbf{w}^{\mathrm{T}}\mathbf{w}$, which proves the equality in (3.18). To estimate the norm of $\widetilde{\mathbb{R}}$, we note that

$$|E|^2\|\widetilde{\mathbb{R}}\mathbf{w}\|^2 = \|\mathbb{N}\mathbf{w}\|^2 = \sum_{s=1}^{k_E}\left\|\int_{e_s}\mathbf{n}_E(\mathbf{w}\cdot\mathbf{x})\,\mathrm{d}\Sigma\right\|^2. \tag{3.19}$$

Recall that we put the origin in the center of mass of E, so that $\|\mathbf{x}\| \leqslant h_E$ for any $\mathbf{x}$ in E. Thus

$$\begin{aligned} |E|^2\|\widetilde{\mathbb{R}}\mathbf{w}\|^2 &\leqslant \|\mathbf{w}\|^2\sum_{s=1}^{k_E}|e_s|\int_{e_s}\|\mathbf{x}\|^2\,\mathrm{d}\Sigma \\ &\leqslant \|\mathbf{w}\|^2 h_E^2\left(\sum_{s=1}^{k_E}|e_s|\right)^2. \end{aligned} \tag{3.20}$$

Now, we consider the pyramids $P_{0,s}$ having $e_{0,s}$ as bases, and the point c_E from Assumption (**M3**) as common vertex. Assumption (**M4**) implies that the height, $h_{0,s}$, of the pyramid $P_{0,s}$ is bigger than $\tau_* h_E$. Assumption (**M3**) implies that the volume of E is bounded by the volume of E_0. More precisely, we have

$$\begin{aligned} |E| &\geqslant \frac{1}{\gamma_*}|E_0| = \frac{1}{\gamma_*}\sum_{s=1}^{k_E}|P_{0,s}| = \frac{1}{3\gamma_*}\sum_{s=1}^{k_E}|e_{0,s}|h_{0,s} \\ &\geqslant \frac{\tau_* h_E}{3\gamma_*}\sum_{s=1}^{k_E}|e_{0,s}| \geqslant \frac{\tau_* h_E}{3\gamma_*^2}\sum_{s=1}^{k_E}|e_s|. \end{aligned}$$

Inserting this in (3.20), we have

$$\|\widetilde{\mathbb{R}}\mathbf{w}\|^2 \leqslant \frac{9\gamma_*^4}{\tau_*^2}\|\mathbf{w}\|^2. \tag{3.21}$$

The proof of the lower bound follows from the Gauss–Green formula

$$\int_{\partial E} n_{E,i}(\mathbf{w}\cdot\mathbf{x})\,\mathrm{d}\Sigma = \int_E \nabla x_i\cdot\mathbf{w}\,\mathrm{d}V = w_i|E|.$$

Applying this result to (3.19), we get

$$\begin{aligned} |E|^2\|\widetilde{\mathbb{R}}\mathbf{w}\|^2 &\geqslant \frac{1}{k_E}\sum_{i=1}^{3}\left(\sum_{s=1}^{k_E}\int_{e_s}n_{E,i}(\mathbf{w}\cdot\mathbf{x})\,\mathrm{d}\Sigma\right)^2 \\ &\geqslant \frac{1}{k_E}\sum_{i=1}^{3}|E|^2 w_i^2 = \frac{|E|^2}{k_E}\|\mathbf{w}\|^2. \end{aligned}$$

This proves the assertion of the lemma. □

From Lemma 3.4 we may easily obtain estimates for the *unscaled* matrices $\mathbb{R}$ and $\mathbb{N}$ and their products with the tensor $\mathbb{K}$. In particular, using Assumption (**M2**), we may prove that

$$\frac{1}{(N_e\kappa^*)^{1/2}}|E| \leqslant \frac{\|\mathbb{K}^{-1/2}\mathbb{R}^{\mathrm{T}}\mathbf{w}\|}{\|\mathbf{w}\|} \leqslant \frac{3\gamma_*^2}{\kappa_*^{1/2}\tau_*}|E|, \quad \forall \mathbf{w} \neq 0. \tag{3.22}$$

It is obvious that the matrix $\mathbb{M}_E$ will satisfy Assumption (**S1**) *if and only if* its inverse matrix satisfies it. Hence, in what follows, we discuss only the case of the matrix $\mathbb{M}_E$. If one decides to follow the path of Theorem 3.3 (constructing directly the matrix $\mathbb{W}_E = \mathbb{M}_E^{-1}$), the same arguments will hold for $\mathbb{W}_E$ as well.

Theorem 3.5. *Let the assumptions of Theorem 3.2 and Lemma 3.4 hold. Assume further that there exist two positive constants s_U^* and S_U^*, independent of E, such that*

$$s_U^*|E|\|\mathbf{v}\|^2 \leqslant \|\mathbb{U}^{1/2}\mathbb{C}^{\mathrm{T}}\mathbf{v}\|^2 \quad \forall \mathbf{v} \in \mathrm{im}(\mathbb{C}) \tag{3.23}$$

and

$$\|\mathbb{U}^{1/2}\mathbb{C}^{\mathrm{T}}\mathbf{v}\|^2 \leqslant S_U^*|E|\|\mathbf{v}\|^2 \quad \forall \mathbf{v} \in \mathfrak{R}^{\ell_E - 3}. \tag{3.24}$$

Then, the matrix $\mathbb{M}_E$ constructed as in (3.10) satisfies Assumption (**S1**). *In particular, we have*

$$\min\left\{\frac{1}{2}s_U^*, \sigma_*\right\}|E|\|\mathbf{v}\|^2 \leqslant \|\mathbb{M}_E^{1/2}\mathbf{v}\|^2 \leqslant \max\{S_U^*, \sigma^*\}|E|\|\mathbf{v}\|^2, \tag{3.25}$$

where

$$\sigma_* = \frac{\kappa_*\tau_*^2 s_U^*}{N_e\kappa^*(18\gamma_*^4 + s_U^*\kappa_*\tau_*^2)} \quad \text{and} \quad \sigma^* = \frac{9\gamma_*^4}{\kappa_*\tau_*^2}.$$

The proof of this theorem follows closely the proof of Theorem 3.6 in [6]; therefore, it is omitted.

In actual numerical computations (based on Theorem 3.2), we recommend to multiply the matrix $\mathbb{U}$ by a characteristics value of $\mathbb{K}_E^{-1}$, for example, its trace. This will improve the spectral properties of the matrix $\mathbb{M}_E$ with respect to material properties. The estimates in (3.25) provide an illustration of the practical role of $\mathbb{U}$ in the conditioning of $\mathbb{M}_E$. As long as the extreme eigenvalues of $\mathbb{U}$ are within those of $\mathbb{K}_E^{-1}$, the conditioning of $\mathbb{M}_E$ is not strongly affected by the choice of $\mathbb{U}$. The same remark obviously applies to the matrix $\widetilde{\mathbb{U}}$, if we decide to use Theorem 3.3 to construct directly the matrix $\mathbb{M}_E^{-1}$. This latter approach is what we have employed in our experiments.

3.3. Computational considerations

According to Theorem 3.3, the most computationally demanding part in building the matrix $\mathbb{M}_E^{-1} = \mathbb{W}_E$ is the construction of the matrix $\mathbb{Q}$. For the particular choice $\mathbb{U} = u\mathbb{I}$, a cheap algorithm was proposed in [6] to directly

build a matrix $\widetilde{\mathbb{Q}} = \mathbb{Q}\mathbb{Q}^{\mathrm{T}}$ with $\mathbb{Q}$ having orthonormal columns. The computation of $\widetilde{\mathbb{Q}}$ in [6] with our notation requires $3\ell_E^2 + \mathrm{O}(\ell_E)$ floating point operations (flops). The same algorithm can be efficiently applied to the present case as well.

Let m_E be the number of internal degrees of freedom for $\mathbf{F}_E$ and $m_E \neq 0$. In this case, only part of matrix $\mathbb{W}_E$ has to be computed. After permutation of columns and rows, matrices $\mathbb{M}_E$ and $\mathbb{W}_E$ may be written in a 2×2 block form:

$$\mathbb{M}_E = \begin{pmatrix} \mathbb{M}_E^{11} & \mathbb{M}_E^{12} \\ \mathbb{M}_E^{21} & \mathbb{M}_E^{22} \end{pmatrix} \quad \text{and} \quad \mathbb{W}_E = \begin{pmatrix} \mathbb{W}_E^{11} & \mathbb{W}_E^{12} \\ \mathbb{W}_E^{21} & \mathbb{W}_E^{22} \end{pmatrix},$$

with the first diagonal blocks corresponding to internal degrees of freedom. The algorithms of static condensation and subsequent hybridization require the inverse of the Schur complement $\mathbb{M}_E^{22} - \mathbb{M}_E^{21}[\mathbb{M}_E^{11}]^{-1}\mathbb{M}_E^{12}$ which is nothing but the matrix $\mathbb{W}_E^{22}$. The corresponding block of $\widetilde{\mathbb{Q}}$ can be computed with $3(\ell_E - m_E)^2 + \mathrm{O}(\ell_E)$ flops. If all faces of element E are moderately curved, $m_E = 2k_E$ and the above modification becomes essential.

In the rest of this subsection we present an alternative strategy for the explicit construction of a matrix $\mathbb{Q}$ satisfying the hypotheses of Theorem 3.3, so that no restrictions are posed on $\widetilde{\mathbb{U}}$, and the full family of MFD methods can be generated.

Proposition 3.6. *Let the matrix $\mathscr{I}$ be defined as follows*:

$$\mathscr{I} = \begin{pmatrix} \mathbb{I}_3 & & & \\ -\mathbb{I}_3 & \mathbb{I}_3 & & \\ & -\mathbb{I}_3 & \ddots & \\ & & \ddots & \mathbb{I}_3 \\ & & & -\mathbb{I}_3 \end{pmatrix} \in \mathfrak{R}^{\ell_E \times (\ell_E - 3)}.$$

Then, the matrices $\mathbb{C}$ and $\mathbb{Q}$ given by

$$\mathbb{C} = \mathscr{D}\mathscr{I} \quad \text{and} \quad \mathbb{Q} = \mathbb{C} - \frac{1}{|E|}\mathbb{D}\mathbb{K}^{-1}\mathbb{N}^{\mathrm{T}}\mathscr{I}, \tag{3.26}$$

respectively, have full column rank and satisfy $\mathbb{C}^{\mathrm{T}}\mathbb{D} = \mathbb{R}^{\mathrm{T}}\mathbb{Q} = 0$. Moreover,

$$\mathrm{cond}(\mathbb{Q}) := \frac{\sigma_{\max}(\mathbb{Q})}{\sigma_{\min}(\mathbb{Q})} \leqslant \frac{1 + 3\sqrt{\ell_E}\gamma_*^2/\tau_*}{\sin(\pi/(2k_E))},$$

where $\sigma_{\max}(\mathbb{Q})$, $\sigma_{\min}(\mathbb{Q})$ are the largest and smallest non-zero singular values of $\mathbb{Q}$, respectively.

Proof. It is obvious that $\mathbb{C}$ has full column rank and spans the null space of $\mathbb{D}^{\mathrm{T}}$. Let us show that $\mathbb{R}^{\mathrm{T}}\mathbb{Q} = 0$. Since $\mathscr{D}$ is an orthogonal matrix, we have

$$\begin{aligned} \mathbb{R}^{\mathrm{T}}\mathbb{Q} &= \mathbb{R}^{\mathrm{T}}\mathbb{C} - \mathbb{R}^{\mathrm{T}}\frac{1}{|E|}\mathbb{D}\mathbb{K}^{-1}\mathbb{N}^{\mathrm{T}}\mathscr{I} = \mathbb{R}^{\mathrm{T}}\mathbb{C} - \mathbb{N}^{\mathrm{T}}\mathscr{I} \\ &= \mathbb{N}^{\mathrm{T}}\mathscr{D}^{\mathrm{T}}\mathscr{D}\mathscr{I} - \mathbb{N}^{\mathrm{T}}\mathscr{I} = 0. \end{aligned}$$

Let us show now that $\mathbb{Q}$ has full column rank. The definition (3.26) yields

$$\mathbb{Q} = \mathscr{D}\mathscr{I} - \frac{1}{|E|}\mathscr{D}\begin{pmatrix} \mathbb{I}_3 \\ \vdots \\ \mathbb{I}_3 \end{pmatrix}\mathbb{N}^{\mathrm{T}}\mathscr{I}.$$

We use again property (3.16) and the simple fact that $[\mathbb{I}_3, \ldots, \mathbb{I}_3]\mathscr{I} = 0$ to show that

$$\mathbb{Q}^{\mathrm{T}}\mathbb{Q} = \mathscr{I}^{\mathrm{T}}\mathscr{I} + \frac{k_E}{|E|^2}\mathscr{I}^{\mathrm{T}}\mathbb{N}\mathbb{N}^{\mathrm{T}}\mathscr{I}.$$

Since the matrix $\mathscr{I}^{\mathrm{T}}\mathscr{I}$ has full rank equal to $\ell_E - 3$ and matrix $\mathscr{I}^{\mathrm{T}}\mathbb{N}\mathbb{N}^{\mathrm{T}}\mathscr{I}$ is symmetric and positive semi-definite, the matrix $\mathbb{Q}^{\mathrm{T}}\mathbb{Q}$ is symmetric and positive definite and has full rank. Therefore, the matrix $\mathbb{Q}$ has full column rank.

We next obtain bounds for the extreme singular values of $\mathbb{Q}$. Straightforward calculations show that $\mathscr{I}^{\mathrm{T}}\mathscr{I}$ is a tensor product of $\mathbb{I}_3$ and a tridiagonal matrix of size $k_E - 1$ with 2 on the main diagonal and -1 on the off diagonals. Thus, $\lambda_j(\mathbb{Q}^{\mathrm{T}}\mathbb{Q}) \geqslant \lambda_j(\mathscr{I}^{\mathrm{T}}\mathscr{I}) = 4\sin^2(j\pi/(2k_E))$ where $j = 1, \ldots, k_E - 1$. Therefore,

$$\sigma_{\min}(\mathbb{Q}) = \lambda_{\min}(\mathbb{Q}^{\mathrm{T}}\mathbb{Q})^{1/2} \geqslant 2\sin\left(\frac{\pi}{2k_E}\right).$$

Noticing that $\|\mathscr{D}\| = 1$ and $\|\mathscr{I}\| \leqslant 2$, and recalling (3.19) and (3.21), we obtain

$$\begin{aligned} \sigma_{\max}(\mathbb{Q}) = \|\mathbb{Q}\| &\leqslant \left\| I - \frac{1}{|E|}\begin{pmatrix} \mathbb{I}_3 \\ \vdots \\ \mathbb{I}_3 \end{pmatrix}\mathbb{N}^{\mathrm{T}} \right\| \|\mathscr{I}\| \\ &\leqslant 2\left(1 + \frac{1}{|E|}\sqrt{\ell_E}\|\mathbb{N}\|\right) \leqslant 2\left(1 + \sqrt{\ell_E}\frac{3\gamma_*^2}{\tau_*}\right). \end{aligned}$$

Collecting the bounds for $\sigma_{\min}(\mathbb{Q})$, $\sigma_{\max}(\mathbb{Q})$ the final result follows. □

The shape regularity constant τ_* makes usually bigger impact on the condition number $\mathrm{cond}(\mathbb{Q})$ than γ_*. For a shape-regular element E, the condition number grows as $\ell_E^{3/2}$. If $\mathrm{cond}(\mathbb{Q})$ becomes too large, the matrix $\mathbb{Q} \in \mathfrak{R}^{\ell_E \times (\ell_E - 3)}$ can be orthogonalized by the modified Gram–Schmidt process, with a computational cost of $2\ell_E(\ell_E - 3)^2$ flops [9]. This approach may be advantageous when ℓ_E is not much greater than 3.

4. Superlinear convergence

In [5], we proved the super-linear convergence of the pressure variable for the case in which the inner product matrix $\mathbb{M}_E$ is constructed as follows. For every element E, we define a lifting operator $\mathscr{R}(\mathbf{G}_E)$ with values in $(L^2(E))^3$ and the following properties:

$$\begin{aligned} &\mathscr{R}(\mathbf{G}_E)|_{\partial E} \equiv \mathbf{G}_E \quad \text{on } \partial E, \\ &\operatorname{div}\mathscr{R}(\mathbf{G}_E) \equiv (\mathscr{D}\mathscr{I}\mathscr{V}^h\mathbf{G})_E \quad \text{in } E \end{aligned} \tag{4.1}$$

for all $\mathbf{G}_E$ and

$$\mathscr{R}_E(\mathbf{G}_E^I) = \vec{G} \tag{4.2}$$

for all constant vector-valued functions $\vec{G}$ (where $\mathbf{G}^I$ is constructed using $\vec{G}$ as in (3.1)). Then, the choice

$$[\mathbf{F},\mathbf{G}]_E := \int_E \mathbb{K}^{-1}\mathscr{R}_E(\mathbf{F_E})\cdot\mathscr{R}_E(\mathbf{G}_E)\,\mathrm{d}V \tag{4.3}$$

allowed us to prove the second order convergence rate for the pressure variable. In practical computations the matrix $\mathbb{M}_E$ was constructed in a different way, essentially following Theorem 3.2 or Theorem 3.3. However the numerical evidence still showed superconvergence for the pressure. In order to obtain a theoretical justification of such numerical evidence, we adopt the following strategy: *For every matrix $\mathbb{M}_E$ given by (3.10), find a lifting operator $\mathscr{R}_E$ such that the matrix $\mathbb{M}_E$ coincides with the matrix induced by $\mathscr{R}_E$ through (4.3)*.

Let us fix a p with $6/5 \leqslant p < 2$, and for every $\mathbf{G}_E$ consider the following Stokes-like problem: find $\boldsymbol{\eta} \in (W^{1,p}(E))^3$ and $\chi \in L^p(E)$ such that

$$\begin{aligned} &-\Delta\boldsymbol{\eta} + \nabla\chi = 0 \quad \text{in } E,\\ &\operatorname{div}\boldsymbol{\eta} = \mathscr{DIV}^h\mathbf{G}_E \quad \text{in } E,\\ &\boldsymbol{\eta} = \mathbf{G}_E \quad \text{on } \partial E. \end{aligned} \tag{4.4}$$

We recall that for $p \geqslant 6/5$, in three dimensions, we have $W^{1,p}(E) \subset L^2(E)$. It is clear that the lifting operator $\widetilde{\mathscr{R}}_E$ defined by $\boldsymbol{\eta} =: \widetilde{\mathscr{R}}_E(\mathbf{G}_E)$ satisfies properties (4.1) and (4.2).

We can now consider the space X_E made by the restrictions of X^h to E, and the space $\mathbf{W}$ obtained as $\mathbf{W} := \widetilde{\mathscr{R}}_E(X_E)$. The dimension of both spaces is equal to ℓ_E. It is clear that the space $\mathbf{W}$ contains the constant vectors.

For notational convenience, we apply a change of basis in $\mathbf{W}$, putting the three constant vectors in the last three positions, and we apply the corresponding change of variables in X_E. Let $\vec{W}_1,\dots,\vec{W}_{\ell_E}$ be the new orthogonal basis in $(L_2(E))^3$, where $\vec{W}_{\ell_E-2}$, $\vec{W}_{\ell_E-1}$, and $\vec{W}_{\ell_E}$ are constant vectors in E. The change of basis in X_E results in an equivalency transformation for the matrix $\mathbb{M}_E$. We denote the transformed matrix by $\widetilde{\mathbb{M}}_E$. The matrix obtained from the lifting operator $\widetilde{\mathscr{R}}_E$ using (4.3) will be given by

$$\widetilde{\mathbb{S}}_{s,r} = \int_E \mathbb{K}^{-1}\vec{W}_s\cdot\vec{W}_r\,\mathrm{d}V.$$

Let $\mathbb{S}$ be the representation of matrix $\widetilde{\mathbb{S}}$ in the original basis of X_E.

We cannot expect that the matrix $\widetilde{\mathbb{S}}$ coincides with $\widetilde{\mathbb{M}}_E$ based on (3.10). We note however that, due to our change of basis, the last three columns and the last three rows of all possible transformed matrices $\widetilde{\mathbb{M}}_E$ obtained through (3.10) will coincide with the corresponding columns and rows of $\widetilde{\mathbb{S}}$. This is due to the fact that all inner products induced by all these matrices will be exact on constant vectors. The rigorous proof is based on Assumption **(S2)** and properties (4.1) and (4.2):

$$\begin{aligned} 0 &= \int_E \mathbb{K}^{-1}\vec{W}_r\cdot\vec{W}_s\,\mathrm{d}V = \int_E \nabla\varphi_r^1\cdot\vec{W}_s\,\mathrm{d}V\\ &= \int_{\partial E}\varphi_r^1\vec{W}_s\cdot\mathbf{n}_E\,\mathrm{d}\Sigma = [(\mathbb{K}\nabla\varphi_r^1)_E^I,(\vec{W}_s)_E^I]_E, \end{aligned}$$

where $1 \leqslant s \leqslant \ell_E$, $\ell_E - 3 < r \leqslant \ell_E$ and we denoted by φ_r^1 a linear function such that $\nabla\varphi_r^1 = \mathbb{K}^{-1}\vec{W}_r$.

Thus, the matrices $\widetilde{\mathbb{S}}$ and $\widetilde{\mathbb{M}}_E$ are block diagonal with two blocks of size $\ell_E - 3$ and 3, respectively. Moreover, in the new basis, $\operatorname{im}(\mathbb{D})$ is spanned by the last three columns of either $\mathbb{S}$ or $\widetilde{\mathbb{M}}_E$. We are going therefore to modify the first $\ell_E - 3$ elements of the basis $\vec{W}_1,\dots,\vec{W}_{\ell_E}$, and then use the new basis to construct a new lifting operator $\mathscr{R}_E$ in such a way that *the matrix obtained from it by (4.3), coincides with the matrix $\widetilde{\mathbb{M}}_E$ based on (3.10)*. This will not be feasible for *all* matrices $\widetilde{\mathbb{M}}_E$, but, as we shall see, for *many* of them.

Lemma 4.1. *The matrix $\mathbb{M}_E$ given by (3.10) is induced by an inner product (4.3) if and only if*

$$\|\widetilde{\mathbb{M}}_E^{1/2}\mathbf{v}\| \geqslant \|\widetilde{\mathbb{S}}^{1/2}\mathbf{v}\| \quad \forall\mathbf{v} \in \operatorname{im}(\widetilde{\mathbb{M}}_E - \widetilde{\mathbb{S}}).$$

Proof. Re-using the original idea from [2], we consider the space of vector-valued functions $\vec{V}$ satisfying

$$\begin{aligned} &\operatorname{div}\vec{V} = 0 \quad \text{in } E,\\ &\vec{V} = 0 \quad \text{on } \partial E,\\ &\int_E \mathbb{K}^{-1}\vec{W}_s\cdot\vec{V}\,\mathrm{d}V = 0 \quad s = 1,2,\dots,\ell_E. \end{aligned} \tag{4.5}$$

It is clear that such space is non-empty, and actually is infinite dimensional. Then we choose $\ell_E - 3$ independent elements $\vec{V}_1,\dots,\vec{V}_{\ell_E-3}$ in this space, and consider the lifting functions

$$\vec{W}_1 + \vec{V}_1, \vec{W}_2 + \vec{V}_2, \dots, \vec{W}_{\ell_E-3} + \vec{V}_{\ell_E-3}, \vec{W}_{\ell_E-2}, \vec{W}_{\ell_E-1}, \vec{W}_{\ell_E}.$$

We denote by $\mathbb{T}$ the matrix induced by (4.3). The orthogonality property gives:

$$\mathbb{T} = \widetilde{\mathbb{S}} + \mathbb{V}, \tag{4.6}$$

where $\mathbb{V}$ is the Gram matrix of functions $\vec{V}_s$ ($s = 1,2,\dots,\ell_E - 3$), completed by zeroes in the last three rows and columns. The matrix $\mathbb{V}$ is symmetric and positive semi-definite. With an abuse of notation, we shall indicate by $\mathbb{V}$ the $\ell_E \times \ell_E$ matrix $\mathbb{V}$ as well as its $(\ell_E - 3)\times(\ell_E - 3)$ principal diagonal block. We note that for any symmetric positive semi-definite $(\ell_E - 3)\times(\ell_E - 3)$ matrix $\mathbb{P}$, it is possible to find functions $\vec{V}_s$ that will generate this matrix. Indeed, if we choose $\ell_E - 3$ orthonormal vectors $\vec{V}_{0,r}$ satisfying (4.5) and a matrix $\mathbb{Z}$ such that $\mathbb{Z}^{\mathrm{T}}\mathbb{Z} = \mathbb{P}$, then taking

$$\vec{V}_s = \sum_{r=1}^{\ell_E-3}\mathbb{Z}_{s,r}\vec{V}_{0,r},$$

we easily obtain $\mathbb{V} = \mathbb{Z}\mathbb{Z}^{\mathrm{T}} = \mathbb{P}$.

Now, the assertion of the lemma can be rephrased as follows: Find necessary and sufficient conditions for the transformed matrix $\widetilde{\mathbb{M}}_E$ (defined above) to be one of the matrices $\mathbb{T}$. It follows from (4.6) that $\widetilde{\mathbb{M}}_E = \mathbb{T}$ if and only if

$$\widetilde{\mathbb{M}}_E - \widetilde{\mathbb{S}} = \mathbb{V} \geqslant 0$$

or

$$((\widetilde{\mathbb{M}}_E - \mathbb{S})\mathbf{v}, \mathbf{v}) \geqslant 0 \quad \forall \mathbf{v} \in \mathfrak{R}^{\ell_E}. \tag{4.7}$$

This proves the assertion of the lemma. □

Corollary 4.2. *The matrix* $\mathbb{M}_E$ *given by (3.10) is induced by an inner product (4.3) if and only if*

$$\|\mathbb{M}_E^{1/2}\mathbf{v}\| \geqslant \|\mathbb{S}^{1/2}\mathbf{v}\| \quad \forall \mathbf{v} \in \mathrm{im}(\mathbb{C}).$$

The proof of this corollary is based on deriving an explicit form for the equivalency transformation mentioned above. We leave it as the exercise for the reader.

When $\mathbb{C}$ has orthonormal columns and $\mathbb{U} = u\mathbb{I}$, the above lemma requires u to be sufficiently large. Indeed, since $\mathbb{C}\mathbb{C}^{\mathrm{T}}\mathbf{v} = \mathbf{v}$, we get

$$(\mathbb{M}_E\mathbf{v}, \mathbf{v}) = \left\| \frac{1}{|E|^{1/2}} \mathbb{K}^{-1/2}\mathbb{R}^{\mathrm{T}}\mathbf{v} \right\|^2 + u\|\mathbf{v}\|^2 \geqslant u\|\mathbf{v}\|^2.$$

On the other hand,

$$(\mathbb{S}\mathbf{v}, \mathbf{v}) \leqslant \lambda_{\max}(\mathbb{S})\|\mathbf{v}\|^2,$$

where $\lambda_{\max}(\mathbb{S})$ is the maximum eigenvalue of $\mathbb{S}$. Thus, it is sufficient to take u larger than $\lambda_{\max}(\mathbb{S})$ to satisfy (4.7) and hence, to guarantee superlinear convergence of the family of MFD methods. This justifies, in some sense, our statement in Section 1 concerning the vector field representation inside each element (the *guardian angel*). Indeed, at least when u is sufficiently large, we can say that our inner product is indeed based on a reconstruction of the vector variable inside each element, but we *actually do not see* how such reconstruction looks like: we only see the resulting inner product.

It is also pertinent to note that the approach based on the lifting operator $\mathscr{R}_E$ is only one of the ways to prove the superconvergence result. Therefore, in practice, the superconvergence may be observed for a wide range of parameters u.

5. Numerical experiments

We shall consider diffusion problems with sufficiently smooth solutions, so that we may expect the second order convergence rate for the scalar variable $\mathbf{p}_h$ and the first order convergence rate for the other primary variable $\mathbf{F}_h$ on generalized polyhedral meshes.

We shall measure the accuracy of the discrete solution $(\mathbf{p}_h, \mathbf{F}_h)$ in the *natural* norms induced by the scalar products (2.6) and (2.8). Let $(\mathbf{p}^I, \mathbf{F}^I)$ be the interpolated solution where $\mathbf{p}^I$ is the vector of mean values of the solution p over the elements and $\mathbf{F}^I$ is given by (3.1). We define the following discrete L_2 errors:

$$|||\mathbf{p}^I - \mathbf{p}_h||| = [\mathbf{p}^I - \mathbf{p}_h, \mathbf{p}^I - \mathbf{p}_h]_Q^{1/2},$$
$$|||\mathbf{F}^I - \mathbf{F}_h||| = [\mathbf{F}^I - \mathbf{F}_h, \mathbf{F}^I - \mathbf{F}_h]_X^{1/2}.$$

For all meshes considered in this section, we have performed the following consistency check. We have solved the Dirichlet boundary value problem with a constant tensor $\mathbb{K}$ and an exact solution p^1 given by $p^1 = x_1 + 2x_2 + 3x_3$. All non-planar mesh faces were classified as strongly curved. As p^1 is linear, it follows from Assumption (**S2**) that the errors should be zero, and this is indeed observed in our experiments.

The discrete problem (for the Lagrange multipliers) was solved with the preconditioned conjugate gradient (PCG) method. A V-cycle of the algebraic multigrid [16] was

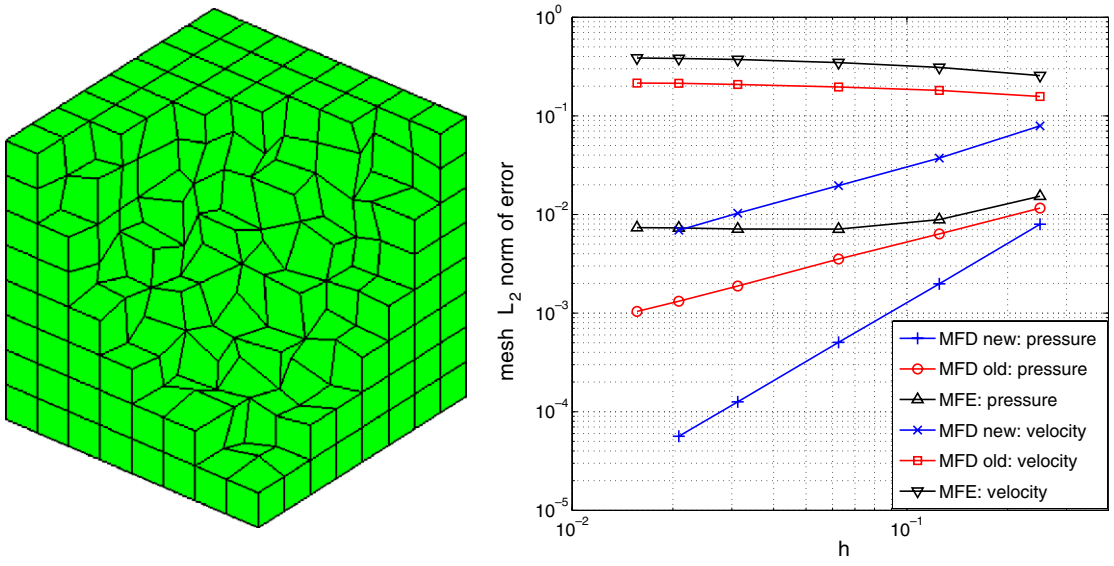

Fig. 1. A logically cubic mesh with randomly perturbed interior points (left picture) and the convergence graphs (right picture) showing optimal convergence rate for the new MFD method (blue), and the lack of convergence for the mixed finite element (black) and the old MFD (red) methods. (For interpretation of the references in colour in this figure legend, the reader is referred to the web version of this article.)

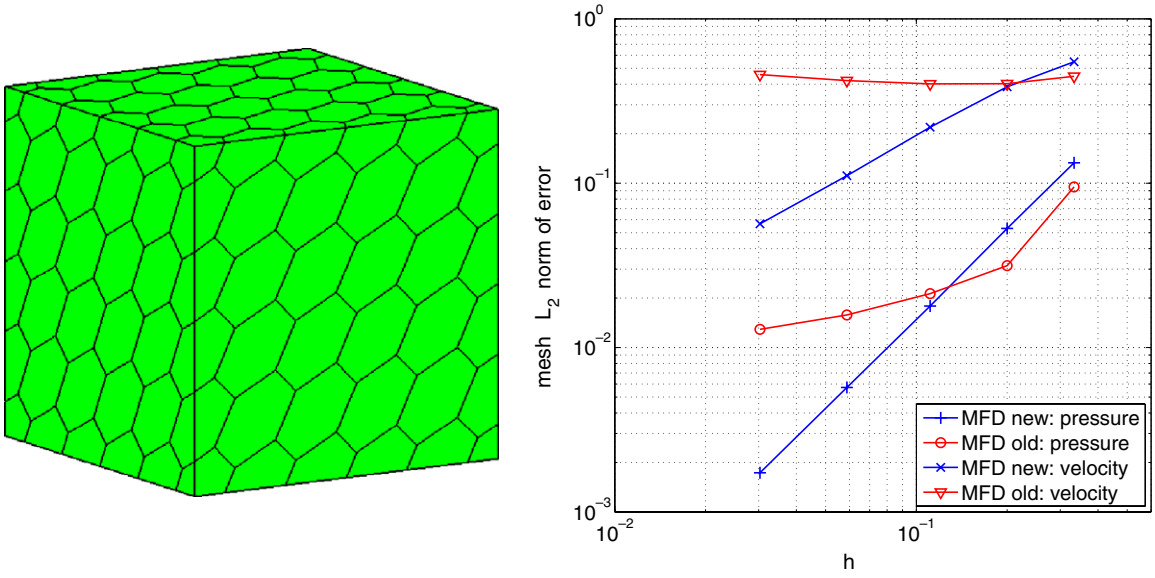

Fig. 2. The trace of the generalized polyhedral mesh (left picture) and convergence graphs (right picture) showing the optimal convergence rates for the new MFD method (blue) and lack of convergence for the old MFD method (red). (For interpretation of the references in colour in this figure legend, the reader is referred to the web version of this article.)

chosen as the preconditioner. The stopping criterion for the PCG method was a reduction of the Euclidean norm of the residual by a factor 10^{-12}. In both experiments below, mesh faces were classified on moderately and strongly curved using $\sigma_* = 0.2$.

Example 1. We consider the Dirichlet boundary value problem (1.1) in the unit cube $[0,1]^3$ with the identity tensor $\mathbb{K}$ and the exact solution

$$p(x,y,z) = x^2y^3z + 3x\sin(yz).$$

We consider a sequence of generalized hexahedral meshes as shown in Fig. 1 where a part of the unit cube was cut out to show the interior mesh. The meshes are generated by moving each mesh point P (of an originally uniform mesh with mesh step h) to a random position inside a cube $C(P)$ centered at the point. The sides of $C(P)$ are aligned with the coordinate axes and their length equals to 0.8 h.

For every element E, we define a scalar matrix $\widetilde{\mathbb{U}} = \tilde{u}_E\mathbb{I}$ where $\tilde{u}_E = \mathrm{trace}(\mathbb{K}_E)/|E|$. The convergence graphs in Fig. 1 show the optimal convergence of the new MFD method and the lack of convergence for the mixed finite element method with the lowest order Raviart–Thomas elements and the old MFD method. Recall that the last two methods use one degree of freedom per mesh face to approximate the flow field. Note that we have the first order convergence rate for the velocity variable and the second order convergence rate for the pressure variable.

Example 2. Let us consider the Dirichlet boundary described in the previous example on a different sequence of generalized polyhedral meshes (see Fig. 2 where we show only the mesh trace). It is pertinent to note that 68% of interior mesh faces are strongly curved according to definition (**M1**) with $\sigma_* = 0.2$.

The mixed finite element method cannot be used on such meshes. The old MFD method lacks convergence for both primary variables. For the new MFD method, we have again the first order convergence rate for the velocity variable and the second order convergence rate for the pressure variable.

6. Conclusion

We gave a rigorous mathematical description of a family of mimetic finite difference methods for diffusion problems on generalized polyhedral meshes. We developed an inexpensive and easy-to-implement numerical algorithm, analyzed it both theoretically and numerically, and proved the superconvergence result for the scalar variable. With this new method, discretizations of elliptic equations on generalized polyhedral meshes becomes as simple as on tetrahedral meshes. The results were obtained for the full material tensor.

Acknowledgements

The work was partly performed at the Los Alamos National Laboratory operated by the University of California for the US Department of Energy under contract W-7405-ENG-36. The first author acknowledges the partial support of the PRIN-2004 Program of Italian MIUR. The second and the third authors acknowledge the partial support of the DOE/ASCR Program in the Applied Mathematical Sciences. The authors thank Dr. V. Dyadechko (LANL) for his help in generating the meshes for Example 2.

References

[1] D.N. Arnold, F. Brezzi, Mixed and non-conforming finite element methods: implementation, post-processing and error estimates, Math. Modell. Numer. Anal. 19 (1985) 7–35.

[2] C. Baiocchi, F. Brezzi, L. Franca, Virtual bubbles and Galerkin-least-squares type methods (Ga.L.S.), Comput. Methods Appl. Mech. Engrg. 105 (1993) 125–144.

[3] F. Brezzi, M. Fortin, Mixed and Hybrid Finite Element Methods, Springer-Verlag, New York, 1991.

[4] F. Brezzi, K. Lipnikov, M. Shashkov, Convergence of mimetic finite difference method for diffusion problems on polyhedral meshes, SIAM J. Numer. Anal. 43 (5) (2005) 1872–1896.

[5] F. Brezzi, K. Lipnikov, M. Shashkov, Convergence of mimetic finite difference method for diffusion problems on polyhedral meshes with curved faces, Math. Mod. Methods Appl. Sci. 16 (2) (2006) 275–297.

[6] F. Brezzi, K. Lipnikov, V. Simoncini, A family of mimetic finite difference methods on polygonal and polyhedral meshes, Math. Mod. Methods Appl. Sci. 15 (10) (2005) 1533–1552.

[7] J. Campbell, M. Shashkov, A tensor artificial viscosity using a mimetic finite difference algorithm, J. Comput. Phys. 172 (2001) 739–765.

[8] B. Fraeijs de Veubeke, Displacement and equilibrium models in the finite element method, in: O.C. Zienkiewicz, G. Holister (Eds.), Stress Analysis, John Wiley and Sons, New York, 1965.

[9] G. Golub, C. Van Loan, Matrix Computations, The John Hopkins University Press, Baltimore and London, 1989.

[10] J. Hyman, M. Shashkov, The approximation of boundary conditions for mimetic finite difference methods, Comput. Math. Appl. 36 (1998) 79–99.

[11] J. Hyman, M. Shashkov, Mimetic discretizations for Maxwell's equations and the equations of magnetic diffusion, Prog. Electromagnetic Res. 32 (2001) 89–121.

[12] J. Hyman, M. Shashkov, S. Steinberg, The numerical solution of diffusion problems in strongly heterogeneous non-isotropic materials, J. Comput. Phys. 132 (1997) 130–148.

[13] J. Nocedal, S. Wright, Numerical Optimization, Springer, Heidelberg, Berlin, New York, 1999.

[14] M. Peric, S. Ferguson, The advantage of polyhedral meshes. Technical report, CD Adapco Group, 2005. www.cd-adapco.com/news/24/TetsvPoly.htm.

[15] M. Shashkov, Conservative Finite-Difference Methods on General Grids, CRC Press, Boca Raton, 1996.

[16] K. Stüben, Algebraic multigrid (AMG): experiences and comparisons, Appl. Math. Comput. 13 (1983) 419–452.