ESAIM: M2AN
Vol. 40, N° 2, 2006, pp. 367–391
DOI: 10.1051/m2an:2006013

ESAIM: Mathematical Modelling and Numerical Analysis
www.edpsciences.org/m2an

EQUIVALENCE BETWEEN LOWEST-ORDER MIXED FINITE ELEMENT AND MULTI-POINT FINITE VOLUME METHODS ON SIMPLICIAL MESHES*

Martin Vohralík[1, 2]

Abstract. We consider the lowest-order Raviart–Thomas mixed finite element method for second-order elliptic problems on simplicial meshes in two and three space dimensions. This method produces saddle-point problems for scalar and flux unknowns. We show how to easily and locally eliminate the flux unknowns, which implies the equivalence between this method and a particular multi-point finite volume scheme, without any approximate numerical integration. The matrix of the final linear system is sparse, positive definite for a large class of problems, but in general nonsymmetric. We next show that these ideas also apply to mixed and upwind-mixed finite element discretizations of nonlinear parabolic convection–diffusion–reaction problems. Besides the theoretical relationship between the two methods, the results allow for important computational savings in the mixed finite element method, which we finally illustrate on a set of numerical experiments.

Mathematics Subject Classification. 76M10, 76M12, 76S05.

Received: April 21, 2005. Revised: December 16, 2005.

1. Introduction

Let us consider the elliptic problem

$$\mathbf{u} = -\mathbf{S}\nabla p \quad \text{in } \Omega, \tag{1.1a}$$

$$\nabla \cdot \mathbf{u} = q \quad \text{in } \Omega, \tag{1.1b}$$

$$p = p_D \quad \text{on } \Gamma_D, \quad \mathbf{u}\cdot\mathbf{n} = u_N \quad \text{on } \Gamma_N, \tag{1.1c}$$

where $\Omega \subset \mathbb{R}^d$, $d = 2, 3$, is a polygonal (polyhedral) domain (open, bounded, and connected set), $\mathbf{S}$ is a bounded, symmetric (this is however not necessary), and uniformly positive definite tensor, $p_D \in \mathrm{H}^{\frac{1}{2}}(\Gamma_D)$, $u_N \in \mathrm{H}^{-\frac{1}{2}}(\Gamma_N)$, $q \in \mathrm{L}^2(\Omega)$, $\Gamma_D \cap \Gamma_N = \emptyset$, $\overline{\Gamma_D} \cup \overline{\Gamma_N} = \partial\Omega$, and $|\Gamma_D| \neq 0$, where $|\Gamma_D|$ is the measure of the set Γ_D.

Keywords and phrases. Mixed finite element method, saddle-point problem, finite volume method, second-order elliptic equation, nonlinear parabolic convection–diffusion–reaction equation.

* *This work was supported by the GdR MoMaS, CNRS-2439, ANDRA, BRGM, CEA, EdF, France.*

[1] Department of Mathematics, Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague, Trojanova 13, 12000 Prague 2, Czech Republic. vohralik@km1.fjfi.cvut.cz

[2] Laboratoire de Mathématiques, Analyse Numérique et EDP, Université de Paris-Sud, Bât. 425, 91405 Orsay, France. martin.vohralik@math.u-psud.fr

Let $\mathcal{T}_h$ be a simplicial triangulation of Ω (consisting of triangles if $d=2$ and of tetrahedra if $d=3$) such that each boundary side (edge if $d=2$, face if $d=3$) lies entirely either in Γ_D or in Γ_N. Let us denote by $\mathcal{E}_h$ the set of all non-Neumann sides of $\mathcal{T}_h$. Let finally $\tilde{\mathbf{u}} \in \mathbf{H}(\mathrm{div},\Omega)$ be such that $\tilde{\mathbf{u}}\cdot\mathbf{n} = u_N$ on Γ_N in the appropriate sense. The approximation of the problem (1.1a)–(1.1c) by means of the mixed finite element method consists in finding $\mathbf{u}_h = \mathbf{u}_{0,h} + \tilde{\mathbf{u}}$, $\mathbf{u}_{0,h} \in \mathbf{V}(\mathcal{E}_h)$, and $p_h \in \Phi(\mathcal{T}_h)$ such that (see [10, 33])

$$(\mathbf{S}^{-1}\mathbf{u}_{0,h}, \mathbf{v}_h)_\Omega - (\nabla\cdot\mathbf{v}_h, p_h)_\Omega = -\langle \mathbf{v}_h\cdot\mathbf{n}, p_D\rangle_{\partial\Omega} - (\mathbf{S}^{-1}\tilde{\mathbf{u}}, \mathbf{v}_h)_\Omega \qquad \forall\, \mathbf{v}_h \in \mathbf{V}(\mathcal{E}_h), \tag{1.2a}$$

$$-(\nabla\cdot\mathbf{u}_{0,h}, \phi_h)_\Omega = -(q,\phi_h)_\Omega + (\nabla\cdot\tilde{\mathbf{u}},\phi_h)_\Omega \qquad \forall \phi_h \in \Phi(\mathcal{T}_h), \tag{1.2b}$$

where $(\mathbf{u}_h,\mathbf{v}_h)_\Omega = \int_\Omega \mathbf{u}_h\cdot\mathbf{v}_h\,\mathrm{d}\mathbf{x}$, $\langle \mathbf{v}_h\cdot\mathbf{n},\varphi\rangle_{\partial\Omega} = \int_{\partial\Omega}\mathbf{v}_h\cdot\mathbf{n}\,\varphi\,\mathrm{d}\gamma(\mathbf{x})$, and $\mathbf{V}(\mathcal{E}_h)$ and $\Phi(\mathcal{T}_h)$ are suitable finite-dimensional spaces defined on $\mathcal{T}_h$. The associated matrix problem is of the saddle-point type and can be written in the form

$$\begin{pmatrix} \mathbb{A} & \mathbb{B}^t \\ \mathbb{B} & 0 \end{pmatrix}\begin{pmatrix} U \\ P \end{pmatrix} = \begin{pmatrix} F \\ G \end{pmatrix}. \tag{1.3}$$

In the lowest-order Raviart–Thomas method [32] and its three-dimensional Nédélec variant [30] the scalar unknowns P are associated with the elements of $\mathcal{T}_h$ and U are the fluxes through the sides of $\mathcal{E}_h$. Using the hybridization technique, one can decrease the number of unknowns to the Lagrange multipliers associated with non-Dirichlet sides and obtain a symmetric and positive definite matrix, *cf.* [8, 10]. In fact, the hybridization is very close to the piecewise linear nonconforming finite element method, *cf.* [8, 14]. The fluxes can then be recovered using the technique first proposed in [29]. Especially in three space dimensions, there are much fewer elements than sides, and hence the long-standing interest in reducing the unknowns to only the scalar unknowns P. This is indeed possible, using approximate numerical integration, see [34] for rectangles and $\mathbf{S}$ diagonal and [4, 9] for rectangles and triangles and $\mathbf{S}$ diagonal and for a limited class of tetrahedra and $\mathbf{S} = Id$. Using the expanded mixed finite element method, these techniques can be extended also onto full-matrix diffusion tensors $\mathbf{S}$ for rectangular parallelepipeds [6] and for "smooth" coefficients and meshes consisting of triangles, quadrilaterals, and hexahedra [7]. To our knowledge, the only technique for reducing the number of unknowns to the number of elements without any numerical integration is proposed and studied in [13, 39, 40]. In two space dimensions, it works on unstructured triangular meshes, but in three space dimensions, it only works on a limited class of structured tetrahedral meshes. Here one associates a *new* unknown to each element.

We present in Section 2 of this paper a new method which permits to exactly and efficiently reduce the system (1.3) to a system for the *original* scalar unknowns P only. We show that, under a condition of the invertibility of some local matrices associated with vertices and only depending on the mesh and on the diffusion tensor, one can express the flux through a given side using the scalar unknowns, sources, and possibly boundary conditions associated with the elements sharing one of the vertices of this side. Recall that expressing the flux through a given side using the scalar unknowns in neighboring elements is the principle of multi-point finite volume schemes, *cf.* [1–3, 15, 20, 21, 26]. Hence the lowest-order Raviart–Thomas mixed finite element method is in the given case equivalent to a particular multi-point finite volume scheme, and this without any numerical integration. We call this scheme a *condensed mixed finite element scheme*. We then discuss the modifications of the proposed scheme if the local matrices are not invertible, consisting namely in considering different sets of elements for the expression of the flux through a given side.

The condensation of the lowest-order Raviart–Thomas method leads to linear systems with sparse but in general nonsymmetric matrices, as we show in Section 3. The system matrix is positive definite under a condition on the mesh and on the tensor $\mathbf{S}$, which can be reduced to a shape criterion allowing for fairly general elements if $\mathbf{S}$ is piecewise constant and scalar. For example, one can deform a square $(0,1)\times(0,1)$, discretized by regular right-angled triangles, until the triangle elements contain angles greater than 130 degrees, see Example 3.8 below. The fulfillment of this condition in particular implies the invertibility of the local matrices from the previous paragraph. Finally, in Section 4, we apply the proposed condensation to mixed (*cf.* [5, 18]) and upwind-mixed (*cf.* [16, 17, 25]) finite element discretizations of nonlinear parabolic convection–diffusion–reaction problems.

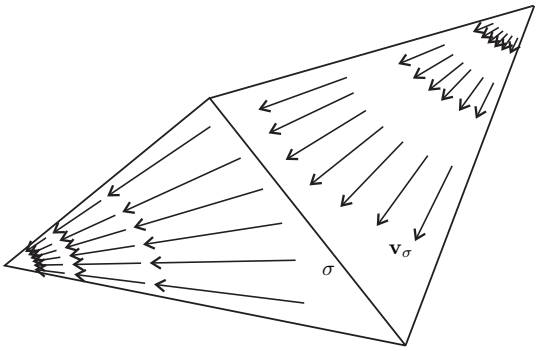


FIGURE 1. RTN basis function $\mathbf{v}_\sigma$ associated with the edge σ.

The essential idea of what we propose can be formulated as follows: given a second-order problem, first decompose it into scalar and flux unknowns and guarantee the fulfillment of the inf-sup (Babuška–Brezzi) condition. Then eliminate the added fluxes. One can in this way obtain the precision of the mixed finite element method for the computational cost of the finite volume one. This is confirmed by numerical experiments carried out in Section 5. Especially for nonlinear parabolic convection–diffusion–reaction problems, one can considerably reduce the CPU time of standard mixed solution approaches. We refer to a more detailed discussion in Section 5.4. Finally, the proposed condensation can easily be implemented in a new self-standing code or in existing mixed finite element codes. This paper is a detailed description of the results previously announced in [37] and [38]. Extension to higher-order schemes is an ongoing work.

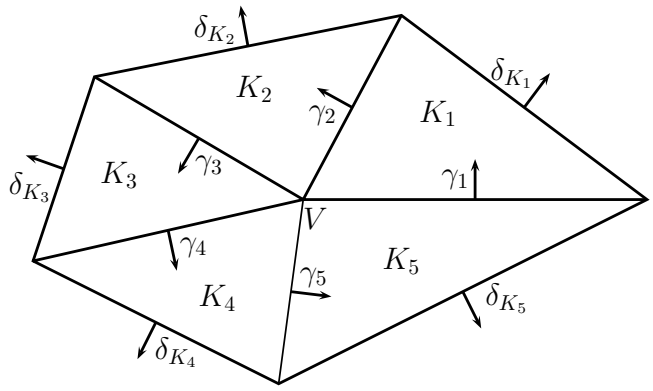
2. The equivalence

We first define the spaces $\mathbf{V}(\mathcal{E}_h)$ and $\Phi(\mathcal{T}_h)$ in this section. We then establish the equivalence between the lowest-order mixed finite element and a particular multi-point finite volume method.

Let us consider simplices $K, L \in \mathcal{T}_h$ sharing an interior side σ. Let V_K be the vertex of K opposite to σ and V_L the vertex of L opposite to σ. Then the RTN (Raviart–Thomas–Nédélec) basis function $\mathbf{v}_\sigma \in \mathbf{V}(\mathcal{E}_h)$ associated with the side σ can be written in the form $\mathbf{v}_\sigma(\mathbf{x}) = \frac{1}{d|K|}(\mathbf{x} - V_K)$, $\mathbf{x} \in K$, $\mathbf{v}_\sigma(\mathbf{x}) = \frac{1}{d|L|}(V_L - \mathbf{x})$, $\mathbf{x} \in L$, $\mathbf{v}_\sigma(\mathbf{x}) = 0$ otherwise. We refer to Figure 1 for a schematic visualization of a RTN basis function in two space dimensions. We fix the orientation of $\mathbf{v}_\sigma$, *i.e.* the order of K and L. For a Dirichlet boundary side σ, the support of $\mathbf{v}_\sigma$ only consists of $K \in \mathcal{T}_h$ such that $\sigma \in \mathcal{E}_K$, where $\mathcal{E}_K$ stands for the sides of the element K. A basis function $\phi_K \in \Phi(\mathcal{T}_h)$ associated with an element $K \in \mathcal{T}_h$ is equal to 1 on K and to 0 otherwise.

Let us denote by $\mathcal{V}_h$ the set of all vertices and consider $V \in \mathcal{V}_h$. We call the set of all elements of $\mathcal{T}_h$ sharing this vertex a *cluster* associated with V and denote it by $\mathcal{C}_V$. Let us denote by $\mathcal{E}_{\mathcal{C}_V}$ the set of all non-Neumann sides of $\mathcal{C}_V$, by $\mathcal{F}_{\mathcal{C}_V}$ the set of all non-Neumann sides sharing V, and by $\mathcal{G}_{\mathcal{C}_V}$ the set of other non-Neumann sides of $\mathcal{C}_V$. Let finally $\mathcal{C}_V^{\mathrm{el}}$ denote the set of elements from the cluster that contain exactly one side from $\mathcal{G}_{\mathcal{C}_V}$. We denote by δ_K the side from $\mathcal{E}_K \cap \mathcal{G}_{\mathcal{C}_V}$ for $K \in \mathcal{C}_V^{\mathrm{el}}$. We have $\mathcal{E}_{\mathcal{C}_V} = \mathcal{F}_{\mathcal{C}_V} \cup \mathcal{G}_{\mathcal{C}_V}$, $\mathcal{F}_{\mathcal{C}_V} \cap \mathcal{G}_{\mathcal{C}_V} = \emptyset$, and $|\mathcal{C}_V^{\mathrm{el}}| = |\mathcal{G}_{\mathcal{C}_V}|$, where we denote by $|A|$ the cardinality of a set A. An example of a cluster $\mathcal{C}_V$ lying in the interior of the domain Ω is given in Figure 2. In this case, $\mathcal{F}_{\mathcal{C}_V}$ are simply the sides sharing V, $\mathcal{G}_{\mathcal{C}_V}$ the other sides of $\mathcal{C}_V$, and $\mathcal{C}_V^{\mathrm{el}} = \mathcal{C}_V$. The situation is more delicate near the boundary, especially if there are Neumann boundary conditions, *cf.* Figure 3. This is also the reason for the quite complex notation introduced. The basic principle of the condensation will however be clear from Figure 2. Finally, we are not interested in the particular and trivial cases where $\mathcal{F}_{\mathcal{C}_V} = \emptyset$ or $\mathcal{G}_{\mathcal{C}_V} = \emptyset$.

Our aim is to easily and locally express $\mathbf{u}_{0,h}$ with the aid of p_h, or, equivalently, the fluxes U with the aid of the scalar unknowns P. For this purpose, we consider equations (1.2a) for the basis functions $\mathbf{v}_\gamma$, $\gamma \in \mathcal{F}_{\mathcal{C}_V}$. We remark that the support of all $\mathbf{v}_\gamma$, $\gamma \in \mathcal{F}_{\mathcal{C}_V}$, is included in $\mathcal{C}_V$ and that $\mathbf{u}_{0,h}|_{\mathcal{C}_V} = \sum_{\sigma \in \mathcal{E}_{\mathcal{C}_V}} U_\sigma \mathbf{v}_\sigma$. This yields,



$$\begin{aligned}\mathcal{C}_V &= \{K_i\}_{i=1}^5\\ \mathcal{C}_V^{\mathrm{el}} &= \{K_i\}_{i=1}^5\\ \mathcal{F}_{\mathcal{C}_V} &= \{\gamma_i\}_{i=1}^5\\ \mathcal{G}_{\mathcal{C}_V} &= \{\delta_{K_i}\}_{i=1}^5\end{aligned}$$

FIGURE 2. Example of a cluster $\mathcal{C}_V$ in the interior of Ω.

using also that $p_h|_K = P_K$,

$$\sum_{\sigma\in\mathcal{E}_{\mathcal{C}_V}} U_\sigma(\mathbf{v}_\sigma, \mathbf{S}^{-1}\mathbf{v}_\gamma)_{\mathcal{C}_V} - \sum_{K\in\mathcal{C}_V} P_K(\nabla\cdot\mathbf{v}_\gamma, 1)_K = -\langle \mathbf{v}_\gamma\cdot\mathbf{n}, p_D\rangle_{\partial\Omega} - (\mathbf{S}^{-1}\tilde{\mathbf{u}}, \mathbf{v}_\gamma)_{\mathcal{C}_V} \qquad \forall\gamma\in\mathcal{F}_{\mathcal{C}_V}, \tag{2.1}$$

i.e. $|\mathcal{F}_{\mathcal{C}_V}|$ equations for the $|\mathcal{E}_{\mathcal{C}_V}|$ unknown fluxes U_σ, $\sigma\in\mathcal{E}_{\mathcal{C}_V}$, where we consider the scalar unknowns P_K, $K\in\mathcal{C}_V$, as parameters. Note that in practice, $p_D|_\sigma \approx \langle p_D, 1\rangle_\sigma/|\sigma|$, $\sigma\subset\Gamma_D$, and $\tilde{\mathbf{u}} \approx \sum_{\sigma\subset\Gamma_N}\langle u_N, 1\rangle_\sigma \mathbf{v}_\sigma$ (for $u_N\in L^2(\Gamma_N)$), so that the above system is completely discrete. The remaining $|\mathcal{G}_{\mathcal{C}_V}|$ equations are given by (1.2b) for all ϕ_K, $K\in\mathcal{C}_V^{\mathrm{el}}$,

$$-\sum_{\sigma\in\mathcal{E}_K,\sigma\not\subset\Gamma_N} U_\sigma(\nabla\cdot\mathbf{v}_\sigma, 1)_K = -(q,1)_K + (\nabla\cdot\tilde{\mathbf{u}}, 1)_K \qquad \forall K\in\mathcal{C}_V^{\mathrm{el}}. \tag{2.2}$$

The matrix problem associated with the set of equations (2.1)–(2.2) can be written in the form

$$\begin{pmatrix} \mathbb{A}_{1,V} & \mathbb{A}_{2,V}\\ \mathbb{B}_{1,V} & \mathbb{B}_{2,V}\end{pmatrix}\begin{pmatrix} U_V^{\mathcal{F}}\\ U_V^{\mathcal{G}}\end{pmatrix} = \begin{pmatrix} F_V - \mathbb{B}_V^t P_V\\ G_V\end{pmatrix}, \tag{2.3}$$

where $U_V^{\mathcal{F}} = \{U_\sigma\}_{\sigma\in\mathcal{F}_{\mathcal{C}_V}}$, $U_V^{\mathcal{G}} = \{U_\sigma\}_{\sigma\in\mathcal{G}_{\mathcal{C}_V}}$, and $P_V = \{P_K\}_{K\in\mathcal{C}_V}$.

We now notice that the matrix $\mathbb{B}_{2,V}$ is square, diagonal, and its entries are equal to ±1 (this follows from the fact that each $K\in\mathcal{C}_V^{\mathrm{el}}$ contains exactly one side from $\mathcal{G}_{\mathcal{C}_V}$ and using that $(\nabla\cdot\mathbf{v}_\sigma, 1)_K = \pm1$ for $\sigma\in\mathcal{E}_K$). Hence we can eliminate the $U_V^{\mathcal{G}}$ unknowns and come to

$$\mathbb{M}_V U_V^{\mathcal{F}} = F_V - \mathbb{B}_V^t P_V - \mathbb{A}_{2,V}\mathbb{B}_{2,V}^{-1} G_V \tag{2.4}$$

for each vertex $V\in\mathcal{V}_h$. Let us call the matrix

$$\mathbb{M}_V := \mathbb{A}_{1,V} - \mathbb{A}_{2,V}\mathbb{B}_{2,V}^{-1}\mathbb{B}_{1,V} \tag{2.5}$$

a *local condensation matrix* associated with the vertex V. Note that as soon as $\mathbb{M}_V$ are invertible for all $V\in\mathcal{V}_h$, we can eliminate the fluxes U from (1.3) and by the well-posedness of (1.3) arrive at a well-posed system for the scalar unknowns P only. We now summarize the obtained results in the following theorem:

Theorem 2.1 (equivalence between MFEM and a particular multi-point FVM)**.** *Let the matrices $\mathbb{M}_V$ given by (2.5) be invertible for all $V \in \mathcal{V}_h$. Then the lowest-order Raviart–Thomas mixed finite element method on simplicial meshes is equivalent to a multi-point finite volume scheme, where the flux through each side can be expressed using the scalar unknowns, sources, and possibly boundary conditions associated with the elements sharing one of the vertices of this side.*

Remark 2.2 (comparison with classical multi-point FVMs)**.** In "classical" multi-point finite volume schemes, *cf.* [1–3, 15, 20, 21, 26], one attempts to express the flux through a given side only using the scalar unknowns associated with the neighboring elements. There are two essential differences between these classical multi-point finite volume schemes and a particular multi-point finite volume scheme—the mixed finite element method. First, in the mixed finite element method, not only the scalar unknowns, but also the sources and possibly boundary conditions associated with the neighboring elements are used to express the flux through a given side. Second, to obtain this expression, one has to solve a local linear problem. In this last feature, the condensed mixed finite element scheme is similar to the "multi-point flux-approximation" scheme proposed and tested in [1, 2] (*cf.* also [26]), see the next remark.

Remark 2.3 (comparison with multi-point flux-approximation schemes)**.** In the multi-point flux-approximation schemes [1, 2, 26, 27], one constructs a polygonal subdomain around each vertex (called an "interaction region"), joining *e.g.* the edge midpoints through triangle barycentres in two space dimensions. Then a piecewise linear nonconforming approximation $\overline{p}_h$ of p in the interaction region is supposed and a full flux continuity for $\overline{\mathbf{u}}_h := -\mathbf{S}\nabla\overline{p}_h$ across the semi-edges sharing the given vertex is imposed, which leads to a local linear system. Solving this system gives the flux through a given semi-edge only using the scalar unknowns associated with the elements sharing the given vertex.

A priori, there are several differences from the condensed mixed finite element scheme. First, the condensed scheme works with the Raviart–Thomas $\mathbf{u}_h$, which is a piecewise linear vector function *a priori* independent of the piecewise constant p_h. Next, the local problems are associated with the whole cluster associated with the given vertex and not only with the interaction region. The conceptual difference however seems to be that the condensed scheme imposes the whole equilibrium (1.1a)–(1.1c), whence the boundary conditions and sources in the flux expressions, whereas the multi-point flux-approximation scheme is only a way how to discretize the relation (1.1a). This may lead to higher precision of the condensed scheme for problems with sources and sinks, but it seems at the same time responsible for the oscillations in mixed finite element discretizations of parabolic problems, see Section 4 for the extension of the condensed scheme to this case.

Some numerical comparisons of the "O" multi-point flux-approximation scheme, of the lowest-order Raviart–Thomas mixed finite element method, and of the piecewise linear continuous Galerkin finite element method on triangular meshes can be found in [27], see also [26]. Relations among several different schemes on quadrilateral grids are also studied in [28].

Let $V \in \mathcal{V}_h$. Let us define a mapping $\Psi_V : \mathbb{R}^{|\mathcal{F}_{\mathcal{C}_V}|} \to \mathbb{R}^{|\mathcal{E}_h|}$, extending a vector $U_V^{\mathcal{F}} = \{U_\sigma\}_{\sigma\in\mathcal{F}_{\mathcal{C}_V}}$ of values associated with the sides from $\mathcal{F}_{\mathcal{C}_V}$ to a vector of values associated with all non-Neumann sides $\mathcal{E}_h$ by

$$[\Psi_V(U_V^{\mathcal{F}})]_\sigma := \begin{cases} U_\sigma & \text{if } \sigma \in \mathcal{F}_{\mathcal{C}_V} \\ 0 & \text{if } \sigma \notin \mathcal{F}_{\mathcal{C}_V} \end{cases} .$$

Since there is no possibility of confusion, we keep the same notation also for a mapping $\mathbb{R}^{|\mathcal{F}_{\mathcal{C}_V}|\times|\mathcal{F}_{\mathcal{C}_V}|} \to \mathbb{R}^{|\mathcal{E}_h|\times|\mathcal{E}_h|}$, extending a local matrix $\mathbb{M}_V$ to a full-size one by zeros by

$$[\Psi_V(\mathbb{M}_V)]_{\sigma,\gamma} := \begin{cases} (\mathbb{M}_V)_{\sigma,\gamma} & \text{if } \sigma \in \mathcal{F}_{\mathcal{C}_V} \text{ and } \gamma \in \mathcal{F}_{\mathcal{C}_V} \\ 0 & \text{if } \sigma \notin \mathcal{F}_{\mathcal{C}_V} \text{ or } \gamma \notin \mathcal{F}_{\mathcal{C}_V} \end{cases} .$$

We finally in the same fashion define a mapping $\Theta_V : \mathbb{R}^{|\mathcal{F}_{\mathcal{C}_V}|\times|\mathcal{C}_V^{\mathrm{el}}|} \to \mathbb{R}^{|\mathcal{E}_h|\times|\mathcal{T}_h|}$, filling a full-size representation of a matrix $\mathbb{J}_V$ by zeros on the rows associated with the sides that are not from $\mathcal{F}_{\mathcal{C}_V}$ and on the columns

associated with the elements that are not from $\mathcal{C}_V^{\mathrm{el}}$,

$$[\Theta_V(\mathbb{J}_V)]_{\sigma,K} := \begin{cases} (\mathbb{J}_V)_{\sigma,K} & \text{if } \sigma \in \mathcal{F}_{\mathcal{C}_V} \text{ and } K \in \mathcal{C}_V^{\mathrm{el}} \\ 0 & \text{if } \sigma \notin \mathcal{F}_{\mathcal{C}_V} \text{ or } K \notin \mathcal{C}_V^{\mathrm{el}} \end{cases}.$$

Let the local condensation matrices $\mathbb{M}_V$ be invertible for all $V \in \mathcal{V}_h$. Let us define $\mathbb{J}_V$ by $\mathbb{J}_V := \mathbb{M}_V^{-1}\mathbb{A}_{2,V}\mathbb{B}_{2,V}^{-1}$. We then can rewrite (2.4) as

$$\Psi_V(U_V^{\mathcal{F}}) = \Psi_V(\mathbb{M}_V^{-1})(F - \mathbb{B}^t P) - \Theta_V(\mathbb{J}_V)G. \tag{2.6}$$

We now notice that

$$\sum_{V \in \mathcal{V}_h} \frac{1}{d} \Psi_V(U_V^{\mathcal{F}}) = U, \tag{2.7}$$

which expresses that if we go through all $V \in \mathcal{V}_h$ and observe the sides in the sets $\mathcal{F}_{\mathcal{C}_V}$, each $\sigma \in \mathcal{E}_h$ appears just d-times (each side has d vertices). Hence we can sum (2.6) over all vertices and divide it by d to find that

$$U = \tilde{\mathbb{A}}^{-1}(F - \mathbb{B}^t P) - \mathbb{J}G, \tag{2.8}$$

where

$$\tilde{\mathbb{A}}^{-1} := \frac{1}{d} \sum_{V \in \mathcal{V}_h} \Psi_V(\mathbb{M}_V^{-1}), \quad \mathbb{J} := \frac{1}{d} \sum_{V \in \mathcal{V}_h} \Theta_V(\mathbb{J}_V). \tag{2.9}$$

Finally, inserting this expression into the second equation of (1.3), we obtain a system for only the scalar unknowns

$$-\mathbb{B}\tilde{\mathbb{A}}^{-1}\mathbb{B}^t P = G - \mathbb{B}\tilde{\mathbb{A}}^{-1}F + \mathbb{B}\mathbb{J}G. \tag{2.10}$$

We now give two remarks.

Remark 2.4 (comparison with the direct elimination of the fluxes)**.** From (1.3), $U = \mathbb{A}^{-1}(F - \mathbb{B}^t P)$. There are two essential differences in comparison with (2.8). First, the matrix $\tilde{\mathbb{A}}^{-1}$ is sparse, whereas $\mathbb{A}^{-1}$ tends to be full. Second, $\tilde{\mathbb{A}}^{-1}$ is obtained for the computational cost of the inverse of $|\mathcal{V}_h|$ local matrices, whereas obtaining $\mathbb{A}^{-1}$ is in general very expensive.

Remark 2.5 (implementation into existing mixed finite element codes)**.** The local problems (2.3) correspond to the rows of (1.3) associated with the sides from $\mathcal{F}_{\mathcal{C}_V}$ and elements from $\mathcal{C}_V^{\mathrm{el}}$. Hence obtaining the final problem (2.10) from (1.3) is immediate.

It appears that in some particular cases, the matrix $\mathbb{M}_V$ is not invertible, *cf.* Example 3.10 below. We give sufficient conditions on the mesh $\mathcal{T}_h$ and on the diffusion tensor $\mathbf{S}$ ensuring that $\mathbb{M}_V$ are invertible for all $V \in \mathcal{V}_h$ below as byproducts of Lemmas 3.6 and 3.9. Finally, we discuss in Section 3.3 the approaches as to how to modify the proposed technique in order to overcome this difficulty.

3. Properties of the condensed mixed finite element scheme

We study in this section the properties of the system matrix of the condensed mixed finite element scheme which are important from the computational point of view, namely its sparsity pattern, symmetry, and positive definiteness. It shows that all these properties are closely related to the properties of the local condensation matrices, which we shall study hereafter. We finally discuss variants and extensions of the proposed technique and open questions.

3.1. **Properties of the system matrix**

Theorem 3.1 (stencil of the system matrix)**.** *Let $\mathbb{M}_V$ be invertible for all $V \in \mathcal{V}_h$. Then on a row of the final system matrix $\mathbb{B}\tilde{\mathbb{A}}^{-1}\mathbb{B}^t$ corresponding to an element $K \in \mathcal{T}_h$, the only possible nonzero entries are on columns corresponding to $L \in \mathcal{T}_h$ such that K and L share a common vertex.*

Proof. The assertion of this theorem follows from the fact that by (2.4), the flux through a side σ is expressed only using the scalar unknowns of the elements $K \in \mathcal{T}_h$ such that K and σ share a common vertex. □

Theorem 3.2 (positive definiteness of the system matrix)**.** *Let $\mathbb{M}_V$ be positive definite for all $V \in \mathcal{V}_h$. Then the final system matrix $\mathbb{B}\tilde{\mathbb{A}}^{-1}\mathbb{B}^t$ is also positive definite.*

Proof. Since $\mathbb{B}$ has a full row rank, $\mathbb{B}\tilde{\mathbb{A}}^{-1}\mathbb{B}^t$ is positive definite as soon as $\tilde{\mathbb{A}}^{-1}$ is positive definite, *i.e.* when

$$X^t\tilde{\mathbb{A}}^{-1}X > 0 \quad \text{for all } X \in \mathbb{R}^{|\mathcal{E}_h|},\ X \neq 0.$$

Let $V \in \mathcal{V}_h$. We define a mapping $\Pi_V : \mathbb{R}^{|\mathcal{E}_h|} \to \mathbb{R}^{|\mathcal{F}_{\mathcal{C}_V}|}$, restricting a vector of values associated with all non-Neumann sides to a vector of values associated with the sides from $\mathcal{F}_{\mathcal{C}_V}$. Let $X \in \mathbb{R}^{|\mathcal{E}_h|}$, $X \neq 0$. Then

$$X^t\tilde{\mathbb{A}}^{-1}X = \frac{1}{d}\sum_{V\in\mathcal{V}_h} X^t\Psi_V(\mathbb{M}_V^{-1})X = \frac{1}{d}\sum_{V\in\mathcal{V}_h} [\Pi_V(X)]^t\mathbb{M}_V^{-1}\Pi_V(X) > 0,$$

using the positive definiteness of the local condensation matrices $\mathbb{M}_V$ and consequently of $\mathbb{M}_V^{-1}$ for all $V \in \mathcal{V}_h$ and the fact that in the above sum, all the terms are non-negative and at least d of them are positive. □

Theorem 3.3 (symmetry of the system matrix)**.** *Let $\mathbb{M}_V$ be invertible and symmetric for all $V \in \mathcal{V}_h$. Then the final system matrix $\mathbb{B}\tilde{\mathbb{A}}^{-1}\mathbb{B}^t$ is also symmetric.*

Proof. If $\mathbb{M}_V$ and consequently $\mathbb{M}_V^{-1}$ are symmetric for all $V \in \mathcal{V}_h$, their extensions $\Psi_V(\mathbb{M}_V^{-1})$ are symmetric as well. Hence $\tilde{\mathbb{A}}^{-1}$, a sum of symmetric matrices by (2.9), is symmetric. Finally, if $\tilde{\mathbb{A}}^{-1}$ is symmetric, $\mathbb{B}\tilde{\mathbb{A}}^{-1}\mathbb{B}^t$ is symmetric as well. □

3.2. **Properties of the local condensation matrices**

The local condensation matrix $\mathbb{M}_V$ (2.5) for $V \in \mathcal{V}_h$ is given by the equations (2.1)–(2.2). It does not depend on the right-hand side, and hence it is connected with the following problem: find $\mathbf{u} \in \mathbf{V}(\mathcal{E}_{\mathcal{C}_V})$ such that

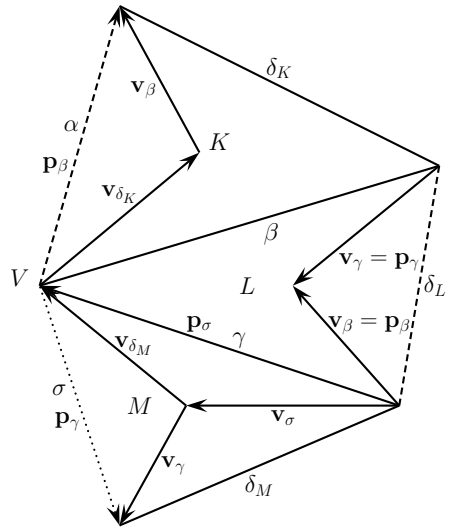
$$(\mathbf{u}, \mathbf{S}^{-1}\mathbf{v})_{\mathcal{C}_V} = 0 \qquad \forall \mathbf{v} \in \mathbf{V}(\mathcal{F}_{\mathcal{C}_V}), \tag{3.1a}$$

$$(\nabla\cdot\mathbf{u}, \phi_K)_K = 0 \qquad \forall K \in \mathcal{C}_V^{\mathrm{el}}. \tag{3.1b}$$

Here, $\mathbf{V}(\mathcal{E}_{\mathcal{C}_V})$ is the space spanned by the RTN basis functions $\mathbf{v}_\sigma$ associated with the non-Neumann sides $\mathcal{E}_{\mathcal{C}_V}$ of the cluster $\mathcal{C}_V$ and $\mathbf{V}(\mathcal{F}_{\mathcal{C}_V})$ is its restriction with the basis functions $\mathbf{v}_\sigma$ associated with the sides from $\mathcal{F}_{\mathcal{C}_V}$. The problem (3.1a)–(3.1b) is further equivalent to the following Petrov–Galerkin problem: find $\mathbf{u} \in \mathbf{V}(\mathrm{div}, \mathcal{E}_{\mathcal{C}_V})$ such that

$$(\mathbf{u}, \mathbf{S}^{-1}\mathbf{v})_{\mathcal{C}_V} = 0 \qquad \forall \mathbf{v} \in \mathbf{V}(\mathcal{F}_{\mathcal{C}_V}),$$

where $\mathbf{V}(\mathrm{div}, \mathcal{E}_{\mathcal{C}_V})$ is the subspace of $\mathbf{V}(\mathcal{E}_{\mathcal{C}_V})$ of the functions whose divergence is equal to 0 on all elements $K \in \mathcal{C}_V^{\mathrm{el}}$. The space $\mathbf{V}(\mathrm{div}, \mathcal{E}_{\mathcal{C}_V})$ is spanned by basis functions $\mathbf{p}_\sigma$ associated with the sides from $\mathcal{F}_{\mathcal{C}_V}$, which have the same support as the RTN basis functions $\mathbf{v}_\sigma$ and whose fluxes through the associated sides equal to those of $\mathbf{v}_\sigma$ (this in particular fixes their orientation). Namely, for $K \in \mathcal{C}_V^{\mathrm{el}}$ and $\sigma \in \mathcal{E}_K \cap \mathcal{F}_{\mathcal{C}_V}$, $\mathbf{p}_\sigma|_K = \mathbf{v}_\sigma - \frac{(\nabla\cdot\mathbf{v}_\sigma, 1)_K}{(\nabla\cdot\mathbf{v}_{\delta_K}, 1)_K}\mathbf{v}_{\delta_K}$. Note that this is a constant function given by $\frac{1}{d|K|}\mathbf{q}_\sigma|_K$, where $\mathbf{q}_\sigma|_K$ is the vector of the edge of K that is not included in the sides σ and δ_K. For $K \in \mathcal{C}_V \setminus \mathcal{C}_V^{\mathrm{el}}$, $\mathbf{p}_\sigma|_K = \mathbf{v}_\sigma|_K$. We refer to Figure 3 for a schematic visualization for $d = 2$.



$$
\begin{aligned}
\mathcal{C}_V &= \{K, L, M\}\\
\mathcal{C}_V^{\mathrm{el}} &= \{K, M\}\\
\mathcal{F}_{\mathcal{C}_V} &= \{\beta, \gamma, \sigma\}\\
\mathcal{G}_{\mathcal{C}_V} &= \{\delta_K, \delta_M\}\\
\Gamma_N &\supset \alpha, \delta_L\\
\Gamma_D &\supset \sigma\\
\mathbf{V}(\mathcal{F}_{\mathcal{C}_V}) &= \mathrm{span}\{\mathbf{v}_\beta, \mathbf{v}_\gamma, \mathbf{v}_\sigma\}\\
\mathbf{V}(\mathrm{div}, \mathcal{E}_{\mathcal{C}_V}) &= \mathrm{span}\{\mathbf{p}_\beta, \mathbf{p}_\gamma, \mathbf{p}_\sigma\}
\end{aligned}
$$

FIGURE 3. Example of a boundary cluster $\mathcal{C}_V$ and schematic representation of the basis functions of the spaces $\mathbf{V}(\mathcal{F}_{\mathcal{C}_V})$ and $\mathbf{V}(\mathrm{div}, \mathcal{E}_{\mathcal{C}_V})$.

Lemma 3.4 (form of the local condensation matrices)**.** *The local condensation matrix $\mathbb{M}_V$ for $V \in \mathcal{V}_h$ can be written in the form*

$$(\mathbb{M}_V)_{\gamma,\sigma} = (\mathbf{p}_\sigma, \mathbf{S}^{-1}\mathbf{v}_\gamma)_{\mathcal{C}_V},$$

where $\mathbf{p}_\sigma$ and $\mathbf{v}_\sigma$, $\sigma \in \mathcal{F}_{\mathcal{C}_V}$, are the basis functions of the spaces $\mathbf{V}(\mathrm{div}, \mathcal{E}_{\mathcal{C}_V})$ and $\mathbf{V}(\mathcal{F}_{\mathcal{C}_V})$, respectively, defined above.

Proof. We can rewrite (3.1a)–(3.1b) as

$$\sum_{\sigma \in \mathcal{E}_{\mathcal{C}_V}} U_\sigma (\mathbf{v}_\sigma, \mathbf{S}^{-1}\mathbf{v}_\gamma)_{\mathcal{C}_V} = 0 \qquad \forall \gamma \in \mathcal{F}_{\mathcal{C}_V}, \tag{3.2a}$$

$$\sum_{\sigma \in \mathcal{E}_K, \sigma \not\subset \Gamma_N} U_\sigma (\nabla \cdot \mathbf{v}_\sigma, 1)_K = 0 \qquad \forall K \in \mathcal{C}_V^{\mathrm{el}}, \tag{3.2b}$$

where $\mathbf{u} = \displaystyle\sum_{\sigma \in \mathcal{E}_{\mathcal{C}_V}} U_\sigma \mathbf{v}_\sigma$. Expressing U_{δ_K} from (3.2b) gives

$$U_{\delta_K} = \frac{- \displaystyle\sum_{\sigma \in \mathcal{E}_K \cap \mathcal{F}_{\mathcal{C}_V}} U_\sigma (\nabla \cdot \mathbf{v}_\sigma, 1)_K}{(\nabla \cdot \mathbf{v}_{\delta_K}, 1)_K}.$$

Inserting this into (3.2a), we have

$$
\begin{aligned}
\sum_{\sigma \in \mathcal{E}_{\mathcal{C}_V}} U_\sigma (\mathbf{v}_\sigma, \mathbf{S}^{-1}\mathbf{v}_\gamma)_{\mathcal{C}_V} &= \sum_{K \in \mathrm{supp}(\mathbf{v}_\gamma)} \left\{ \sum_{\sigma \in \mathcal{E}_K \cap \mathcal{F}_{\mathcal{C}_V}} U_\sigma (\mathbf{v}_\sigma, \mathbf{S}^{-1}\mathbf{v}_\gamma)_K + U_{\delta_K} (\mathbf{v}_{\delta_K}, \mathbf{S}^{-1}\mathbf{v}_\gamma)_K \right\}\\
&= \sum_{K \in \mathrm{supp}(\mathbf{v}_\gamma)} \sum_{\sigma \in \mathcal{E}_K \cap \mathcal{F}_{\mathcal{C}_V}} U_\sigma \Big(\mathbf{v}_\sigma - \frac{(\nabla \cdot \mathbf{v}_\sigma, 1)_K}{(\nabla \cdot \mathbf{v}_{\delta_K}, 1)_K} \mathbf{v}_{\delta_K}, \mathbf{S}^{-1}\mathbf{v}_\gamma \Big)_K,
\end{aligned}
$$

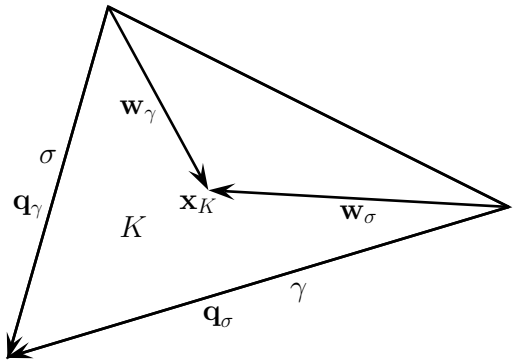


FIGURE 4. Triangle K for the simplified elementwise positive definiteness criterion.

where we have defined for simplification $\mathbf{v}_{\delta_K} = 0$ if $K \in \mathcal{C}_V \setminus \mathcal{C}_V^{\mathrm{el}}$ (*i.e.* if $\mathcal{E}_K \cap \mathcal{G}_{\mathcal{C}_V} = \emptyset$). Hence, using the definition of the basis functions of the spaces $\mathbf{V}(\mathrm{div}, \mathcal{E}_{\mathcal{C}_V})$ and $\mathbf{V}(\mathcal{F}_{\mathcal{C}_V})$, the assertion of the lemma follows. □

Remark 3.5 (implementation)**.** Let $\mathbf{S}$ be piecewise constant and let $\Gamma_N = \emptyset$. Then

$$\begin{aligned}(\mathbb{M}_V)_{\gamma,\sigma} &= \sum_{K\in\mathcal{C}_V;\,\sigma,\gamma\in\mathcal{E}_K} (\mathbf{p}_\sigma, \mathbf{S}^{-1}\mathbf{v}_\gamma)_K = \sum_{K\in\mathcal{C}_V;\,\sigma,\gamma\in\mathcal{E}_K} \frac{(\nabla\cdot\mathbf{v}_\gamma, 1)_K}{d^2|K|^2}(\mathbf{S}^{-1}\mathbf{q}_\sigma, \mathbf{x} - V_{\gamma,K})_K \\ &= \sum_{K\in\mathcal{C}_V;\,\sigma,\gamma\in\mathcal{E}_K} \frac{1}{d^2|K|}\mathbf{S}|_K^{-1}\mathbf{q}_\sigma|_K \cdot \mathbf{w}_\gamma|_K,\end{aligned}$$

where $\sigma, \gamma \in \mathcal{F}_{\mathcal{C}_V}$ and $\mathbf{w}_\gamma|_K := (\nabla\cdot\mathbf{v}_\gamma, 1)_K(\mathbf{x}_K - V_{\gamma,K})$ with $\mathbf{x}_K$ the barycentre of K and $V_{\gamma,K}$ the vertex of K opposite to the side γ, *cf.* Figure 4. We have used the facts that $\{K \in \mathcal{C}_V;\, \sigma, \gamma \in \mathcal{E}_K\} = \mathrm{supp}(\mathbf{p}_\sigma) \cap \mathrm{supp}(\mathbf{v}_\gamma)$ and that $\mathbf{x}_K = (\mathbf{x}, 1)_K/|K|$. Hence, to implement the condensed mixed finite element scheme when in addition $q = 0$, everything we need are the edge and vertex–barycentre vectors in each simplex and its measure.

We now give two lemmas that guarantee the positive definiteness of the local condensation matrices, the assumption of Theorem 3.2. Since positive definiteness implies invertibility, the local condensation matrices are under the following conditions in particular invertible, which guarantees the feasibility of the condensation in the proposed form. The given conditions are sufficient but not necessary to ensure the positive definiteness—they can be used as a simple elementwise or sidewise criterion, in order to avoid the direct checking of the positive definiteness of the local condensation matrices.

Lemma 3.6 (positive definiteness of the local condensation matrices—elementwise criterion)**.** *Let the matrices* $\mathbb{E}_{V,K} \in \mathbb{R}^{|\mathcal{E}_K\cap\mathcal{F}_{\mathcal{C}_V}|\times|\mathcal{E}_K\cap\mathcal{F}_{\mathcal{C}_V}|}$ *given by*

$$(\mathbb{E}_{V,K})_{\gamma,\sigma} := (\mathbf{p}_\sigma, \mathbf{S}^{-1}\mathbf{v}_\gamma)_K,$$

where $\mathbf{p}_\sigma$ *and* $\mathbf{v}_\sigma$, $\sigma \in \mathcal{E}_K \cap \mathcal{F}_{\mathcal{C}_V}$, *are the basis functions of the spaces* $\mathbf{V}(\mathrm{div}, \mathcal{E}_{\mathcal{C}_V})$ *and* $\mathbf{V}(\mathcal{F}_{\mathcal{C}_V})$, *respectively, be positive definite for all* $K \in \mathcal{T}_h$ *and for all vertices* V *of* K. *Then the local condensation matrices* $\mathbb{M}_V$ *are positive definite for all* $V \in \mathcal{V}_h$.

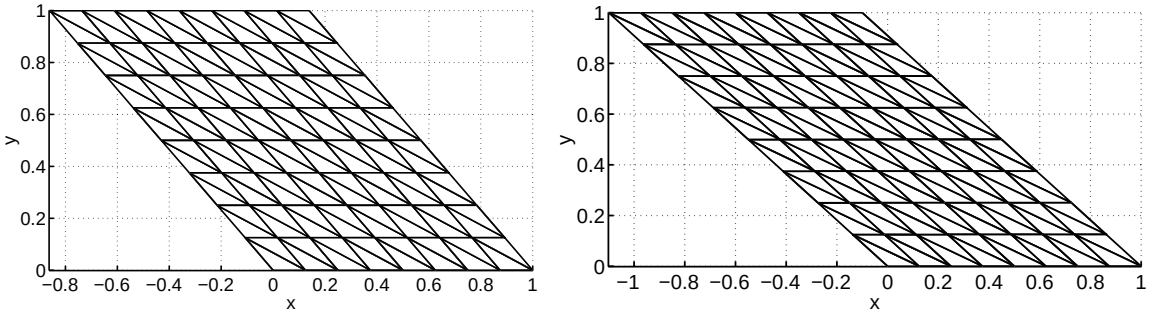


FIGURE 5. Theoretical (left) and experimental (right) limit mesh for the positive definiteness of the system matrix for a deformed square and $\mathbf{S} = Id$.

Proof. Let $V \in \mathcal{V}_h$ and let $X \in \mathbb{R}^{|\mathcal{F}_{\mathcal{C}_V}|}$, $X \neq 0$. We then have, with $\mathbf{p} = \sum_{\sigma \in \mathcal{F}_{\mathcal{C}_V}} X_\sigma \mathbf{p}_\sigma$, $\mathbf{v} = \sum_{\sigma \in \mathcal{F}_{\mathcal{C}_V}} X_\sigma \mathbf{v}_\sigma$,

$$\begin{aligned} X^t \mathbb{M}_V X &= (\mathbf{p}, \mathbf{S}^{-1}\mathbf{v})_{\mathcal{C}_V} = \sum_{K \in \mathcal{C}_V^{\mathrm{el}}} (\mathbf{p}, \mathbf{S}^{-1}\mathbf{v})_K + \sum_{K \in \mathcal{C}_V \setminus \mathcal{C}_V^{\mathrm{el}}} (\mathbf{p}, \mathbf{S}^{-1}\mathbf{v})_K \\ &= \sum_{K \in \mathcal{C}_V^{\mathrm{el}}} [\Pi_{V,K}(X)]^t \mathbb{E}_{V,K} \Pi_{V,K}(X) + \sum_{K \in \mathcal{C}_V \setminus \mathcal{C}_V^{\mathrm{el}}} (\mathbf{v}, \mathbf{S}^{-1}\mathbf{v})_K > 0, \end{aligned}$$

where the mapping $\Pi_{V,K} : \mathbb{R}^{|\mathcal{F}_{\mathcal{C}_V}|} \to \mathbb{R}^{|\mathcal{E}_K \cap \mathcal{F}_{\mathcal{C}_V}|}$ restricts a vector of values associated with the sides from $\mathcal{F}_{\mathcal{C}_V}$ to a vector of values associated with the sides from $\mathcal{E}_K \cap \mathcal{F}_{\mathcal{C}_V}$, and using the fact that the two last terms are non-negative and at least one of them is positive. □

Remark 3.7 (simple elementwise positive definiteness criterion in two space dimensions)**.** Let $d = 2$ and let $\mathbf{S}$ be piecewise constant. Let $\mathbf{q}_\sigma, \mathbf{q}_\gamma, \mathbf{w}_\sigma, \mathbf{w}_\gamma$ be the constant edge and vertex–barycentre vectors of a triangle K as in Figure 4. Then a simplified criterion for the positive definiteness of the local condensation matrices is

$$\left| \mathbf{S}|_K^{-1} \mathbf{q}_\sigma \cdot \mathbf{w}_\gamma + \mathbf{S}|_K^{-1} \mathbf{q}_\gamma \cdot \mathbf{w}_\sigma \right|^2 < 4 (\mathbf{S}|_K^{-1} \mathbf{q}_\sigma \cdot \mathbf{w}_\sigma)(\mathbf{S}|_K^{-1} \mathbf{q}_\gamma \cdot \mathbf{w}_\gamma) \tag{3.3}$$

for all $K \in \mathcal{T}_h$ and for all denotation σ, γ of two edges of K. Notice that $\mathbf{q}_\sigma \cdot \mathbf{w}_\gamma = 0$ for an equilateral triangle and that this quantity grows in the absolute value while deforming the triangle. On the contrary, $\mathbf{q}_\sigma \cdot \mathbf{w}_\sigma$ decreases with the angle between $\mathbf{q}_\sigma$ and $\mathbf{w}_\sigma$ and it is positive only if this angle is less than $\pi/2$. This criterion is a consequence of Remark 3.5 and of Lemma 3.6 with a tighten up criterion for triangles with Neumann edges.

Example 3.8 (positive definiteness for a triangulation of a deformed square)**.** Let $\mathbf{S} = Id$, let Ω be a square $(0,1) \times (0,1)$, and let $\mathcal{T}_h$ be its triangulation by regular right-angled triangles. Let us deform the domain and the mesh by shifting horizontally the upper edge of the square. Criterion (3.3) gives that the local condensation matrices (and consequently the system matrix) are positive definite up to the mesh given in Figure 5 on the left-hand side. The experimental limit mesh is still a bit less restrictive and is given in Figure 5 on the right-hand side.

Lemma 3.9 (positive definiteness of the local condensation matrices—sidewise criterion)**.** *Let* $\mathbf{S}$ *be piecewise constant and let* $\Gamma_N = \emptyset$*. Let for all* $\gamma \in \mathcal{E}_h$ *and for all vertices* V *of* γ*,*

$$\sum_{K \in \mathrm{supp}(\mathbf{v}_\gamma)} \frac{1}{d^2|K|} \mathbf{S}|_K^{-1} \mathbf{q}_\gamma|_K \cdot \mathbf{w}_\gamma|_K > \sum_{K \in \mathrm{supp}(\mathbf{v}_\gamma)} \frac{1}{d^2|K|} \sum_{\sigma \in \mathcal{E}_K \cap \mathcal{F}_{\mathcal{C}_V},\, \sigma \neq \gamma} \left| \frac{1}{2} \mathbf{S}|_K^{-1} (\mathbf{q}_\sigma|_K \cdot \mathbf{w}_\gamma|_K + \mathbf{q}_\gamma|_K \cdot \mathbf{w}_\sigma|_K) \right|,$$

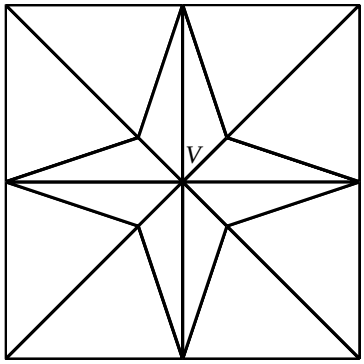


FIGURE 6. A mesh where the local condensation matrix $\mathbb{M}_V$ is singular.

where the constant edge and vertex–barycentre vectors $\mathbf{q}_\sigma|_K$, $\mathbf{w}_\sigma|_K$, *respectively, are derived from the basis functions of the spaces* $\mathbf{V}(\mathrm{div}, \mathcal{E}_{\mathcal{C}_V})$ *and* $\mathbf{V}(\mathcal{F}_{\mathcal{C}_V})$ $\mathbf{p}_\sigma$ *and* $\mathbf{v}_\sigma$ *as in Remark 3.5. Then the local condensation matrices* $\mathbb{M}_V$ *are positive definite for all* $V \in \mathcal{V}_h$.

Proof. The assumption of the lemma ensures that the matrices $\frac{1}{2}(\mathbb{M}_V + \mathbb{M}_V^t)$ for all $V \in \mathcal{V}_h$ have positive diagonal entries and are strictly diagonally dominant and hence they are positive definite. To conclude, it suffices to note that the matrix $\mathbb{M}_V$ is positive definite if and only if its symmetric part $\frac{1}{2}(\mathbb{M}_V + \mathbb{M}_V^t)$ is positive definite. □

Example 3.10 (singular local condensation matrix)**.** We give in Figure 6 an example of a mesh where the local condensation matrix $\mathbb{M}_V$ is singular for $\mathbf{S} = Id$. All the triangles sharing the vertex V have exactly one edge σ such that $\mathbf{q}_\sigma \cdot \mathbf{w}_\sigma = 0$ with the notation of Figure 4. Hence, in particular, the assumptions of Lemma 3.6 are not verified. This singularity is not local—it suffices to modify the coordinates of one point to make $\mathbb{M}_V$ invertible, *cf.* the detailed numerical study in Section 5.3.

We now state under which conditions the assumption of Theorem 3.3 is satisfied.

Lemma 3.11 (symmetry of the local condensation matrices)**.** *Let* $\mathcal{T}_h$ *consist of equilateral simplices and let* $\mathbf{S}$ *be a piecewise constant scalar function. Then* $\mathbb{M}_V$ *are symmetric for all* $V \in \mathcal{V}_h$.

Proof. We have

$$(\mathbb{M}_V)_{\gamma,\sigma} = \Big(\mathbf{v}_\sigma - \frac{(\nabla \cdot \mathbf{v}_\sigma, 1)_K}{(\nabla \cdot \mathbf{v}_{\delta_K}, 1)_K}\mathbf{v}_{\delta_K}, \mathbf{S}^{-1}\mathbf{v}_\gamma\Big)_K,$$

where $K \in \mathrm{supp}(\mathbf{p}_\sigma) \cap \mathrm{supp}(\mathbf{v}_\gamma)$, $\sigma, \gamma \in \mathcal{F}_{\mathcal{C}_V}$, $\sigma \neq \gamma$. If $K \in \mathcal{C}_V \setminus \mathcal{C}_V^{\mathrm{el}}$ and thus $\mathbf{v}_{\delta_K} = 0$ by the definition, $(\mathbb{M}_V)_{\gamma,\sigma}$ is clearly equal to $(\mathbb{M}_V)_{\sigma,\gamma}$ for a general $\mathbf{S}$ by its symmetry. If $K \in \mathcal{C}_V^{\mathrm{el}}$, $(\mathbb{M}_V)_{\gamma,\sigma} = (\mathbb{M}_V)_{\sigma,\gamma}$ as soon as

$$\big(\mathbf{S}^{-1}\mathbf{v}_{\delta_K}, \mathbf{v}_\gamma(\nabla \cdot \mathbf{v}_\gamma, 1)_K\big)_K = \big(\mathbf{S}^{-1}\mathbf{v}_{\delta_K}, \mathbf{v}_\sigma(\nabla \cdot \mathbf{v}_\sigma, 1)_K\big)_K,$$

which is the case of an equilateral simplex and $\mathbf{S}$ a piecewise constant scalar function. □

Remark 3.12 (equilateral simplices and a piecewise constant scalar diffusion tensor)**.** Let $\mathcal{T}_h$ consist of equilateral simplices, let $\mathbf{S}$ be a piecewise constant scalar function, and let $\Gamma_N = \emptyset$. Then it follows from Remark 3.5 that $(\mathbb{M}_V)_{\gamma,\sigma} = 0$ if $\sigma \neq \gamma$ (since the vectors $\mathbf{q}_\sigma|_K$ and $\mathbf{w}_\gamma|_K$ are orthogonal), and hence the local condensation matrices are diagonal. Thus to express the flux though an interior side γ in this case, we only need the scalar unknowns associated with the two elements that share this side. As a consequence, the final system matrix has only a 4-point stencil in two space dimensions and a 5-point stencil in three space dimensions and is moreover symmetric and positive definite. A simple computation shows that this matrix is equivalent to that of the

standard finite volume scheme [19] when $\mathbf{S} = Id$. Note however that the right-hand side is generally different in the presence of a source term!

3.3. **Variants, extensions, and open problems**

The essential idea of the proposed elimination, briefly said, consists in considering such sets of elements that it is possible to eliminate the fluxes through the exterior sides of these sets by the divergence equations on the exterior elements. The clusters defined by all elements sharing a given vertex represent just the most basic possibility. We now precise on this point.

Let $\mathcal{C}$ be a set of elements of $\mathcal{T}_h$ and let $\mathcal{G}_\mathcal{C}$ be the set of sides of $\mathcal{C}$ between an element $K \in \mathcal{C}$ and $L \notin \mathcal{C}$. Let each $K \in \mathcal{C}$ contain at most one $\sigma \in \mathcal{G}_\mathcal{C}$ and let us denote the subset of $\mathcal{C}$ of elements containing a $\sigma \in \mathcal{G}_\mathcal{C}$ by $\mathcal{C}^{\mathrm{el}}$. Clearly, $|\mathcal{C}^{\mathrm{el}}| = |\mathcal{G}_\mathcal{C}|$, and we denote by δ_K the side of $K \in \mathcal{C}^{\mathrm{el}}$ such that $\delta_K \in \mathcal{G}_\mathcal{C}$. Finally, let $\mathcal{E}_\mathcal{C}$ stand for all non-Neumann sides of $\mathcal{C}$ and $\mathcal{F}_\mathcal{C}$ for $\mathcal{E}_\mathcal{C} \setminus \mathcal{G}_\mathcal{C}$. A particular example is the cluster $\mathcal{C}_V$ associated with a vertex V. We have the spaces $\mathbf{V}(\mathcal{F}_{\mathcal{C}_V})$ and $\mathbf{V}(\mathrm{div}, \mathcal{E}_{\mathcal{C}_V})$ as in Section 3.2 and the following generalization of Lemma 3.6:

Lemma 3.13 (positive definiteness of local condensation matrices on general clusters)**.** *Let the matrices* $\mathbb{E}_{\mathcal{C},K} \in \mathbb{R}^{|\mathcal{E}_K \cap \mathcal{F}_\mathcal{C}| \times |\mathcal{E}_K \cap \mathcal{F}_\mathcal{C}|}$ *given by*

$$(\mathbb{E}_{\mathcal{C},K})_{\gamma,\sigma} := (\mathbf{p}_\sigma, \mathbf{S}^{-1}\mathbf{v}_\gamma)_K,$$

where $\mathbf{p}_\sigma$ *and* $\mathbf{v}_\sigma$, $\sigma \in \mathcal{E}_K \cap \mathcal{F}_\mathcal{C}$, *are the basis functions of the spaces* $\mathbf{V}(\mathrm{div}, \mathcal{E}_\mathcal{C})$ *and* $\mathbf{V}(\mathcal{F}_\mathcal{C})$, *respectively, be positive definite for all* $K \in \mathcal{C}^{\mathrm{el}}$. *Then the local condensation matrix* $\mathbb{M}_\mathcal{C}$ *associated with the cluster* $\mathcal{C}$, $(\mathbb{M}_\mathcal{C})_{\gamma,\sigma} = (\mathbf{p}_\sigma, \mathbf{S}^{-1}\mathbf{v}_\gamma)_\mathcal{C}$, *is positive definite.*

Proof. Let $X \in \mathbb{R}^{|\mathcal{F}_\mathcal{C}|}$, $X \neq 0$. We then have, with $\mathbf{p} = \sum_{\sigma \in \mathcal{F}_\mathcal{C}} X_\sigma \mathbf{p}_\sigma$, $\mathbf{v} = \sum_{\sigma \in \mathcal{F}_\mathcal{C}} X_\sigma \mathbf{v}_\sigma$,

$$\begin{aligned} X^t \mathbb{M}_\mathcal{C} X = (\mathbf{p}, \mathbf{S}^{-1}\mathbf{v})_\mathcal{C} &= \sum_{K \in \mathcal{C}^{\mathrm{el}}} (\mathbf{p}, \mathbf{S}^{-1}\mathbf{v})_K + \sum_{K \in \mathcal{C} \setminus \mathcal{C}^{\mathrm{el}}} (\mathbf{p}, \mathbf{S}^{-1}\mathbf{v})_K \\ &= \sum_{K \in \mathcal{C}^{\mathrm{el}}} [\Pi_{\mathcal{C},K}(X)]^t \mathbb{E}_{\mathcal{C},K} \Pi_{\mathcal{C},K}(X) + \sum_{K \in \mathcal{C} \setminus \mathcal{C}^{\mathrm{el}}} (\mathbf{v}, \mathbf{S}^{-1}\mathbf{v})_K > 0, \end{aligned}$$

where the mapping $\Pi_{\mathcal{C},K} : \mathbb{R}^{|\mathcal{F}_\mathcal{C}|} \to \mathbb{R}^{|\mathcal{E}_K \cap \mathcal{F}_\mathcal{C}|}$ restricts a vector of values associated with the sides from $\mathcal{F}_\mathcal{C}$ to a vector of values associated with the sides from $\mathcal{E}_K \cap \mathcal{F}_\mathcal{C}$, and using the fact that the two last terms are non-negative and at least one of them is positive. □

The above lemma shows that the positive definiteness of local condensation matrices only depends on the elements from $\mathcal{C}^{\mathrm{el}}$. Hence, in particular, should the local condensation matrix associated with a cluster of a vertex V be singular, we can resort to a wider cluster. This namely functions in the case of Example 3.10, *cf.* Section 5.3 below. Finally, to expose the problem in its full complexity, it appears that it is not necessary to consider the divergence equations on the elements of $\mathcal{C}$ sharing a side with an element $L \notin \mathcal{C}$. Let again $\mathcal{C}$ be a set of elements of $\mathcal{T}_h$ and let $\mathcal{G}_\mathcal{C}$ be the set of sides of $\mathcal{C}$ between an element $K \in \mathcal{C}$ and $L \notin \mathcal{C}$. Let $\mathcal{E}_\mathcal{C}$ stand for all non-Neumann sides of $\mathcal{C}$ and $\mathcal{F}_\mathcal{C}$ for $\mathcal{E}_\mathcal{C} \setminus \mathcal{G}_\mathcal{C}$. We notice that on the rows of the submatrix $\mathbb{A}$ of (1.3) associated with the sides from $\mathcal{F}_\mathcal{C}$ and on the rows of the submatrix $\mathbb{B}$ associated with the elements from $\mathcal{C}$, the only nonzero entries are on the columns associated with the sides from $\mathcal{E}_\mathcal{C}$. Hence, to carry out the condensation, it is sufficient if the submatrix consisting of the above rows has a rank equal to $|\mathcal{E}_\mathcal{C}|$. The main open problem, which resembles the existence of "singular triangles" in [13, 40], is whether there always has to exist a system of clusters covering $\mathcal{T}_h$ with the above property. Next, in the case of clusters associated with vertices, we have the simple expression (2.7) for the fluxes through all non-Neumann sides. For general clusters, however, we have to associate a weight α_σ^i to each side $\sigma \in \mathcal{E}_h$ and i-th out of b clusters $\mathcal{C}$ where σ belongs to $\mathcal{F}_\mathcal{C}$, such that $\sum_{i=1}^b \alpha_\sigma^i = 1$, in order to have $\sum_{i=1}^b \alpha_\sigma^i U_\sigma^i = U_\sigma$, where U_σ^i is the expression of the flux through σ from the i-th cluster. Another interesting open problem is whether one could influence the stencil, symmetry,

and positive definiteness of the system matrix by a suitable choice of these weights. For the moment, we have only focused on the basic case. Throughout all the tests presented in Sections 5.1 and 5.2 below, which involve general meshes and inhomogeneous and anisotropic (nonconstant full-matrix) diffusion tensors, we have used the local condensation matrices associated with vertices. These were always invertible, although not always positive definite.

In the lowest-order Raviart–Thomas mixed finite element method on rectangular meshes or in the lowest-order Brezzi–Douglas–Marini mixed finite element method [11,12] on simplicial meshes, it is either not possible to create subsets $\mathcal{C}$ of $\mathcal{T}_h$ such that each element of $\mathcal{C}$ shares at most one side with an element $L \notin \mathcal{C}$, or the number of degrees of freedom of vector unknowns per side is greater than the number of degrees of freedom of scalar unknowns per element. Hence the basic form of the condensation with clusters around vertices does not apply. On the other hand, for both Raviart–Thomas and Brezzi–Douglas–Marini mixed finite elements of second order on simplicial meshes, the two above properties are satisfied. The extension of the basic condensation to this case, which may lead to an interesting relation between these second-order mixed finite element methods and the discontinuous Galerkin method, is an ongoing work.

4. Application to nonlinear parabolic problems

We show in this section that the above ideas easily apply also to the discretization of nonlinear parabolic convection–diffusion–reaction problems. We consider in particular the problem

$$\frac{\partial \beta(p)}{\partial t} + \nabla \cdot \mathbf{u} + F(p) = q \quad \text{in } \Omega \times (0,T), \tag{4.1a}$$

$$\mathbf{u} = -\mathbf{S}\nabla p + \psi(p)\mathbf{w} \quad \text{in } \Omega \times (0,T), \tag{4.1b}$$

$$p(\cdot,0) = p_0 \quad \text{in } \Omega, \tag{4.1c}$$

$$p = p_D \quad \text{on } \Gamma_D \times (0,T), \tag{4.1d}$$

$$\mathbf{u}\cdot\mathbf{n} = u_N \quad \text{on } \Gamma_N \times (0,T), \tag{4.1e}$$

where β, ψ, and F are monotone nonlinear functions, $\mathbf{S}$ is again a bounded, symmetric, and uniformly positive definite tensor, $\mathbf{w}$ is a velocity field, and q represents a source term.

Let again $\tilde{\mathbf{u}}$ be such that $\tilde{\mathbf{u}}\cdot\mathbf{n} = u_N$ on Γ_N in the appropriate sense. We split up the time interval $(0,T)$ such that $0 = t_0 < \ldots < t_n < \ldots < t_N = T$ and define $\triangle t_n := t_n - t_{n-1}$, $n \in \{1,2,\ldots,N\}$, and $p_h^0|_K$ by $(p_0,1)_K/|K|$ for all $K \in \mathcal{T}_h$. The fully implicit lowest-order Raviart–Thomas mixed finite element approximation of the problem (4.1a)–(4.1e), *cf.* [5], consists in finding on each time level t_n, $n \in \{1,2,\ldots,N\}$, the functions $\mathbf{u}_h^n = \mathbf{u}_{0,h}^n + \tilde{\mathbf{u}}^n$, $\mathbf{u}_{0,h}^n \in \mathbf{V}(\mathcal{E}_h)$, and $p_h^n \in \Phi(\mathcal{T}_h)$ such that

$$(\mathbf{S}^{-n}\mathbf{u}_{0,h}^n, \mathbf{v}_h)_\Omega - (\nabla\cdot\mathbf{v}_h, p_h^n)_\Omega - (\psi(p_h^n)\mathbf{w}^n, \mathbf{S}^{-n}\mathbf{v}_h)_\Omega = -\langle \mathbf{v}_h\cdot\mathbf{n}, p_D^n\rangle_{\partial\Omega} - (\mathbf{S}^{-n}\tilde{\mathbf{u}}^n, \mathbf{v}_h)_\Omega \qquad \forall\, \mathbf{v}_h \in \mathbf{V}(\mathcal{E}_h), \tag{4.2a}$$

$$\Big(\frac{\beta(p_h^n) - \beta(p_h^{n-1})}{\triangle t_n}, \phi_h\Big)_\Omega + (\nabla\cdot\mathbf{u}_{0,h}^n, \phi_h)_\Omega + (F(p_h^n), \phi_h)_\Omega = (q,\phi_h)_\Omega - (\nabla\cdot\tilde{\mathbf{u}}^n, \phi_h)_\Omega \qquad \forall \phi_h \in \Phi(\mathcal{T}_h), \tag{4.2b}$$

where

$$\mathbf{S}^{-n} := \frac{1}{\triangle t_n}\int_{t_{n-1}}^{t_n} \mathbf{S}^{-1}(\cdot,t)\,\mathrm{d}t, \quad \mathbf{w}^n := \frac{1}{\triangle t_n}\int_{t_{n-1}}^{t_n} \mathbf{w}(\cdot,t)\,\mathrm{d}t,$$

$$p_D^n := \frac{1}{\triangle t_n}\int_{t_{n-1}}^{t_n} p_D(\cdot,t)\,\mathrm{d}t, \quad \tilde{\mathbf{u}}^n := \frac{1}{\triangle t_n}\int_{t_{n-1}}^{t_n} \tilde{\mathbf{u}}(\cdot,t)\,\mathrm{d}t \qquad n \in \{1,2,\ldots,N\}.$$

Note that if $\beta = F = \psi = 0$, the matrix form of the problem (4.2a)–(4.2b) is given by (1.3), where the second equation is multiplied by -1. Such system matrix is not symmetric, but is positive definite, which is a favorable starting form for (4.2a)–(4.2b).

Everything we have to say about the application of the proposed condensation to the system (4.2a)–(4.2b) is that the terms where the unknown discrete velocity function $\mathbf{u}_{0,h}^n$ appears are exactly the same as in the linear elliptic case, see (1.2a)–(1.2b). Hence one can eliminate $\mathbf{u}_{0,h}^n$ on each discrete time level as in Section 2. This time, the flux unknowns are *nonlinear* functions of the scalar unknowns, convection velocity field, sources, and boundary conditions. The system (4.2a)–(4.2b), linearized by *e.g.* the Newton method, can be written in the matrix form as

$$\begin{pmatrix} \mathbb{A} & \mathbb{C} \\ \mathbb{B} & \mathbb{D} \end{pmatrix} \begin{pmatrix} U \\ P \end{pmatrix} = \begin{pmatrix} F \\ G \end{pmatrix}. \tag{4.3}$$

Let $V \in \mathcal{V}_h$ be a vertex and $\mathcal{C}_V$ the associated cluster and let us consider the linearized equations (4.2a) for the basis functions $\mathbf{v}_\gamma$, $\gamma \in \mathcal{F}_{\mathcal{C}_V}$, and the linearized equations (4.2b) for all ϕ_K, $K \in \mathcal{C}_V^{\mathrm{el}}$. This gives

$$\begin{pmatrix} \mathbb{A}_{1,V} & \mathbb{A}_{2,V} \\ \mathbb{B}_{1,V} & \mathbb{B}_{2,V} \end{pmatrix} \begin{pmatrix} U_V^{\mathcal{F}} \\ U_V^{\mathcal{G}} \end{pmatrix} = \begin{pmatrix} F_V - \mathbb{C}_V P_{1,V} \\ G_V - \mathbb{D}_V P_{2,V} \end{pmatrix}. \tag{4.4}$$

In fact, in the present case, $P_{1,V} = P_{2,V} = \{P_K\}_{K \in \mathcal{C}_V}$. We shall need the form (4.4) below for the upwind-mixed method. The matrix $\mathbb{B}_{2,V}$ is still diagonal, and hence we easily have

$$\mathbb{M}_V U_V^{\mathcal{F}} = F_V - \mathbb{C}_V P_{1,V} - \mathbb{A}_{2,V} \mathbb{B}_{2,V}^{-1} (G_V - \mathbb{D}_V P_{2,V}), \tag{4.5}$$

where the local condensation matrix associated with the vertex V, $\mathbb{M}_V = \mathbb{A}_{1,V} - \mathbb{A}_{2,V} \mathbb{B}_{2,V}^{-1} \mathbb{B}_{1,V}$, is the same as in the linear elliptic case. Hence its invertibility and the feasibility of the condensation in this form is determined by the rules studied in Section 3. Should $\mathbb{M}_V$ be invertible for all $V \in \mathcal{V}_h$, we have

$$U = \tilde{\mathbb{A}}^{-1}(F - \mathbb{C}P) - \mathbb{J}(G - \mathbb{D}P),$$

using (2.7). Here $\tilde{\mathbb{A}}^{-1}$ and $\mathbb{J}$ are given by (2.9). It now suffices to insert this expression for U into the second equation of (4.3) to obtain the final system for the scalar unknowns P only,

$$(-\mathbb{B}\tilde{\mathbb{A}}^{-1}\mathbb{C} + \mathbb{B}\mathbb{J}\mathbb{D} + \mathbb{D})P = G - \mathbb{B}\tilde{\mathbb{A}}^{-1}F + \mathbb{B}\mathbb{J}G. \tag{4.6}$$

This transcription enables in particular a straightforward implementation of the proposed condensation in existing mixed finite element codes.

Remark 4.1 (assemblage of $\tilde{\mathbb{A}}^{-1}$ and $\mathbb{J}$)**.** We note that the matrices $\tilde{\mathbb{A}}^{-1}$ and $\mathbb{J}$ only depend on the matrices $\mathbb{A}, \mathbb{B}$ of (4.3). Hence, if these matrices do not change (*i.e.* when the diffusion tensor $\mathbf{S}$ is constant with respect to time), the assemblage of $\tilde{\mathbb{A}}^{-1}$ and $\mathbb{J}$ can be done only once before the start of the calculation. On each time and linearization step, one then needs only $\mathbb{C}$, $\mathbb{D}$, F, and G from (4.3) to assemble the final linear system (4.6).

We now finally turn to the upwind-mixed lowest-order Raviart–Thomas method, *cf.* [16, 17, 25]. For this purpose, we first rewrite (4.1a)–(4.1b) as

$$\begin{aligned} \frac{\partial \beta(p)}{\partial t} + \nabla \cdot \mathbf{r} + \nabla \cdot (\psi(p)\mathbf{w}) + F(p) &= q \quad \text{in } \Omega \times (0,T), \\ \mathbf{r} &= -\mathbf{S}\nabla p \quad \text{in } \Omega \times (0,T). \end{aligned}$$

Whereas the initial and Dirichlet boundary conditions (4.1c) and (4.1d) stay the same, we rewrite the Robin boundary condition (4.1e) as a Neumann one,

$$\mathbf{r} \cdot \mathbf{n} = v_N \quad \text{on } \Gamma_N \times (0,T).$$

Let again $\tilde{\mathbf{r}}$ be such that $\tilde{\mathbf{r}} \cdot \mathbf{n} = v_N$ on Γ_N in the appropriate sense and define $\tilde{\mathbf{r}}^n := \frac{1}{\triangle t_n} \int_{t_{n-1}}^{t_n} \tilde{\mathbf{r}}(\cdot, t)\, \mathrm{d}t$, $n \in \{1, 2, \ldots, N\}$. The fully implicit upwind-mixed finite element method then reads: on each time level t_n, $n \in \{1, 2, \ldots, N\}$, find the functions $\mathbf{r}_h^n = \mathbf{r}_{0,h}^n + \tilde{\mathbf{r}}^n$, $\mathbf{r}_{0,h}^n \in \mathbf{V}(\mathcal{E}_h)$, and $p_h^n \in \Phi(\mathcal{T}_h)$ such that

$$(\mathbf{S}^{-n}\mathbf{r}_{0,h}^n, \mathbf{v}_h)_\Omega - (\nabla \cdot \mathbf{v}_h, p_h^n)_\Omega = -\langle \mathbf{v}_h \cdot \mathbf{n}, p_D^n \rangle_{\partial\Omega} - (\mathbf{S}^{-n}\tilde{\mathbf{r}}^n, \mathbf{v}_h)_\Omega \qquad \forall\, \mathbf{v}_h \in \mathbf{V}(\mathcal{E}_h), \tag{4.8a}$$

$$\Big(\frac{\beta(P_K^n) - \beta(P_K^{n-1})}{\triangle t_n}, \phi_K\Big)_K + (\nabla \cdot \mathbf{r}_{0,h}^n, \phi_K)_K + \sum_{\sigma \in \mathcal{E}_K} \psi(\widehat{p_\sigma^n}) \mathbf{w}_{K,\sigma}^n + (F(P_K^n), \phi_K)_K$$
$$= (q, \phi_K)_K - (\nabla \cdot \tilde{\mathbf{r}}^n, \phi_K)_K \qquad \forall K \in \mathcal{T}_h, \tag{4.8b}$$

where $\mathbf{w}_{K,\sigma}^n = \langle \mathbf{w}^n \cdot \mathbf{n}, 1 \rangle_\sigma$ and $\widehat{p_\sigma^n}$ is the upwind value defined respectively by

$$\widehat{p_\sigma^n} := \begin{cases} P_K^n & \text{if } \mathbf{w}_{K,\sigma}^n \geq 0 \\ P_L^n & \text{if } \mathbf{w}_{K,\sigma}^n < 0, \end{cases} \qquad \widehat{p_\sigma^n} := \begin{cases} P_K^n & \text{if } \mathbf{w}_{K,\sigma}^n \geq 0 \\ \langle p_D^n, 1 \rangle_\sigma / |\sigma| & \text{if } \mathbf{w}_{K,\sigma}^n < 0, \end{cases} \qquad \widehat{p_\sigma^n} := P_K^n$$

for σ an interior side between the elements K and L, for σ a Dirichlet boundary side, and for σ a Neumann boundary side. The linearization of the system (4.8a)–(4.8b) has again the form (4.3), with this time $\mathbb{C} = -\mathbb{B}^t$. The condensation applies again directly and in particular the final system has the form (4.6). The only difference is that because of the upstream weighting, $P_{1,V} \neq P_{2,V}$ in (4.4). In the expression for the fluxes through the $\mathcal{F}_{\mathcal{C}_V}$ sides, all the scalar unknowns associated with the elements sharing a side with an element from the cluster $\mathcal{C}_V$ may appear. Hence also the stencil of the final matrix is in this case wider: on a row of the final matrix corresponding to an element $K \in \mathcal{T}_h$, the only possible nonzero entries are on columns corresponding to $L \in \mathcal{T}_h$ such that L shares a common side with an element $M \in \mathcal{T}_h$ such that M and K share a common vertex. Finally, a similar observation to Remark 4.1 holds also in this case. Should $\mathbb{A}$ and $\mathbb{B}$ be constant, we only need to upload $\mathbb{D}$, F, and G on each time and linearization step, as in the finite volume method.

5. Numerical experiments

We give the results of several numerical experiments in two space dimensions in this section. We first compare the arising linear systems properties and the computational cost of the proposed condensation of the lowest-order Raviart–Thomas mixed finite element method with the hybridization approach for elliptic problems. We next compare the condensation with standard mixed solution approaches for nonlinear parabolic convection–diffusion–reaction problems. In all these tests, we employ the local condensation matrices associated with vertices. We finally numerically study the stability of this basic form of the condensation with respect to nearly singular cases and show that resorting to the clusters defined by all elements sharing a vertex with a given element (*cf.* Sect. 3.3) can eliminate this problem.

We employ two iterative methods for the solution of the arising sparse linear systems. If the matrix is symmetric and positive definite, we use the conjugate gradients (CG) method [23, 31]. For nonsymmetric matrices, we employ the bi-conjugate gradients stabilized (Bi-CGStab) method [31, 36]. To accelerate the convergence of these methods, we use incomplete Cholesky (IC) and incomplete LU (ILU) factorizations with a specified drop tolerance, *cf.* [35]. The drop tolerance is always chosen in such a way that the sum of CPU times of the preconditioning and of the solution of the preconditioned system was minimal. We denote the preconditioned methods by PCG and PBi-CGStab, respectively. We always use a zero start vector and stop the iterative process as soon as the relative residual $\|Y - \mathbb{M}\tilde{X}\|_2 / \|Y\|_2$, where $\tilde{X}$ is the approximate solution to the system $\mathbb{M}X = Y$, decreases below 1e-8. We focus on iterative solvers since they have reasonable memory requirements and, combined with *e.g.* the Newton method and a suitable preconditioning, allow for an efficient solution of nonlinear problems. Next, in the tables with results, we shall use the abbreviation SPD for a symmetric positive definite matrix, NPD for a nonsymmetric but positive definite matrix (recall that a real matrix $\mathbb{M} \in \mathbb{R}^{M \times M}$ is positive definite if $X^t \mathbb{M} X > 0$ for all $X \in \mathbb{R}^M$, $X \neq 0$), NNS for a nonsymmetric negative-stable matrix (a matrix whose all eigenvalues have positive real parts, which is in particular the case

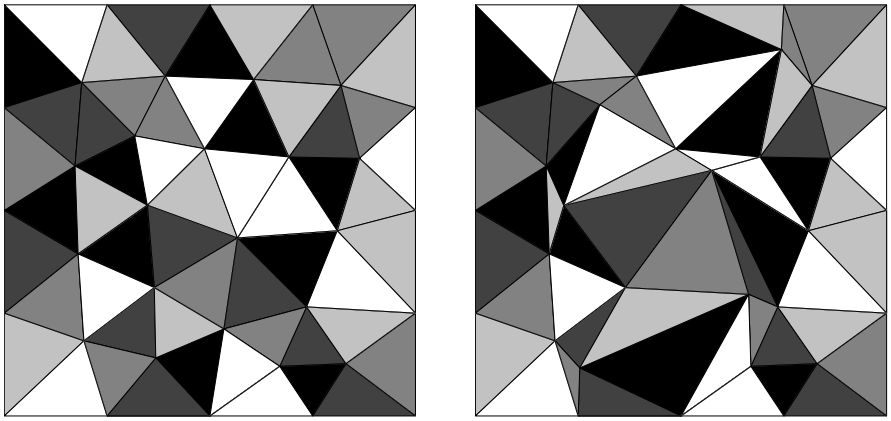


FIGURE 7. Initial meshes A (left) and B (right).

for positive definite matrices), and NID for a general nonsymmetric indefinite matrix. We further use st. for the stencil, *i.e.* for the maximum number of nonzero entries on each matrix row, and cond. for the 2-norm condition number (defined for a matrix $\mathbb{M}$ by $\|\mathbb{M}\|_2\|\mathbb{M}^{-1}\|_2$, or equivalently by the ratio of its largest and smallest singular value).

All the computations presented in this section were performed in double precision on a notebook with Intel Pentium 4-M 1.8 GHz processor and MS Windows XP operating system. Machine precision was in the power of 1e-16. All the matrix operations were done with the help of MATLAB 6.1.

5.1. **Condensed mixed finite element method for elliptic problems**

We consider in this section the problem (1.1a)–(1.1c) on $\Omega = (0,1) \times (0,1)$, where on the left edge, homogeneous Neumann boundary condition is prescribed and on the rest of the boundary, p is given by $p(x,y) = 0.1y + 0.9$. We perform the calculations on one to five regular refinements (each triangle is refined into four triangles by joining its edges midpoints) of the meshes viewed in Figure 7. In the mesh A, the minimal and maximal angles are equal to 35.4 and 88.7 degrees, and in the mesh B to 9.24 and 139 degrees, respectively. Note in particular that the mesh B is not Delaunay. A sink term $q = -0.001$ is prescribed on two elements of the initial mesh. Finally, the tensor $\mathbf{S}$ is given by

$$\mathbf{S}|_K = \begin{pmatrix} \cos(\theta_K) & -\sin(\theta_K) \\ \sin(\theta_K) & \cos(\theta_K) \end{pmatrix} \begin{pmatrix} s_K & 0 \\ 0 & \nu s_K \end{pmatrix} \begin{pmatrix} \cos(\theta_K) & \sin(\theta_K) \\ -\sin(\theta_K) & \cos(\theta_K) \end{pmatrix} \text{ for } K \in \mathcal{T}_h,$$

where we distinguish its following five different forms:

$$s_K = 1 \quad \forall K \in \mathcal{T}_h, \quad \nu = 1, \tag{5.1}$$

i.e. the homogeneous isotropic case ($\mathbf{S} = Id$), or

$$s_K = 1 \quad \forall K \in \mathcal{T}_h, \quad \theta_K \in \left\{ \frac{\pi}{5}, \frac{3\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{5}, \frac{\pi}{3} \right\}, \quad \nu = 0.2, \tag{5.2}$$

i.e. the (homogeneous with respect to s_K) anisotropic case ($\mathbf{S}$ is a full-matrix tensor), or

$$s_K \in \{10, 1, 0.1, 0.01, 0.001\}, \quad \nu = 1, \tag{5.3}$$

i.e. the inhomogeneous isotropic case (**S** is a varying scalar), or

$$s_K \in \{10, 5, 1, 0.5, 0.1\}, \quad \theta_K \in \left\{\frac{\pi}{5}, \frac{3\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{5}, \frac{\pi}{3}\right\}, \quad \nu = 0.2, \tag{5.4}$$

$$s_K \in \{10, 5, 1, 0.5, 0.1\}, \quad \theta_K \in \left\{\frac{\pi}{5}, \frac{3\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{5}, \frac{\pi}{3}\right\}, \quad \nu = 0.05, \tag{5.5}$$

i.e. the inhomogeneous anisotropic cases (**S** is a varying full matrix). The corresponding distributions of s_K and θ_K are viewed in Figure 7.

In Table 1, we compare different properties of the hybridization (implemented as the nonconforming finite element method, *cf.* [14]) with the condensed scheme (2.10). The number of unknowns in the hybridization is given by the number of non-Dirichlet edges. In the condensed scheme, this number is decreased by approximately 1/3 and is equal to the number of triangles. The stencil in the hybridization is equal to 5 and it was equal to 14 for the matrices issued from the condensation. The system matrix of the mixed-hybrid method is always symmetric and positive definite. In the condensation, the system matrix is nonsymmetric and its positive definiteness is only guaranteed if all the local condensation matrices $\mathbb{M}_V$ given by (2.5) are positive definite, which is in particular the case under condition (3.3). In the first three tested cases, this condition was verified by all $K \in \mathcal{T}_h$. Note in particular that this is true even for the quite strong anisotropy ratio $\nu = 0.2$ prescribed in the second case. In the fourth case (deformed mesh B, $\mathbf{S} = Id$), this criterion was violated by 11% of elements, but the system matrix still was positive definite. Positive definiteness was lost under subsequent increase of the anisotropy (mesh B, coefficients (5.4), 20% of elements violating condition (3.3)), but the matrix still was negative-stable. It is interesting that for the first refinement of the initial mesh, the system matrix was in fact even positive definite. Still increasing the anisotropy ratio ($\nu = 0.05$) in the last case, there were 72% of elements violating condition (3.3) and the condensed mixed finite element matrix was no more negative-stable. Nevertheless, the local condensation matrices associated with vertices were in all cases invertible and hence the condensation approach in its simplest form was feasible. The maximums of condition numbers of the matrices $\mathbb{M}_V$ over all $V \in \mathcal{V}_h$ were, respectively, 12, 48, 30 560, 108, 1372, and 125 825 in the six tested cases. Finally, the system matrices condition numbers are indicated in Table 1. It appears that in first five considered cases, the conditioning was 2 to 3-times better for the condensation, whereas in the last case, the conditioning was much worse for the condensation.

The above-discussed properties of the system matrices fundamentally influence the solution of the associated systems of linear equations. We first give the CPU time in seconds and the number of iterations necessary for the CG and BiCGStab methods, respectively. In rather homogeneous but possibly quite anisotropic cases (meshes A or B with $\mathbf{S} = Id$ or mesh A with **S** given by (5.2)), the CPU times of the condensation were from 1.5 to 1.85-times lower than those of the hybridization. For the strongly inhomogeneous case (5.3), the CPU times of the two methods were comparable. Finally, for the mesh B and **S** given by (5.5), BiCGStab converged very slowly for the condensation, since the system matrix was in this case not negative-stable. IC or ILU preconditioning of the above methods (whose time in seconds we report as well) enabled a considerable reduction in the necessary CPU times (but increased the memory requirements). It in particular also worked in the case where BiCGStab converged very slowly. Hence, in all considered cases, using this type of preconditioning, the condensation enabled a reduction in the CPU time necessary to solve the linear systems arising from the lowest-order mixed finite element method in comparison with the hybridization by a factor comprised between 1.2 and 1.6.

5.2. **Condensed mixed finite element method for nonlinear parabolic problems**

We compare in this section the condensed and standard mixed finite element methods for two nonlinear parabolic problems. We perform the simulations on one to five-times refined initial meshes from Figure 8. In the mesh C, the minimal and maximal angles are equal to 29.1 and 84.8 degrees, and in the mesh D to 15.3 and 135 degrees, respectively; the mesh D is again not Delaunay. The initial time step is equal to $T/2$ and is divided by two each time the space mesh is refined.

TABLE 1. Comparison of matrix properties and of the computational cost of the hybridized and condensed mixed finite element methods, elliptic problem (1.1a)–(1.1c).

	MHFE							Condensed MFE							
			CG		PCG						Bi-CGStab		PBi-CGStab		
Case	Unkn.	Cond.	CPU	Iter.	CPU	IC	Iter.	Unkn.	Matr.	Cond.	CPU	Iter.	CPU	ILU	Iter.
mesh A	316	314	0.03	82	0.02	0.01	5	216	NPD	183	0.03	55.0	0.02	0.01	1.5
cfs. (5.1)	1280	1352	0.21	171	0.05	0.02	8	864	NPD	725	0.16	87.0	0.04	0.02	2.5
	5152	5527	1.35	348	0.40	0.18	13	3456	NPD	2910	1.16	162.5	0.28	0.16	4.5
	20672	22260	12.35	691	3.02	1.48	18	13824	NPD	11673	8.86	295.5	1.99	1.03	8.5
	82816	89165	135.83	1358	22.48	11.90	27	55296	NPD	46762	89.17	615.0	16.82	9.31	14.5
mesh A	316	517	0.05	114	0.02	0.01	5	216	NPD	213	0.03	47.5	0.02	0.01	1.5
cfs. (5.2)	1280	2134	0.27	231	0.06	0.02	7	864	NPD	813	0.16	86.0	0.04	0.02	2.5
	5152	8676	1.75	466	0.40	0.20	12	3456	NPD	3195	1.32	181.5	0.29	0.15	5.0
	20672	34856	16.62	926	3.35	1.69	17	13824	NPD	12698	10.80	335.0	2.10	0.97	10.0
	82816	139477	192.23	1824	36.02	26.30	20	55296	NPD	50685	102.78	755.5	22.12	6.65	34.5
mesh A	316	165877	0.47	1123	0.02	0.01	5	216	NPD	86082	0.46	677.5	0.02	0.01	1.5
cfs. (5.3)	1280	606930	2.95	3327	0.05	0.02	9	864	NPD	302639	2.79	1579.5	0.04	0.02	2.5
	5152	2.25e+6	27.91	7066	0.41	0.18	12	3456	NPD	1.15e+6	29.62	4028.5	0.28	0.13	5.0
	20672	8.49e+6	268.41	14009	3.11	1.54	17	13824	NPD	4.38e+6	279.78	9203.5	2.25	1.00	10.0
	82816	3.23e+7	2514.15	26728	24.57	10.90	33	55296	NPD	1.68e+7	2662.26	20148.5	20.37	13.23	12.0
mesh B	316	1204	0.05	129	0.02	0.01	5	216	NPD	278	0.04	48.5	0.02	0.01	1.5
cfs. (5.1)	1280	5236	0.33	300	0.06	0.02	9	864	NPD	1424	0.23	116.5	0.04	0.02	2.5
	5152	21382	2.55	660	0.40	0.17	14	3456	NPD	6495	2.01	268.5	0.28	0.16	4.5
	20672	86007	25.88	1370	3.05	1.36	21	13824	NPD	27151	17.55	558.5	2.13	1.23	7.5
	82816	344380	261.47	2740	27.81	12.99	34	55296	NPD	110002	173.92	1146.5	21.08	13.54	12.5
mesh B	316	10398	0.14	402	0.02	0.01	5	216	NPD	2479	0.10	148.5	0.02	0.01	2.5
cfs. (5.4)	1280	38970	0.87	885	0.06	0.02	10	864	NNS	10699	0.66	346.5	0.04	0.02	3.5
	5152	146249	7.80	1881	0.41	0.15	14	3456	NNS	45264	5.60	741.5	0.29	0.17	4.5
	20672	561083	69.92	3796	3.44	1.60	19	13824	NNS	183063	49.39	1635.0	2.28	1.33	7.0
	82816	2.18e+6	708.13	7410	33.40	22.86	22	55296	NNS	723345	539.41	4056.5	21.39	15.05	8.5
mesh B	316	21846	0.21	591	0.02	0.01	5	216	NID	238294	0.32	282.5	0.02	0.01	2.0
cfs. (5.5)	1280	68078	1.28	1284	0.07	0.02	11	864	NID	861441	1.92	1053.5	0.04	0.02	3.0
	5152	241229	9.80	2616	0.44	0.15	15	3456	NID	2.60e+6	410.13	54541.5	0.29	0.17	4.0
	20672	906060	91.60	5141	3.60	2.09	18	13824	NID	9.68e+6	–	–	2.42	1.61	5.5
	82816	3.50e+6	981.17	10052	33.75	22.36	24	55296	NID	3.74e+7	–	–	22.41	17.55	6.5

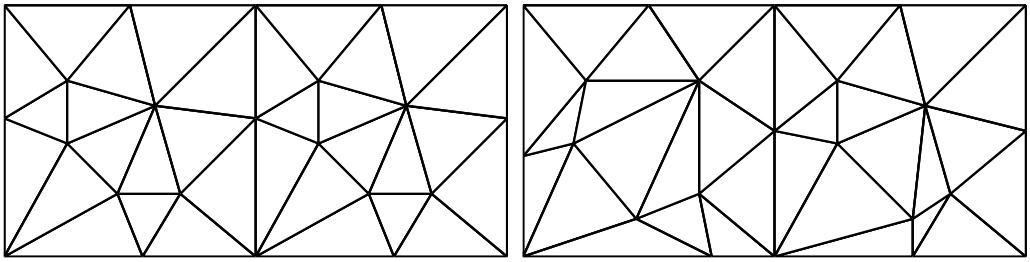


FIGURE 8. Initial meshes C (left) and D (right).

5.2.1. *A reaction–diffusion problem*

For $\Omega = (0,2) \times (0,1)$ and $T = 1$, we consider the nonlinear reaction–diffusion problem

$$\frac{\partial (p + p^\alpha)}{\partial t} - \nabla \cdot (\mathbf{S} \nabla p) + 3p + \alpha p^\alpha = 0 \tag{5.6}$$

with $\alpha = 0.5$ and either

$$\mathbf{S} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ in } \Omega \tag{5.7}$$

or

$$\mathbf{S} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ for } x < 1, \quad \mathbf{S} = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} \text{ for } x > 1. \tag{5.8}$$

Initial and Dirichlet boundary conditions are given by the exact solution $p(x,y,t) = e^x e^y e^{-t}/e^3$. Notice that the flux of the solution given by $-\mathbf{S}\nabla p$ has a continuous normal trace across the discontinuity line $x = 1$ for the diffusion tensor (5.8). The derivative of the function p^α, $\alpha = 0.5$, blows up in 0 but the problem is not really degenerate parabolic, since the exact solution does not take the value of 0. We consider the condensation of the mixed finite element method (4.6) and the mixed finite element method (4.2a)–(4.2b). We notice that the system of equations of the mixed method has on each time and linearization step the form (4.3), where $\mathbb{D}$ is a diagonal matrix. Hence a standard solution approach is to inverse $\mathbb{D}$, then solve for U the system $(\mathbb{A} - \mathbb{C}\mathbb{D}^{-1}\mathbb{B})U = F - \mathbb{C}\mathbb{D}^{-1}G$, and finally recover P from $P = \mathbb{D}^{-1}(G - \mathbb{B}U)$. In fact, in the present case, $\mathbb{C} = \mathbb{B}^t$, and thus the final system matrix is symmetric. It is noted in [24] that this approach is not suitable when the term occurring in the time derivative and the reaction term are too small in comparison with the other terms, which is however not the present case. On the contrary, according to [24], such solution approach is more reliable than the hybridization of the mixed finite element method for parabolic problems with general diffusion tensors.

We compare the properties of the system matrices and the computational cost for the first time and Newton linearization steps in Table 2. The CPU time of the condensed mixed finite element method is about 2-times shorter than the CPU time of the standard approach in the case of the tensor (5.7) and the initial mesh C. When full-matrix and discontinuous diffusion tensor (5.8) and a less regular mesh D are used, then the CPU time of the condensed version is more than 4-times shorter when no preconditioning is used and more than 2-times shorter with preconditioning. Note the important increase of the condition number of the system matrix of the standard mixed finite element method for the tensor (5.8).

5.2.2. *A convection–diffusion–reaction problem*

For $\Omega = (0,2) \times (0,1)$ and $T = 1$, we consider the nonlinear convection–diffusion–reaction problem

$$\frac{\partial (p + p^\alpha)}{\partial t} - \nabla \cdot (\mathbf{S} \nabla p) + \nabla \cdot (p\mathbf{w}) + \alpha p^\alpha = 0 \tag{5.9}$$

with $\alpha = 0.5$ and either

$$\mathbf{S} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ in } \Omega, \quad \mathbf{w} = (3,0) \text{ in } \Omega \tag{5.10}$$

or

$$\mathbf{S} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ for } x < 1, \quad \mathbf{S} = \begin{pmatrix} 8 & -7 \\ -7 & 20 \end{pmatrix} \text{ for } x > 1,$$
$$\mathbf{w} = (3,0) \text{ for } x < 1, \quad \mathbf{w} = (3,12) \text{ for } x > 1. \tag{5.11}$$

Initial and Dirichlet boundary conditions are again given by the exact solution $p(x,y,t) = e^x e^y e^{-t}/e^3$. Notice that in the case of the coefficients given by (5.11), the velocity field $\mathbf{w}$ as well as the flux of the solution given by $-\mathbf{S}\nabla p + (p\mathbf{w})$ have a continuous normal trace across the discontinuity line $x = 1$. The problem is not convection-dominated, and hence we can use the mixed finite element method (4.2a)–(4.2b). Notice that the associated linear system on each time and linearization step has again the form (4.3) with $\mathbb{D}$ a diagonal matrix. Hence the same solution approach as in the previous section can be used. In this case however $\mathbb{C} \neq \mathbb{B}^t$, and thus the final system for U is nonsymmetric.

We compare the properties of the linear systems and the computational cost for the first time and Newton linearization steps in Table 3. The settings are the same as in the previous section, except for the fact that we have to use the Bi-CGStab method and the LU incomplete factorization also for the standard mixed approach in view of the nonsymmetry of its system matrices. One can observe that the increase of the condition number of the system matrix of the condensed mixed finite element method with less regular coefficients and mesh is much less important than that of the standard mixed finite element method. Hence the CPU time of the unpreconditioned Bi-CGStab method for the condensed version is about 3-times shorter for the coefficients (5.10) and mesh C, but about 10-times shorter for the coefficients (5.11) and mesh D. Using the preconditioning considerably smears the difference. The CPU time of the condensed version is then about 2-times shorter.

5.2.3. *A convection–diffusion–reaction problem and the upwind-mixed method*

We consider here once more the problem (5.9) with coefficients (5.10) and mesh C. This time, we employ the upwind-mixed finite element method (4.8a)–(4.8b) and the corresponding condensed version.

We compare the properties of the linear systems on the first time and Newton linearization steps in Table 4. Although there is an increase in the stencil of the condensed upwind-mixed finite element method, the system matrix condition number and CPU times are very similar to the condensed mixed finite element method, *cf.* Table 3. The system for the upwind-mixed finite element method on each time and linearization step has again the form (4.3). The matrix $\mathbb{D}$ is however in this case not diagonal, and hence we cannot easily eliminate the scalar unknowns P. We thus consider the whole matrix for the unknowns U and P. This matrix is very well conditioned, nonsymmetric, and positive definite, but the direct application of the Bi-CGStab method does not lead to satisfactory results, *cf.* Table 4. Also the direct LU incomplete factorization is almost impossible, since the LU factors tend to considerably increase the fill-in. A suitable solution approach however seems to be to first perform the column minimum degree permutation [22]. The matrix with permuted columns then has sparser LU incomplete factors, which can in turn be successfully used as preconditioners. The memory requirements of such approach are however still considerably higher than those of the condensation, which may limit its use for large problems. We report in Table 4 the CPU times necessary for finding the column minimum degree permutation and LU incomplete factorization of the matrix with permuted columns, in addition to the total CPU time. In the present case, the condensation reduces the CPU time again by a factor better than 2.

5.3. Stability with respect to nearly singular local condensation matrices, practical remedies of this problem

We have shown in Example 3.10 that the local condensation matrices associated with vertices given by (2.5) may become singular for a deformed mesh when $\mathbf{S} = Id$. This may likewise happen for a "nice" mesh but

TABLE 2. Comparison of matrix properties and of the computational cost of the standard and condensed mixed finite element methods, first time and linearization step, parabolic reaction–diffusion problem (5.6).

	MFE								Condensed MFE							
				CG		PCG						Bi-CGStab		PBi-CGStab		
Case	Unkn.	Matr.	Cond.	CPU	Iter.	CPU	IC	Iter.	Unkn.	Matr.	Cond.	CPU	Iter.	CPU	ILU	Iter.
mesh C	204	SPD	290	0.05	110	0.02	0.01	5	128	NPD	37	0.02	29.5	0.02	0.01	2.0
cfs. (5.7)	792	SPD	764	0.14	206	0.04	0.02	7	512	NPD	109	0.06	50.0	0.02	0.01	2.5
	3120	SPD	1770	0.95	333	0.18	0.08	11	2048	NPD	298	0.37	80.5	0.10	0.06	3.0
	12 384	SPD	3820	5.36	508	1.21	0.58	14	8192	NPD	747	2.45	122.5	0.68	0.38	5.0
	49 344	SPD	7974	34.45	743	8.17	3.83	18	32 768	NPD	1753	14.24	175.0	4.75	2.95	7.0
mesh D	204	SPD	1358	0.07	170	0.02	0.01	6	128	NPD	61	0.02	31.0	0.02	0.01	2.0
cfs. (5.8)	792	SPD	4314	0.36	409	0.04	0.02	10	512	NPD	225	0.07	62.0	0.02	0.01	2.5
	3120	SPD	11 506	2.23	836	0.28	0.12	16	2048	NPD	676	0.49	119.0	0.12	0.07	3.5
	12 384	SPD	28 188	15.93	1456	1.61	0.68	20	8192	NPD	1814	3.76	212.0	0.75	0.39	6.0
	49 344	SPD	65 024	108.51	2279	11.75	5.76	27	32 768	NPD	4603	26.62	335.5	5.31	3.02	8.0

TABLE 3. Comparison of matrix properties and of the computational cost of the standard and condensed mixed finite element methods, first time and linearization step, parabolic convection–diffusion–reaction problem (5.9).

	MFE								Condensed MFE							
				Bi-CGStab		PBi-CGStab						Bi-CGStab		PBi-CGStab		
Case	Unkn.	Matr.	Cond.	CPU	Iter.	CPU	ILU	Iter.	Unkn.	Matr.	Cond.	CPU	Iter.	CPU	ILU	Iter.
mesh C	204	NPD	405	0.06	95.5	0.02	0.01	2.0	128	NPD	39	0.02	27.0	0.02	0.01	2.0
cfs. (5.10)	792	NPD	917	0.22	153.0	0.04	0.03	3.0	512	NPD	116	0.07	56.5	0.02	0.01	2.5
	3120	NPD	1949	1.36	282.0	0.22	0.14	4.0	2048	NPD	311	0.38	82.5	0.11	0.06	3.5
	12 384	NPD	4016	8.47	406.5	1.51	0.94	5.0	8192	NPD	768	2.65	139.0	0.75	0.41	5.5
	49 344	NPD	8181	51.18	553.0	10.26	6.94	6.0	32 768	NPD	1782	17.14	191.5	4.85	2.95	7.0
mesh D	204	NPD	13 849	0.23	412.5	0.02	0.01	2.0	128	NPD	470	0.04	70.0	0.02	0.01	2.0
cfs. (5.11)	792	NPD	39 935	1.38	1105.5	0.03	0.02	2.5	512	NPD	1665	0.21	149.5	0.03	0.01	2.5
	3120	NPD	103 342	12.12	2419.5	0.22	0.18	3.0	2048	NPD	4824	1.47	322.5	0.12	0.07	3.5
	12 384	NPD	250 923	103.42	5390.5	1.84	1.32	4.0	8192	NPD	12 523	8.66	474.5	0.88	0.56	5.0
	49 344	NPD	586 375	617.26	7145.5	16.04	11.04	7.0	32 768	NPD	31368	61.53	787.5	7.47	5.46	5.5

TABLE 4. Comparison of the computational cost of the standard and condensed upwind-mixed finite element methods, first time and linearization step, parabolic convection–diffusion–reaction problem (5.9), coefficients (5.10), mesh C.

Upwind-MFE									
				Bi-CGStab		PBi-CGStab			
Unkn.	Matr.	St.	Cond.	CPU	Iter.	CPU	Per.	ILU	Iter.
332	NPD	7	17	0.18	235.5	0.03	0.01	0.01	2.0
1304	NPD	7	29	1.17	549.5	0.06	0.01	0.03	2.5
5168	NPD	7	67	13.01	1540.5	0.31	0.03	0.15	3.5
20 576	NPD	7	168	124.06	3561.5	1.98	0.32	0.98	4.0
82 112	NPD	7	393	3233.05	16 763.5	10.58	1.35	4.92	6.0
Condensed upwind-MFE									
128	NPD	19	42	0.02	25.5	0.02		0.01	2.0
512	NPD	19	120	0.09	57.0	0.02		0.01	2.5
2048	NPD	19	318	0.46	88.0	0.11		0.06	3.0
8192	NPD	19	777	2.99	138.5	0.68		0.36	5.0
32 768	NPD	19	1792	18.86	210.5	4.89		2.87	7.5

TABLE 5. Condensation with local condensation matrices associated with vertices versus condensation with local condensation matrices associated with elements for perturbations of the mesh from Figure 6.

	Condensation around vertices				Condensation around elements		
Shift	Cond. loc.	Cond.	L^∞ error	L^2 error	Cond. loc.	Cond.	Error
1e-1	1.07e+3	2.30e+4	5.64e-14	4.71e-13	177	2318	machine precision
1e-2	9.09e+4	2.34e+6	1.19e-11	8.71e-11	118	2270	machine precision
1e-4	9.00e+8	2.34e+10	2.71e-7	2.11e-6	115	2269	machine precision
1e-6	9.00e+12	2.34e+14	1.69e-3	1.28e-2	115	2269	machine precision
1e-8	9.12e+16	2.36e+18	4.82e-1	1.95e+0	115	2269	machine precision

a general tensor $\mathbf{S}$. We have next discussed in Section 3.3 the variants of the basic form of the condensation aiming to overcome this possible difficulty. We now illustrate these affairs practically.

We report in Table 5 the stability of the basic form of the condensation while approaching the singular case of Example 3.10. We suppose that Figure 6 represents a mesh of a domain $\Omega = (0,1) \times (0,1)$ and vary by 1e-1 to 1e-8 the x coordinate of the point lying in the middle of the bottom edge. Hence the local condensation matrices associated with the vertex V are no more singular, but they are increasingly close to. We consider three-times refined original mesh from Figure 6 and indicate for this case the condition numbers of the local condensation matrices associated with V, as well as the condition numbers of the system matrices. We next give L^∞ and L^2 errors between the approximate solution vectors P arising from the condensation in this form and from the standard formulation (1.3), while using the exact solver of MATLAB 6.1 based on LU factorization. We can see that the important increase in the conditioning destroys the precision of the results. However, using the clusters defined by all elements sharing a vertex with a given element (*cf.* Sect. 3.3), the condensation is easily feasible, the condition numbers of local condensation matrices as well as those of the system matrices are bounded, and we again end up with the right solution up to the machine precision. Hence, resorting to such clusters eliminates the problem in the given case.

5.4. **Conclusions**

We have studied in this section the computational cost of the proposed condensation of the mixed finite element method for elliptic and (nonlinear) parabolic problems.

For elliptic problems in two space dimensions, the standard hybridization leads to systems for the number of unknowns equal to the number of edges with symmetric positive definite matrices with a 5-point stencil. In the proposed condensation, the number of unknowns is reduced to the number of elements (which is approximately 2/3 of the number of edges), but the system matrices are in general nonsymmetric, have a wider (about 14-point) stencil, and are positive definite only under a condition on the mesh and the diffusion tensor. This condition however allows for quite deformed triangles in the case of a piecewise constant scalar diffusion tensor. The CPU time speed-ups for the test cases were comprised between 1.2 and 1.85. The finite volume reformulation of the mixed finite element method proposed and studied in [13,39,40] leads to symmetric matrices with the number of unknowns equal to the number of elements and a 4-point stencil. The matrices are positive definite for Delaunay triangulations and constant scalar diffusion tensors but generally indefinite otherwise. Hence the computational savings of the reformulation will be very probably more important than those of the condensation for Delaunay triangulations and constant scalar diffusion tensors. The situation should be much more favorable for the condensation when the mesh is not Delaunay or when the diffusion tensor is inhomogeneous and anisotropic. The performed tests in particular show that the matrix problems arising from the condensation in such cases may easily be solved by the usual iterative solvers, which is far from being the case in the method proposed in [13, 39, 40], *cf.* the numerical experiments carried out in these references. In three space dimensions, the finite volume reformulation is in general not possible, see [40]. In contrast, the condensation applies directly as in two space dimensions. Moreover, the number of unknowns is in this case only about 1/2 of that of the hybridization. Hence one can expect even more important computational savings than in the two-dimensional case.

Next, the proposed condensation applies as well to mixed finite element discretizations of (nonlinear) parabolic convection–diffusion–reaction problems. The resulting matrices are still sparse, positive definite for a large class of meshes and diffusion tensors, nonsymmetric, and seem to be very well conditioned. Moreover, if the diffusion tensor is constant with respect to time, one can assemble and invert the local condensation matrices only once before the start of the calculation and then only work with the scalar unknowns as in the finite volume method, which still reduces the computational complexity. In two space dimensions, the number of unknowns is equal to approximately 2/3 of that of standard solution approaches in the mixed finite element method and to approximately 2/5 of that of the upwind-mixed method. The CPU times necessary for the solution of the associated linear systems in the presented test cases were reduced by a factor 2 for parabolic reaction–diffusion problems. When convection is present, nonsymmetric matrices arise naturally also in the mixed and upwind-mixed schemes, which can further increase the speed-up. The finite volume reformulation of the mixed finite element method is possible for parabolic reaction–diffusion problems, but leads in general to indefinite nonsymmetric systems with a limited gain in the terms of the computational cost, *cf.* [13, 39]. Hence the condensation seems to be much more attractive in this case. This is still emphasized by the fact that it can be very easily implemented into existing mixed finite element codes. Finally, the speed-up should be even more important in three space dimensions, where the number of unknowns of the condensation is about 1/2 of that of the mixed and 1/3 of that of the upwind-mixed schemes.

Acknowledgements. The author would like to thank his Ph.D. advisor Danielle Hilhorst from the University of Paris-South and Professor Robert Eymard from the University of Marne-la-Vallée for their valuable advice and hints.

References

[1] I. Aavatsmark, T. Barkve, Ø. Bøe and T. Mannseth, Discretization on unstructured grids for inhomogeneous, anisotropic media. Part I: Derivation of the methods. *SIAM J. Sci. Comput.* **19** (1998) 1700–1716.

[2] I. Aavatsmark, T. Barkve, Ø. Bøe and T. Mannseth, Discretization on unstructured grids for inhomogeneous, anisotropic media. Part II: Discussion and numerical results. *SIAM J. Sci. Comput.* **19** (1998) 1717–1736.

[3] M. Aftosmis, D. Gaitonde and T. Sean Tavares, On the accuracy, stability and monotonicity of various reconstruction algorithms for unstructured meshes. *AIAA* (1994), paper No. 94-0415.
[4] A. Agouzal, J. Baranger, J.-F. Maître and F. Oudin, Connection between finite volume and mixed finite element methods for a diffusion problem with nonconstant coefficients. Application to a convection diffusion problem. *East-West J. Numer. Math.* **3** (1995) 237–254.
[5] T. Arbogast, M.F. Wheeler and N. Zhang, A nonlinear mixed finite element method for a degenerate parabolic equation arising in flow in porous media. *SIAM J. Numer. Anal.* **33** (1996) 1669–1687.
[6] T. Arbogast, M.F. Wheeler and I. Yotov, Mixed finite elements for elliptic problems with tensor coefficients as cell-centered finite differences. *SIAM J. Numer. Anal.* **34** (1997) 828–852.
[7] T. Arbogast, C.N. Dawson, P.T. Keenan, M.F. Wheeler and I. Yotov, Enhanced cell-centered finite differences for elliptic equations on general geometry. *SIAM J. Sci. Comput.* **19** (1998) 404–425.
[8] D.N. Arnold and F. Brezzi, Mixed and nonconforming finite element methods: Implementation, postprocessing and error estimates. *RAIRO Modél. Math. Anal. Numér.* **19** (1985) 7–32.
[9] J. Baranger, J.-F. Maître and F. Oudin, Connection between finite volume and mixed finite element methods. *RAIRO Modél. Math. Anal. Numér.* **30** (1996) 445–465.
[10] F. Brezzi and M. Fortin, *Mixed and Hybrid Finite Element Methods*. Springer-Verlag, New York (1991).
[11] F. Brezzi, J. Douglas Jr. and L.D. Marini, Two families of mixed finite elements for second order elliptic problems. *Numer. Math.* **47** (1985) 217–235.
[12] F. Brezzi, J. Douglas Jr., R. Duran and M. Fortin, Mixed finite elements for second order elliptic problems in three variables. *Numer. Math.* **51** (1987) 237–250.
[13] G. Chavent, A. Younès and Ph. Ackerer, On the finite volume reformulation of the mixed finite element method for elliptic and parabolic PDE on triangles. *Comput. Methods Appl. Mech. Engrg.* **192** (2003) 655–682.
[14] Z. Chen, Equivalence between and multigrid algorithms for nonconforming and mixed methods for second-order elliptic problems. *East-West J. Numer. Math.* **4** (1996) 1–33.
[15] Y. Coudière, J.-P. Vila and Villedieu Ph., Convergence rate of a finite volume scheme for a two dimensional convection-diffusion problem. *ESAIM: M2AN* **33** (1999) 493–516.
[16] C. Dawson, Analysis of an upwind-mixed finite element method for nonlinear contaminnat transport equations. *SIAM J. Numer. Anal.* **35** (1998) 1709–1724.
[17] C. Dawson and V. Aizinger, Upwind-mixed methods for transport equations. *Comput. Geosci.* **3** (1999) 93–110.
[18] J. Douglas Jr. and J.E. Roberts, Global estimates for mixed methods for second order elliptic equations. *Math. Comp.* **44** (1985) 39–52.
[19] R. Eymard, T. Gallouët and R. Herbin, Finite volume methods, in *Handbook of Numerical Analysis*, Ph.G. Ciarlet and J.-L. Lions Eds. Elsevier Science B.V., Amsterdam **7** (2000) 713–1020.
[20] R. Eymard, T. Gallouët and R. Herbin, A cell-centred finite-volume approximation for anisotropic diffusion operators on unstructured meshes in any space dimension. *IMA J. Numer. Anal.* **26** (2006) 326–353.
[21] I. Faille, A control volume method to solve an elliptic equation on a two-dimensional irregular mesh. *Comput. Methods Appl. Mech. Engrg.* **100** (1992) 275–290.
[22] J.R. Gilbert, C. Moler and R. Schreiber, Sparse matrices in MATLAB: Design and implementation. *SIAM J. Matrix Anal. Appl.* **13** (1992) 333–356.
[23] M.R. Hestenes and E. Stiefel, Methods of conjugate gradients for solving linear systems. *J. Res. Nat. Bur. Stand.* **49** (1952) 409–436.
[24] H. Hoteit, J. Erhel, R. Mosé, B. Philippe and Ph. Ackerer, Numerical reliability for mixed methods applied to flow problems in porous media. *Comput. Geosci.* **6** (2002) 161–194.
[25] J. Jaffré, Éléments finis mixtes et décentrage pour les équations de diffusion-convection. *Calcolo* **23** (1984) 171–197.
[26] L. Jeannin, I. Faille and T. Gallouët, Comment modéliser les écoulements diphasiques compressibles sur des grilles hybrides ? *Oil & Gas Science and Technology – Rev. IFP* **55** (2000) 269–279.
[27] R.A. Klausen and G.T. Eigestad, Multi point flux approximations and finite element methods; practical aspects of discontinuous media, *Proc. 9th European Conference on the Mathematics of Oil Recovery*, Cannes, France, B003 (2004).
[28] R.A. Klausen and T.F. Russell, Relationships among some locally conservative discretization methods which handle discontinuous coefficients. *Comput. Geosci.* **8** (2004) 341–377.
[29] L.D. Marini, An inexpensive method for the evaluation of the solution of the lowest order Raviart–Thomas mixed method. *SIAM J. Numer. Anal.* **22** (1985) 493–496.
[30] J.C. Nédélec, Mixed finite elements in $\mathbb{R}^3$. *Numer. Math.* **35** (1980) 315–341.
[31] A. Quarteroni and A. Valli, *Numerical Approximation of Partial Differential Equations*. Springer-Verlag, Berlin (1994).
[32] P.-A. Raviart and J.-M. Thomas, A mixed finite element method for 2-nd order elliptic problems, in *Mathematical Aspects of Finite Element Methods*. Galligani I., Magenes E. Eds., *Lect. Notes Math.*, Springer, Berlin **606** (1977) 292–315.
[33] J.E. Roberts and J.-M. Thomas, Mixed and hybrid methods, in *Handbook of Numerical Analysis*, Ph.G. Ciarlet and J.-L. Lions Eds., Elsevier Science B.V., Amsterdam **2** (1991) 523–639.

[34] T.F. Russell and M.F. Wheeler, Finite element and finite difference methods for continuous flows in porous media, in *The Mathematics of Reservoir Simulation*, R.E. Ewing Ed., SIAM, Philadelphia (1983) 35–106.
[35] Y. Saad, *Iterative Methods for Sparse Linear Systems*. PWS Publishing Company (1996).
[36] H.A. van der Vorst, Bi-CGSTAB: A fast and smoothly converging variant of Bi-CG for the solution of non-symmetric linear systems. *SIAM J. Sci. Stat. Comput.* **13** (1992) 631–644.
[37] M. Vohralík, Equivalence between mixed finite element and multi-point finite volume methods. *C. R. Acad. Sci. Paris., Ser. I* **339** (2004) 525–528.
[38] M. Vohralík, Equivalence between mixed finite element and multi-point finite volume methods. Derivation, properties, and numerical experiments, in *Proceedings of ALGORITMY 2005*, Slovak University of Technology, Slovakia (2005) 103–112.
[39] A. Younès, R. Mose, Ph. Ackerer and G. Chavent, A new formulation of the mixed finite element method for solving elliptic and parabolic PDE with triangular elements. *J. Comput. Phys.* **149** (1999) 148–167.
[40] A. Younès, Ph. Ackerer and G. Chavent, From mixed finite elements to finite volumes for elliptic PDEs in two and three dimensions. *Internat. J. Numer. Methods Engrg.* **59** (2004) 365–388.