

EQUIVALENCE BETWEEN LOWEST-ORDER MIXED FINITE ELEMENT AND MULTI-POINT FINITE VOLUME METHODS. DERIVATION, PROPERTIES, AND NUMERICAL EXPERIMENTS*

MARTIN VOHRALÍK[†]

Abstract. We consider the lowest-order Raviart–Thomas mixed finite element method for elliptic diffusion problems on simplicial meshes in two or three space dimensions. This method produces saddle-point problems for scalar and flux unknowns. We show how to easily eliminate the flux unknowns, which implies an 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–reaction–diffusion problems. We finally present a set of numerical experiments confirming important computational savings while using the equivalent finite volume form of the lowest-order mixed finite element method.

Key words. lowest-order Raviart–Thomas mixed finite element method, saddle-point problem, multi-point finite volume method, elliptic diffusion equation, nonlinear parabolic convection–reaction–diffusion equation

AMS subject classifications. 76M10, 76M12, 76S05

1. Introduction. Let us consider the elliptic problem

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

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

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

where $\Omega \subset \mathbb{R}^d$, $d = 2, 3$, is a polygonal domain (open, bounded, and connected set), \mathbf{S} is a bounded, symmetric (this is however not necessary), and uniformly positive definite tensor, $p_D \in H^{\frac{1}{2}}(\Gamma_D)$, $u_N \in H^{-\frac{1}{2}}(\Gamma_N)$, $q \in L_2(\Omega)$, $\Gamma_D \cap \Gamma_N = \emptyset$, $\overline{\Gamma_D} \cup \overline{\Gamma_N} = \partial\Omega$, and $|\Gamma_D| \neq 0$, where $|\Gamma_D|$ is the measure of the set Γ_D .

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 lies entirely either in Γ_D or in Γ_N . Let us denote by \mathcal{E}_h the set of all non-Neumann sides (edges if $d = 2$, faces if $d = 3$) of \mathcal{T}_h . Let finally $\tilde{\mathbf{u}} \in \mathbf{H}(\text{div}, \Omega)$ 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 [5, 12])

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

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

*This work was supported by Ministry of Education of the Czech Republic, project code MSM 242200001.

[†]Laboratoire de Mathématiques, Analyse Numérique et EDP, Université de Paris-Sud et CNRS, Bât. 425, 91405 Orsay, France & Department of Modelling of Processes, Faculty of Mechatronics, Technical University of Liberec, Hálkova 6, 461 17 Liberec, Czech Republic
martin.vohralik@math.u-psud.fr

where $(\mathbf{u}, \mathbf{v})_\Omega = \int_\Omega \mathbf{u} \cdot \mathbf{v} \, d\mathbf{x}$, $\langle \mathbf{v} \cdot \mathbf{n}, \varphi \rangle_{\partial\Omega} = \int_{\partial\Omega} \mathbf{v} \cdot \mathbf{n} \varphi \, d\gamma(\mathbf{x})$, 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 saddle-point and can be written in the form

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

In the lowest-order Raviart–Thomas method [11] and its three-dimensional Nédélec variant [10] 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 Lagrange multipliers associated with non-Dirichlet sides and obtain a symmetric and positive definite matrix, cf. [3, 5]. Especially in three space dimensions, there are much less elements than sides, and hence the long-standing interest in reducing the unknowns to only the scalar unknowns P . When \mathbf{S} is diagonal, this is indeed possible, using approximate numerical integration, cf. [4, 13]. Using the expanded mixed finite element method, these techniques can be extended also onto full-matrix diffusion tensors, cf. [2]. To our knowledge, the only technique for reducing the number of unknowns to the number of elements without any numerical integration is studied in [14]. 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. One associates here to each element a *new* unknown.

We present in this paper a new method which permits to exactly and efficiently reduce the system (1.3) onto a system for the *original* scalar unknowns P only. It shows that in the lowest-order Raviart–Thomas mixed finite element method, one can express, solving only local problems, the flux through each side using the scalar unknowns, sources, and possibly boundary conditions associated with the elements in a neighborhood of this side. This method is thus 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 describe the stencil of the final matrix and give sufficient conditions for its symmetry and positive definiteness. We next apply the condensation to mixed and upwind-mixed (cf. [7]) finite element discretizations of nonlinear parabolic convection–reaction–diffusion problems. We finally present numerical examples confirming considerable computational savings while using the condensation. Extension to higher-order schemes is an ongoing work.

2. The equivalence. In this section, we first define the spaces $\mathbf{V}(\mathcal{E}_h)$ and $\Phi(\mathcal{T}_h)$ and then establish the equivalence.

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 . The 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 fix its orientation, i.e. the order of K and L . For a Dirichlet boundary side σ , the support of \mathbf{v}_σ 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 $\mathcal{C}_V^{\text{el}}$ denote the set of elements from the cluster which contain exactly

one side from $\mathcal{G}_{\mathcal{C}_V}$, which we denote by δ_K for $K \in \mathcal{C}_V^{\text{el}}$. 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 $|\mathcal{C}_V^{\text{el}}| = |\mathcal{G}_{\mathcal{C}_V}|$, where we denote by $|A|$ the cardinality of a set A . We are not interested in the trivial cases where $\mathcal{F}_{\mathcal{C}_V} = \emptyset$ or $\mathcal{G}_{\mathcal{C}_V} = \emptyset$.

Let us now consider the equations (1.2a) for the basis functions \mathbf{v}_γ , $\gamma \in \mathcal{F}_{\mathcal{C}_V}$. We remark that the support of all \mathbf{v}_γ , $\gamma \in \mathcal{F}_{\mathcal{C}_V}$, 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, using also that $p_h|_K = P_K$,

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

i.e. $|\mathcal{F}_{\mathcal{C}_V}|$ equations for the $|\mathcal{E}_{\mathcal{C}_V}|$ unknown fluxes, where we consider the scalar unknowns P_K , $K \in \mathcal{C}_V$, as parameters. The remaining $|\mathcal{G}_{\mathcal{C}_V}|$ equations are given by (1.2b) for all ϕ_K , $K \in \mathcal{C}_V^{\text{el}}$,

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

The matrix problem associated with the set of equations (2.1)–(2.2) has the form

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

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 $(\nabla \cdot \mathbf{v}_\sigma, 1)_K = \pm 1$ for $\sigma \in \mathcal{E}_K$). Hence we can eliminate the $U_V^{\mathcal{G}}$ unknowns and come to

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

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

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

a *local condensation matrix* associated with the vertex V . We have:

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 2.1 (Comparison with a classical multi-point FVM). *In “classical” multi-point finite volume schemes, cf. [1, 6, 8, 9], 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 scheme proposed in [1].*

Let $V \in \mathcal{V}_h$. Let us define a mapping $\Psi_V : \mathbb{R}^{|\mathcal{F}_{C_V}|} \rightarrow \mathbb{R}^{|\mathcal{E}_h|}$, extending a vector $U_V^{\mathcal{F}} = \{U_\sigma\}_{\sigma \in \mathcal{F}_{C_V}}$ of values associated with the sides from \mathcal{F}_{C_V} to a vector of values associated with all non-Neumann sides \mathcal{E}_h by $[\Psi_V(U_V^{\mathcal{F}})]_\sigma := U_\sigma$ if $\sigma \in \mathcal{F}_{C_V}$, $[\Psi_V(U_V^{\mathcal{F}})]_\sigma := 0$ if $\sigma \notin \mathcal{F}_{C_V}$. Since there is no possibility of confusion, we keep the same notation also for a mapping $\mathbb{R}^{|\mathcal{F}_{C_V}| \times |\mathcal{F}_{C_V}|} \rightarrow \mathbb{R}^{|\mathcal{E}_h| \times |\mathcal{E}_h|}$, extending a local matrix \mathbb{M}_V to a full-size one by zeros by $[\Psi_V(\mathbb{M}_V)]_{\sigma,\gamma} := (\mathbb{M}_V)_{\sigma,\gamma}$ if $\sigma \in \mathcal{F}_{C_V}$ and $\gamma \in \mathcal{F}_{C_V}$, $[\Psi_V(\mathbb{M}_V)]_{\sigma,\gamma} := 0$ if $\sigma \notin \mathcal{F}_{C_V}$ or $\gamma \notin \mathcal{F}_{C_V}$. We finally in the same fashion define a mapping $\Theta_V : \mathbb{R}^{|\mathcal{F}_{C_V}| \times |\mathcal{C}_V^{\text{el}}|} \rightarrow \mathbb{R}^{|\mathcal{E}_h| \times |\mathcal{T}_h|}$, filling a full-size representation of a matrix \mathbb{J}_V by zeros on the rows associated with the sides which are not from \mathcal{F}_{C_V} and on the columns associated with the elements which are not from $\mathcal{C}_V^{\text{el}}$, $[\Theta_V(\mathbb{J}_V)]_{\sigma,K} := (\mathbb{J}_V)_{\sigma,K}$ if $\sigma \in \mathcal{F}_{C_V}$ and $K \in \mathcal{C}_V^{\text{el}}$, $[\Theta_V(\mathbb{J}_V)]_{\sigma,K} := 0$ if $\sigma \notin \mathcal{F}_{C_V}$ or $K \notin \mathcal{C}_V^{\text{el}}$. Let the local condensation matrices \mathbb{M}_V be invertible for all $V \in \mathcal{V}_h$. Let us finally 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

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

We now notice that $\sum_{V \in \mathcal{V}_h} \frac{1}{d} \Psi_V(U_V^{\mathcal{F}}) = U$, which expresses that if we go through all $V \in \mathcal{V}_h$ and observe the sides in the sets \mathcal{F}_{C_V} , each $\sigma \in \mathcal{E}_h$ appears just d -times. Hence we can sum (2.6) over all vertices and divide it by d to find that

$$U = \tilde{\mathbb{A}}^{-1}(F - \mathbb{B}^t P) - \mathbb{J}G, \quad (2.7)$$

where

$$\tilde{\mathbb{A}}^{-1} := \frac{1}{d} \sum_{V \in \mathcal{V}_h} \Psi_V(\mathbb{M}_V^{-1}), \quad \mathbb{J} := \frac{1}{d} \sum_{V \in \mathcal{V}_h} \Theta_V(\mathbb{J}_V). \quad (2.8)$$

Finally, inserting the expression (2.7) into the second equation of (1.3), we obtain a system for only the scalar unknowns

$$-\mathbb{B} \tilde{\mathbb{A}}^{-1} \mathbb{B}^t P = G - \mathbb{B} \tilde{\mathbb{A}}^{-1} F + \mathbb{B} \mathbb{J} G. \quad (2.9)$$

REMARK 2.2 (Comparison with direct elimination of the fluxes). *From (1.3), $U = \mathbb{A}^{-1}(F - \mathbb{B}^t P)$. There are two essential differences from (2.7). 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 price of the inverse of $|\mathcal{V}_h|$ local matrices, whereas obtaining \mathbb{A}^{-1} is in general very expensive.*

REMARK 2.3 (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}_{C_V} and elements from $\mathcal{C}_V^{\text{el}}$. Hence obtaining (2.9) from (1.3) is immediate.*

It appears that in some particular cases, the matrix \mathbb{M}_V is not invertible. 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 a byproduct of Lemma 3.5. If a given local condensation matrix is not invertible, one can resort to a wider set of elements than the clusters defined above.

3. Properties of the condensed mixed finite element scheme. We study in this section the properties of the system matrix of the condensed scheme.

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. \square

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

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

Let $V \in \mathcal{V}_h$. We define a mapping $\Pi_V : \mathbb{R}^{|\mathcal{E}_h|} \rightarrow \mathbb{R}^{|\mathcal{F}_{C_V}|}$, restricting a vector of values associated with all non-Neumann sides to a vector of values associated with the sides from \mathcal{F}_{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 nonnegative and at least d of them are positive. \square

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 (2.8), is symmetric. Finally, if $\tilde{\mathbb{A}}^{-1}$ is symmetric, $\mathbb{B}\tilde{\mathbb{A}}^{-1}\mathbb{B}^t$ is symmetric as well. \square

3.2. Properties of the local condensation matrices. Let $\mathbf{V}(\mathcal{E}_{C_V})$ be the space spanned by the basis functions \mathbf{v}_σ associated with the non-Neumann sides \mathcal{E}_{C_V} of the cluster C_V and $\mathbf{V}(\mathcal{F}_{C_V})$ its restriction with the basis functions \mathbf{v}_σ associated with the sides from \mathcal{F}_{C_V} . Let further $\mathbf{V}(\text{div}, \mathcal{E}_{C_V})$ be the subspace of $\mathbf{V}(\mathcal{E}_{C_V})$ of the functions whose divergence is equal to 0 on all elements $K \in C_V^{\text{el}}$. The space $\mathbf{V}(\text{div}, \mathcal{E}_{C_V})$ is spanned by basis functions \mathbf{p}_σ associated with the sides from \mathcal{F}_{C_V} , which have the same support as the basis functions \mathbf{v}_σ and whose fluxes across the associated sides equal those of \mathbf{v}_σ (this namely fixes their orientation). In particular, for $K \in C_V^{\text{el}}$ and $\sigma \in \mathcal{E}_K \cap \mathcal{F}_{C_V}$, $\mathbf{p}_\sigma|_K = \mathbf{v}_\sigma - \frac{(\nabla \cdot \mathbf{v}_\sigma, 1)_K}{(\nabla \cdot \mathbf{v}_{\delta_K}, 1)_K} \mathbf{v}_{\delta_K}$. 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 which is not included in the sides σ and δ_K . For $K \in C_V \setminus C_V^{\text{el}}$, $\mathbf{p}_\sigma|_K = \mathbf{v}_\sigma|_K$. We refer to [15, Chapter 3] for the proofs of the following assertions.

LEMMA 3.4 (Form of the local condensation matrices). *The local condensation matrix \mathbb{M}_V for $V \in \mathcal{V}_h$ is given by*

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

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

LEMMA 3.5 (Positive definiteness of the local condensation matrices). *Let the matrices $\mathbb{E}_{V, K} \in \mathbb{R}^{|\mathcal{E}_K \cap \mathcal{F}_{C_V}| \times |\mathcal{E}_K \cap \mathcal{F}_{C_V}|}$ given by*

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

where \mathbf{p}_σ and \mathbf{v}_σ , $\sigma \in \mathcal{E}_K \cap \mathcal{F}_{C_V}$, are the basis functions of the spaces $\mathbf{V}(\text{div}, \mathcal{E}_{C_V})$ and $\mathbf{V}(\mathcal{F}_{C_V})$, 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 $V \in \mathcal{V}_h$.

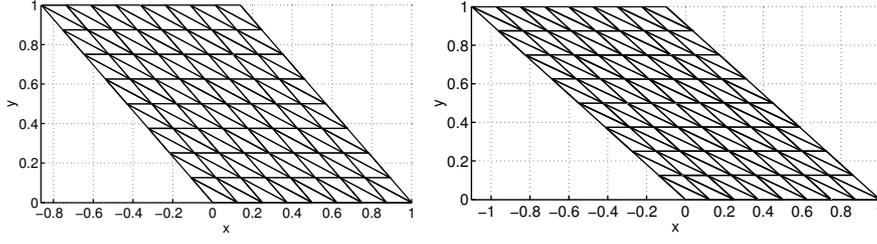


FIG. 3.1. Theoretical (left) and experimental (right) limit mesh for the positive definiteness of the system matrix for a deformed square and $\mathbf{S} = Id$

EXAMPLE 3.1 (Positive definiteness for a triangulation of a deformed square). Let $\mathbf{S} = Id$ and let $\Omega = (0, 1) \times (0, 1)$. Let us deform Ω and \mathcal{T}_h given by right-angled triangles by shifting horizontally the upper edge. Then the theoretical and experimental limits for the positive definiteness of the system matrix are given in Figure 3.1.

LEMMA 3.6 (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 $V \in \mathcal{V}_h$.

REMARK 3.1 (Equilateral simplices and a piecewise constant scalar diffusion tensor). Let \mathcal{T}_h consist of equilateral simplices, let \mathbf{S} be piecewise constant and scalar, and let $\Gamma_N = \emptyset$. Then the local condensation matrices are diagonal and consequently 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.

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–reaction–diffusion problems. We consider in particular the problem

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

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

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

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

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

where β , ψ , and F are nonlinear functions, \mathbf{S} is a bounded, symmetric, and uniformly positive definite tensor, \mathbf{w} is a velocity field, and q represents the 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 $0 = t_0 < \dots < t_n < \dots < t_N = T$ and define $\Delta t_n := t_n - t_{n-1}$, $n \in \{1, 2, \dots, N\}$, 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) consists in finding on each time level t_n , $n \in \{1, 2, \dots, N\}$, 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

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

$$\begin{aligned} \left(\frac{\beta(p_h^n) - \beta(p_h^{n-1})}{\Delta t_n}, \phi_h \right)_\Omega + (\nabla \cdot \mathbf{u}_{0,h}^n, \phi_h)_\Omega + (F(p_h^n), \phi_h)_\Omega &= (q, \phi_h)_\Omega \\ &- (\nabla \cdot \tilde{\mathbf{u}}^n, \phi_h)_\Omega \quad \forall \phi_h \in \Phi(\mathcal{T}_h), \end{aligned} \quad (4.2b)$$

where

$$\begin{aligned} \mathbf{S}^{-n} &:= \frac{1}{\Delta t_n} \int_{t_{n-1}}^{t_n} \mathbf{S}^{-1}(\cdot, t) dt, & \mathbf{w}^n &:= \frac{1}{\Delta t_n} \int_{t_{n-1}}^{t_n} \mathbf{w}(\cdot, t) dt, \\ p_D^n &:= \frac{1}{\Delta t_n} \int_{t_{n-1}}^{t_n} p_D(\cdot, t) dt, & \tilde{\mathbf{u}}^n &:= \frac{1}{\Delta t_n} \int_{t_{n-1}}^{t_n} \tilde{\mathbf{u}}(\cdot, t) dt \quad n \in \{1, 2, \dots, N\}. \end{aligned}$$

We now notice that the terms where the unknown discrete velocity function $\mathbf{u}_{0,h}^n$ appears are exactly the same as in the linear elliptic diffusion case, see (1.2a)–(1.2b). Hence one can eliminate $\mathbf{u}_{0,h}^n$ on each discrete time level as in Section 2. The system (4.2a)–(4.2b), linearized by the Newton method, can be written in the matrix form as

$$\begin{pmatrix} \mathbb{A} & \mathbb{C} \\ \mathbb{B} & \mathbb{D} \end{pmatrix} \begin{pmatrix} U \\ P \end{pmatrix} = \begin{pmatrix} F \\ G \end{pmatrix} \quad (4.3)$$

and the final system for the scalar unknowns P only writes in the form

$$(-\mathbb{B}\tilde{\mathbb{A}}^{-1}\mathbb{C} + \mathbb{B}\mathbb{J}\mathbb{D} + \mathbb{D})P = G - \mathbb{B}\tilde{\mathbb{A}}^{-1}F + \mathbb{B}\mathbb{J}G. \quad (4.4)$$

This transcription enables in particular a straightforward implementation of the condensation in any mixed finite element code. Moreover, if the diffusion tensor \mathbf{S} is constant with respect to time, \mathbb{A}, \mathbb{B} do not change and hence the assemblage of $\tilde{\mathbb{A}}^{-1}$ and \mathbb{J} can be done only once before the start of the calculation.

We now finally turn to the upwind-mixed lowest-order Raviart–Thomas method, cf. [7]. 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 Neumann boundary condition (4.1e) as $\mathbf{r} \cdot \mathbf{n} = v_N$ 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 $\tilde{\mathbf{r}}^n := \frac{1}{\Delta t_n} \int_{t_{n-1}}^{t_n} \tilde{\mathbf{r}}(\cdot, t) dt$, $n \in \{1, 2, \dots, N\}$. The fully implicit upwind-mixed finite element method then reads: on each time level t_n , $n \in \{1, 2, \dots, N\}$, 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

$$\begin{aligned} (\mathbf{S}^{-n}\mathbf{r}_{0,h}^n, \mathbf{v}_h)_\Omega - (\nabla \cdot \mathbf{v}_h, p_h^n)_\Omega &= -\langle \mathbf{v}_h \cdot \mathbf{n}, p_D^n \rangle_{\partial\Omega} \\ &\quad - (\mathbf{S}^{-n}\tilde{\mathbf{r}}^n, \mathbf{v}_h)_\Omega \quad \forall \mathbf{v}_h \in \mathbf{V}(\mathcal{E}_h), \end{aligned} \quad (4.6a)$$

$$\begin{aligned} \left(\frac{\beta(P_K^n) - \beta(P_K^{n-1})}{\Delta t_n}, \phi_K \right)_K + (\nabla \cdot \mathbf{r}_{0,h}^n, \phi_K)_K + \sum_{\sigma \in \mathcal{E}_K} \psi(\widehat{p}_\sigma^n) \mathbf{w}_{K,\sigma}^n + (F(P_K^n), \phi_K)_K \\ = (q, \phi_K)_K - (\nabla \cdot \tilde{\mathbf{r}}^n, \phi_K)_K \quad \forall K \in \mathcal{T}_h, \end{aligned} \quad (4.6b)$$

where $\mathbf{w}_{K,\sigma}^n = \langle \mathbf{w}^n \cdot \mathbf{n}, 1 \rangle_\sigma$ and \widehat{p}_σ^n is the upwind value defined by P_K^n if $\mathbf{w}_{K,\sigma}^n \geq 0$ and by P_L^n otherwise for σ an interior side between the elements K and L , by P_K^n if $\mathbf{w}_{K,\sigma}^n \geq 0$ and by $\langle p_D^n, 1 \rangle_\sigma / |\sigma|$ otherwise for σ a Dirichlet boundary side, and by P_K^n for σ a Neumann boundary side. The linearization of the system (4.6a)–(4.6b) 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 (4.4). The final matrix has however in this case a wider stencil due to the upstream weighting.

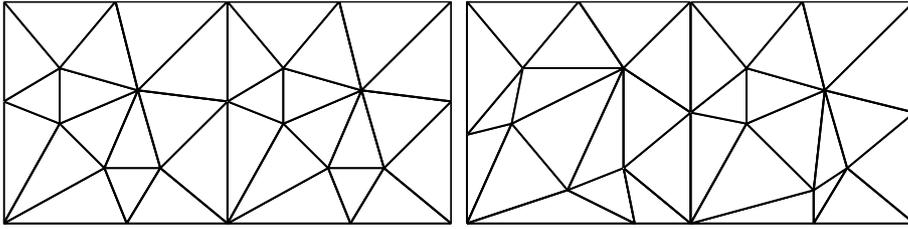


FIG. 5.1. Initial meshes A (left) and B (right)

TABLE 5.1
Condensed mixed method, problem (5.1), left part of the mesh A

Ref.	Unkn.	St.	Cond.	Bi-CGS	Iter.
3	1024	14	721	0.20	76.5
4	4096	14	2882	1.43	147.5
5	16384	14	11523	12.55	295.5
6	65536	14	46093	117.58	555.5

TABLE 5.2
Hybridized mixed method, problem (5.1), left part of the mesh A

Ref.	Unkn.	St.	Cond.	Bi-CGS	Iter.	CG	Iter.
3	1504	5	1397	0.31	118.0	0.22	157
4	6080	5	5616	2.43	230.5	1.75	316
5	24448	5	22499	23.40	449.5	16.87	623
6	98048	5	89995	227.04	864.0	162.09	1226

5. Numerical experiments. We give the results of several numerical experiments in two space dimensions in this section. All the computations were done in a C++ code 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 power of 10^{-16} . All the matrix operations were done with the help of MATLAB 6.1.

5.1. An elliptic problem. For $\Omega = (0, 1) \times (0, 1)$, we consider the problem

$$-\Delta p = -2e^x e^y \quad (5.1)$$

with a Dirichlet boundary condition given by the exact solution $p(x, y) = e^x e^y$ on regular refinements of the initial mesh given in the left square of the mesh A from Figure 5.1. We compare the computational cost of the condensation and that of the hybridization onto Lagrange multipliers associated with the edges. In both cases the system matrices are positive definite but they are symmetric only in the latter case. In Tables 5.1 and 5.2, we give the number of unknowns, the system matrices stencil and 2-norm condition number, and the CPU time in seconds and the number of iterations of the Bi-CGStab method necessary to decrease the 2-norm relative residual below $1e-10$, using a zero start vector. For the hybridization, we consider also the CG method. In this case, the condensation CPU time is about 1.35-times shorter.

5.1.1. A convection–reaction–diffusion problem. For $\Omega = (0, 2) \times (0, 1)$ and a time interval $(0, 1)$, we consider a nonlinear convection–reaction–diffusion problem which involves discontinuous coefficients and an inhomogeneous and anisotropic

TABLE 5.3
Condensed mixed method, problem (5.2), coefficients (5.3), mesh B

Ref.	Unkn.	St.	Cond.	Bi-CGS	Iter.	CPU	ILU	PBi-CGS	Iter.
3	2048	14	4824	1.47	322.5	0.12	0.07	0.05	3.5
4	8192	14	12523	8.66	474.5	0.88	0.56	0.32	5.0
5	32768	14	31368	61.53	787.5	7.47	5.46	2.01	5.5

TABLE 5.4
Standard mixed method, problem (5.2), coefficients (5.3), mesh B

Ref.	Unkn.	St.	Cond.	Bi-CGS	Iter.	CPU	ILU	PBi-CGS	Iter.
3	3120	5	131073	12.12	2419.5	0.41	0.18	0.23	3.0
4	12384	5	250923	103.42	5390.5	3.06	1.32	1.74	3.5
5	49344	5	586375	617.26	7145.5	29.88	14.96	14.92	4.0

diffusion tensor,

$$\frac{\partial(p + p^\alpha)}{\partial t} - \nabla \cdot (\mathbf{S}\nabla p) + \nabla \cdot (p\mathbf{w}) + \alpha p^\alpha = 0 \quad (5.2)$$

with $\alpha = 0.5$ and

$$\mathbf{S} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ for } x < 1, \quad \mathbf{S} = \begin{pmatrix} 8 & -7 \\ -7 & 20 \end{pmatrix} \text{ for } x > 2, \\ \mathbf{w} = (3, 0) \text{ for } x < 1, \quad \mathbf{w} = (3, 12) \text{ for } x > 2. \quad (5.3)$$

Initial and Dirichlet boundary conditions are given by the exact solution $p(x, y, t) = e^x e^y e^{-t}/e^3$. 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 $P = \mathbb{D}^{-1}(G - \mathbb{B}U)$. We compare the properties of the linear systems on the first time and Newton linearization steps in Tables 5.3 and 5.4, considering regular refinements of the mesh B. We consider the Bi-CGStab method without any preconditioning first and then use the incomplete LU factorizations as preconditioners (in both cases the system matrices are nonsymmetric but positive definite). The drop tolerance is chosen in a such way that the sum of times in seconds (CPU) for the preconditioning and the solution of the preconditioned system was minimal. The CPU time of the condensation is in this case about 4-times shorter.

5.1.2. A convection–reaction–diffusion problem and the upwind-mixed method. We consider here once more the problem (5.2), this time with coefficients

$$\mathbf{S} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ in } \Omega, \quad \mathbf{w} = (3, 0) \text{ in } \Omega \quad (5.4)$$

and mesh A. We employ the upwind-mixed finite element method (4.6a)–(4.6b) and the corresponding condensed version. For the upwind-mixed method, we cannot easily eliminate the scalar unknowns P , as the matrix \mathbb{D} from (4.3) is in this case not diagonal. The direct application of the Bi-CGStab method to the system (4.3) does not lead to satisfactory results, cf. Table 5.6. A suitable solution approach however seems to be to first perform the column minimum degree permutation and then use the incomplete LU factorization for preconditioning; we report in Table 5.6 the separate times in seconds as well as their sum (CPU). In the given case, the condensation reduces the CPU time for one linear system by a factor better than 3.

TABLE 5.5
Condensed upwind-mixed method, problem (5.2), coefficients (5.4), mesh A

Ref.	Unkn.	St.	Cond.	Bi-CGS	Iter.	CPU	ILU	PBi-CGS	Iter.
3	2048	19	318	0.46	88.0	0.11	0.06	0.05	3.0
4	8192	19	777	2.99	138.5	0.68	0.36	0.32	5.0
5	32768	19	1792	18.86	210.5	4.89	2.87	2.02	7.5

TABLE 5.6
Standard upwind-mixed method, problem (5.2), coefficients (5.4), mesh A

Ref.	Unkn.	St.	Cond.	Bi-CGS	Iter.	CPU	Per.	ILU	PBi-CGS	Iter.
3	5168	7	67	13.01	1540.5	0.48	0.03	0.15	0.30	3.5
4	20576	7	168	124.06	3561.5	2.83	0.32	0.98	1.53	4.0
5	82112	7	393	3233.05	16763.5	15.70	1.35	4.92	9.43	6.0

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), pp. 1700–1716.
- [2] 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), pp. 404–425.
- [3] D. N. ARNOLD AND F. BREZZI, *Mixed and nonconforming finite element methods: Implementation, postprocessing and error estimates*, RAIRO Modél. Math. Anal. Numér., 19 (1985), pp. 7–32.
- [4] J. BARANGER, J.-F. MAÎTRE, AND F. OUDIN, *Connection between finite volume and mixed finite element methods*, M2AN Math. Model. Numer. Anal., 30 (1996), pp. 445–465.
- [5] F. BREZZI AND M. FORTIN, *Mixed and Hybrid Finite Element Methods*, Springer, New York, 1991.
- [6] Y. COUDIÈRE, J.-P. VILA, AND PH. VILLEDIEU, *Convergence rate of a finite volume scheme for a two dimensional convection–diffusion problem*, M2AN Math. Model. Numer. Anal., 33 (1999), pp. 493–516.
- [7] C. DAWSON, *Analysis of an upwind-mixed finite element method for nonlinear contaminant transport equations*, SIAM J. Numer. Anal., 35 (1998), pp. 1709–1724.
- [8] R. EYMARD, T. GALLOUËT, AND R. HERBIN, *A finite volume scheme for anisotropic diffusion problems*, submitted to C. R. Math. Acad. Sci. Paris, 2004.
- [9] I. FAILLE, *A control volume method to solve an elliptic equation on a two-dimensional irregular mesh*, Comput. Methods Appl. Mech. Engrg., 100 (1992), pp. 275–290.
- [10] J. C. NÉDÉLEC, *Mixed finite elements in \mathbb{R}^3* , Numer. Math., 35 (1980), pp. 315–341.
- [11] 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, I. Galligani and E. Magenes, eds., Lecture Notes in Math. 606, pp. 292–315, Springer, Berlin, 1977.
- [12] J. E. ROBERTS AND J.-M. THOMAS, *Mixed and hybrid methods*, in Handbook of Numerical Analysis, vol. 2, P. G. Ciarlet and J. L. Lions, eds., pp. 523–639, Elsevier, Amsterdam, 1991.
- [13] 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., pp. 35–106, SIAM, Philadelphia, 1983.
- [14] 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), pp. 365–388.
- [15] M. VOHRALÍK, *Numerical methods for elliptic and nonlinear parabolic equations. Application to flow problems in porous and fractured media*, PhD. Thesis, Université de Paris-Sud & Czech Technical University in Prague, 2004.