From face to element unknowns by local static condensation with application to nonconforming finite elements[*]

Martin Vohralík[1] and Barbara Wohlmuth[2]

[1]UPMC Univ. Paris 06, UMR 7598, Laboratoire Jacques-Louis Lions, 75005, Paris, France
&
CNRS, UMR 7598, Laboratoire Jacques-Louis Lions, 75005, Paris, France
e-mail: vohralik@ann.jussieu.fr

[2] Fakultät für Mathematik, Lehrstuhl für Numerische Mathematik
Boltzmannstrasse 3, 85748 Garching bei München, Germany
e-mail: wohlmuth@ma.tum.de

Abstract

We derive in this paper a new local static condensation strategy which allows to reduce significantly the number of unknowns in algebraic systems arising in discretization of partial differential equations. We apply it to the discretization of a model linear elliptic diffusion and a model nonlinear parabolic advection–diffusion–reaction problem by Crouzeix–Raviart nonconforming finite elements. Herein, the unknowns, originally associated with the mesh faces, can be reduced to new unknowns associated with the mesh elements. The resulting matrices are sparse, with possibly only four nonzero entries per row in two space dimensions, positive definite in dependence on the mesh geometry and the diffusion–dispersion tensor, but in general nonsymmetric. Our approach consists in introducing new element unknowns, the identification of suitable local vertex-based subproblems, and the inversion of the corresponding local matrices. We give sufficient conditions for the well-posedness of the local problems, as well as for the resulting global one. In addition, we provide a geometrical interpretation which suggests how to influence the form of the local and global matrices depending on the local mesh and data. We finally present an abstract generalization allowing for a further reduction of the number of unknowns, typically to one unknown per a set of mesh elements. We conclude by numerical experiments which show that the condition number of the resulting matrices is robust with respect to the mesh anisotropies and the diffusion tensor inhomogeneities.

Key words: local static condensation, nonconforming finite element method, diffusion equation, nonlinear parabolic advection–diffusion–reaction equation

1 Introduction

Let $\mathbb{Z}^{i}$ and $\mathbb{Z}^{b}$ be two given matrices and let E be a given right-hand side vector. We consider in this paper the following problem: find Λ, $\Lambda := (\Lambda^{i}, \Lambda^{b})^{t}$, such that

$$\begin{pmatrix} \mathbb{Z}^{i} & \mathbb{Z}^{b} \\ 0 & \mathbb{I} \end{pmatrix} \begin{pmatrix} \Lambda^{i} \\ \Lambda^{b} \end{pmatrix} = \begin{pmatrix} E \\ 0 \end{pmatrix}. \tag{1.1}$$

Here $\mathbb{I}$ stands for the identity matrix. Note that it follows from (1.1) that $\Lambda^{\mathrm{b}} = 0$, so that (1.1) can be equivalently rewritten as: find Λ^{i} such that

$$\mathbb{Z}^{\mathrm{i}}\Lambda^{\mathrm{i}} = E. \tag{1.2}$$

System (1.1) typically results in the discretization of elliptic or parabolic problems by the Crouzeix–Raviart nonconforming finite element method (cf. Crouzeix and Raviart [6]) or by the mixed finite element method (cf. Raviart–Thomas [10] and Arnold and Brezzi [3]), see Sections 3 and 4 for details. Therein, $\mathbb{Z}^{\mathrm{i}}$ is a sparse (symmetric) positive definite matrix.

The purpose of this paper is to devise a general principle allowing to *reduce equivalently* system (1.1) to a system

$$\mathbb{S}P = H, \tag{1.3}$$

with a sparse and easily computable matrix $\mathbb{S}$ and *much fewer unknowns P*. Let Ω be a computational domain and $\mathcal{T}_h$ its simplicial mesh. We suppose that the unknowns Λ are associated with the mesh faces, whereas the new unknowns P are associated with the mesh elements.

In Section 2, we introduce an *abstract* and *algebraic* principle for reducing (1.1) to (1.3). We introduce an *arbitrary matrix* $\mathbb{N}$ and augment (1.1) with the relation $\mathbb{N}\Lambda = P$. We then identify suitable local subproblems of the augmented system, invert the corresponding local matrices, obtain local expressions of the unknowns Λ^{i} in terms of the new unknowns P, and finally identify the matrix $\mathbb{S}$ and the right-hand side vector H of (1.3). Under Assumption 2.2 below, we also prove the well-posedness of (1.3), the equivalence of (1.3) with (1.1), and characterize the stencil (maximal number of nonzero entries per each matrix row) of $\mathbb{S}$. This approach generalizes those obtained in the framework of the mixed finite element method in Younès *et al.* [19, 18], Chavent *et al.* [4], and [14, 17]. In contrast to [19, 4, 18], our starting point is purely algebraic, and no geometrical mesh or discretization information is used. Moreover, different choices of $\mathbb{N}$ are possible and a family of reformulations can be obtained.

Enlightening the abstract algebraic approach of Section 2, we provide in Section 3 its specification for the discretization of a model linear elliptic diffusion problem by the Crouzeix–Raviart nonconforming finite element method. We also present its geometric interpretation. Here, the new element values are the values of the Crouzeix–Raviart approximation in points associated with the elements (not necessarily inside the elements). We give in Section 3 sufficient conditions in terms of the diffusion tensor $\underline{\mathbf{S}}$ and of the geometry of the mesh $\mathcal{T}_h$ for Assumption 2.2 to hold. We also investigate how the properties of the local problems and of the matrix $\mathbb{S}$ can be influenced by the choice of the element points. In Section 4, we then present similar developments for a model nonlinear parabolic advection–diffusion–reaction problem discretized by the Crouzeix–Raviart method.

In Section 5, we report results of several numerical experiments. The matrices $\mathbb{S}$ resulting from our approach are in general positive definite but nonsymmetric. Their condition number is in our numerical examples insensitive to the anisotropies of the mesh $\mathcal{T}_h$ and inhomogeneities of the tensor $\underline{\mathbf{S}}$ for linear diffusion problems. We demonstrate that CPU gains for both direct and iterative solvers in range 1.5-times to 3-times, 30-times in particular situations, can be achieved. A concluding discussion is given in Section 6.

We finish the paper by Appendix A which gives a generalization of the approach of Section 2, weakens Assumption 2.2, and enables a further reduction of the number of unknowns.

2 Static condensation from edges to elements

We introduce in this section our basic static condensation principle, which enables us to rewrite (1.1) equivalently as (1.3), reducing the number of unknowns from mesh faces to mesh elements.
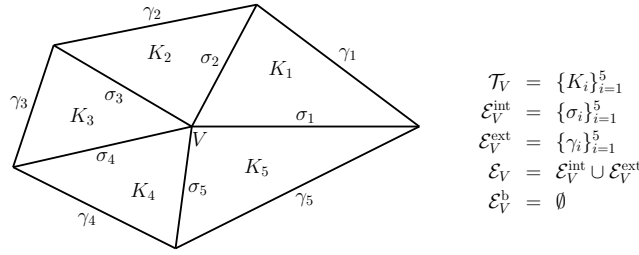
$$
\begin{aligned}
\mathcal{T}_V &= \{K_i\}_{i=1}^5 \\
\mathcal{E}_V^{\mathrm{int}} &= \{\sigma_i\}_{i=1}^5 \\
\mathcal{E}_V^{\mathrm{ext}} &= \{\gamma_i\}_{i=1}^5 \\
\mathcal{E}_V &= \mathcal{E}_V^{\mathrm{int}} \cup \mathcal{E}_V^{\mathrm{ext}} \\
\mathcal{E}_V^{\mathrm{b}} &= \emptyset
\end{aligned}
$$

Figure 1: An example of a patch $\mathcal{T}_V$ around a vertex V in the interior of Ω

$$
\begin{aligned}
\mathcal{T}_V &= \{K_i\}_{i=1}^5 \\
\mathcal{E}_V^{\mathrm{int}} &= \{\sigma_i\}_{i=1}^5 \\
\mathcal{E}_V^{\mathrm{ext}} &= \{\gamma_i\}_{i=1}^5 \\
\mathcal{E}_V &= \mathcal{E}_V^{\mathrm{int}} \cup \mathcal{E}_V^{\mathrm{ext}} \\
\mathcal{E}_V^{\mathrm{b}} &= \emptyset
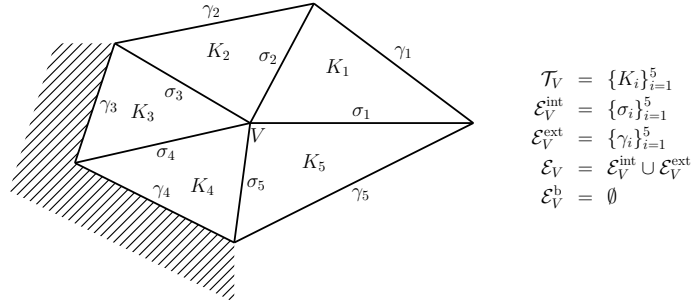\end{aligned}
$$

Figure 2: An example of a patch $\mathcal{T}_V$ around a vertex V close to the boundary of Ω

2.1 The domain and its mesh

Let $\Omega \subset \mathbb{R}^d$, $d \geq 2$, be a polygonal (polyhedral) domain and let $\mathcal{T}_h$ be a matching (containing no hanging nodes) simplicial mesh of Ω in the sense of Ciarlet [5]. We denote by $\mathcal{E}_h$ the set of all $(d-1)$-dimensional faces of $\mathcal{T}_h$. We divide $\mathcal{E}_h$ into interior faces $\mathcal{E}_h^{\mathrm{i}}$ and boundary faces $\mathcal{E}_h^{\mathrm{b}}$. For $\sigma \in \mathcal{E}_h$, let $\mathbf{x}_\sigma$ stand for the barycenter of the face σ. We next denote by $\mathcal{V}_h$ the set of vertices of $\mathcal{T}_h$. For a given vertex $V \in \mathcal{V}_h$, we shall denote by $\mathcal{T}_V$ the patch of the elements of $\mathcal{T}_h$ which share V, by $\mathcal{E}_V^{\mathrm{int}} \subset \mathcal{E}_h^{\mathrm{i}}$ the interior faces of $\mathcal{T}_V$, and by $\mathcal{E}_V^{\mathrm{ext}}$ the faces of $\mathcal{T}_V$ not having V as vertex. We set $\mathcal{E}_V := \mathcal{E}_V^{\mathrm{int}} \cup \mathcal{E}_V^{\mathrm{ext}}$ and we also denote by $\mathcal{E}_V^{\mathrm{b}}$ the faces of $\mathcal{T}_V$ which lie on the boundary $\partial\Omega$ and not in $\mathcal{E}_V$. We refer to Figures 1–3 for an illustration in two space dimensions. Let $K \in \mathcal{T}_h$. By $\mathcal{E}_K$, we denote the set of all faces of K and by $\mathcal{E}_K^{\mathrm{i}}$ the set of such faces of $\mathcal{E}_K$ which lie in $\mathcal{E}_h^{\mathrm{i}}$. Let V be a vertex of K. We will also employ the notation $\mathcal{E}_{V,K}$ for the faces of K which have V as vertex. Finally, $\mathbf{x}_K$ denotes the barycenter of K. The symbol $|S|$ stands for the cardinality (the number of elements) of a set S.

$$
\begin{aligned}
\mathcal{T}_V &= \{K_i\}_{i=1}^2 \\
\mathcal{E}_V^{\mathrm{int}} &= \sigma_2 \\
\mathcal{E}_V^{\mathrm{ext}} &= \{\gamma_i\}_{i=1}^2 \\
\mathcal{E}_V &= \mathcal{E}_V^{\mathrm{int}} \cup \mathcal{E}_V^{\mathrm{ext}} \\
\mathcal{E}_V^{\mathrm{b}} &= \{\sigma_1, \sigma_3\}
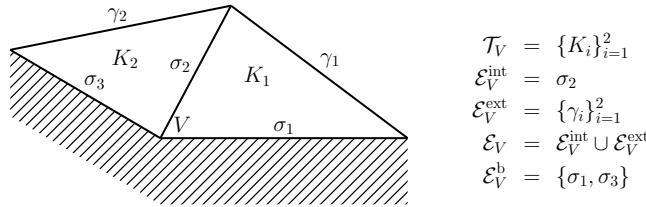\end{aligned}
$$

Figure 3: An example of a patch $\mathcal{T}_V$ around a vertex V on the boundary of Ω

3

2.2 Augmented problem setting

Let $\mathbb{Z}^{\mathrm{i}} \in \mathbb{R}^{|\mathcal{E}_h^{\mathrm{i}}| \times |\mathcal{E}_h^{\mathrm{i}}|}$, $\mathbb{Z}^{\mathrm{b}} \in \mathbb{R}^{|\mathcal{E}_h^{\mathrm{i}}| \times |\mathcal{E}_h^{\mathrm{b}}|}$ be given matrices. Both have the number of rows equal to the number of mesh interior faces; the number of columns of $\mathbb{Z}^{\mathrm{i}}$ is given by the number of mesh interior faces, whereas that of $\mathbb{Z}^{\mathrm{b}}$ by that of mesh boundary faces. Let a right-hand side vector $E \in \mathbb{R}^{|\mathcal{E}_h^{\mathrm{i}}|}$ be also given. We consider the following problem: find $\Lambda \in \mathbb{R}^{|\mathcal{E}_h|}$, $\Lambda = \{\Lambda_\sigma\}_{\sigma \in \mathcal{E}_h} = (\Lambda^{\mathrm{i}}, \Lambda^{\mathrm{b}})^t$, such that (1.1) holds. Herein, we merely suppose that (1.1) is well posed, i.e., that the system matrix of (1.1) is nonsingular, and that on a row associated with a face $\sigma \in \mathcal{E}_h^{\mathrm{i}}$, the only nonzero entries of $(\mathbb{Z}^{\mathrm{i}}, \mathbb{Z}^{\mathrm{b}})$ lie on columns associated with faces $\gamma \in \mathcal{E}_h$ such that σ and γ belong to the same simplex.

Let $\mathbb{N} \in \mathbb{R}^{|\mathcal{T}_h| \times |\mathcal{E}_h|}$. We suppose that on a row of $\mathbb{N}$ associated with an element $K \in \mathcal{T}_h$, the only nonzero entries are on columns associated with the faces $\mathcal{E}_K$ of K; apart from this assumption, the matrix $\mathbb{N}$ is *arbitrary*. This assumption will ensure locality of our approach and sparsity of the final matrix $\mathbb{S}$. Introduce one new unknown P_K for each mesh element $K \in \mathcal{T}_h$. Let P be the corresponding algebraic vector, $P = \{P_K\}_{K \in \mathcal{T}_h}$. Consider now the following *augmented problem*: find $\Lambda \in \mathbb{R}^{|\mathcal{E}_h|}$ and $P \in \mathbb{R}^{|\mathcal{T}_h|}$ such that (1.1) holds together with

$$\mathbb{N}\Lambda = P. \tag{2.1}$$

This problem is well posed in the sense that there exists a unique solution (Λ, P) of (1.1), (2.1). Indeed, (1.1) defines Λ in a unique way thanks to its well-posedness. As (2.1) is completely uncoupled from (1.1), P is simply prescribed by $P := \mathbb{N}\Lambda$.

2.3 Structure of the algorithm

In order to present the key idea of our approach as clearly as possible and to underline its simplicity, we now present the structure of our algorithm:

1. Assemble the matrices $\mathbb{Z}^{\mathrm{i}}$, $\mathbb{Z}^{\mathrm{b}}$, and $\mathbb{N}$ from (1.1), (2.1).
2. Consider certain lines from the first block row of (1.1) and from (2.1) on patches of elements around vertices and assemble the local problems (2.4), see Section 2.4.
3. Run through all vertices, invert the local matrices $\mathbb{M}_V$, and assemble the matrices $\widetilde{\mathbb{M}}^{\mathrm{inv}}$ and $\overline{\mathbb{M}}^{\mathrm{inv}}$ from (2.10), see Section 2.6.
4. Solve the reduced system (2.11), see Section 2.6.
5. Get Λ^{i} from (2.12), see Section 2.6.

2.4 Definition of the local problems

Consider now a vertex V from the set of vertices $\mathcal{V}_h$. Recall that we have denoted by $\mathcal{T}_V$ the elements sharing the vertex V and by $\mathcal{E}_V^{\mathrm{int}}$ the interior faces of this patch. Consider the lines from the first block row of (1.1) associated with the faces from $\mathcal{E}_V^{\mathrm{int}}$ and the lines from (2.1) associated with the elements from $\mathcal{T}_V$. The appearing unknowns correspond to the faces of the set $\mathcal{E}_V \cup \mathcal{E}_V^{\mathrm{b}}$, $\Lambda_V := \{\Lambda_\sigma\}_{\sigma \in \mathcal{E}_V \cup \mathcal{E}_V^{\mathrm{b}}}$, see Figures 1–3. Remark that $\mathcal{E}_V \cup \mathcal{E}_V^{\mathrm{b}}$ corresponds to all faces of the patch $\mathcal{T}_V$. We denote the corresponding submatrices by $\mathbb{Z}_V$ and $\mathbb{N}_V$ and the right-hand side vectors by E_V, $E_V := \{E_\sigma\}_{\sigma \in \mathcal{E}_V^{\mathrm{int}}}$ and P_V, $P_V := \{P_K\}_{K \in \mathcal{T}_V}$, respectively. This gives rise to the following *local linear system* for all $V \in \mathcal{V}_h$: given the vectors E_V and P_V, find the vector Λ_V such that

$$\begin{pmatrix} \mathbb{Z}_V \\ \mathbb{N}_V \end{pmatrix} \Lambda_V = \begin{pmatrix} E_V \\ P_V \end{pmatrix}. \tag{2.2}$$

The size of the matrix $\mathbb{Z}_V$ is $|\mathcal{E}_V^{\mathrm{int}}| \times |\mathcal{E}_V \cup \mathcal{E}_V^{\mathrm{b}}|$, whereas the size of $\mathbb{N}_V$ is $|\mathcal{T}_V| \times |\mathcal{E}_V \cup \mathcal{E}_V^{\mathrm{b}}|$. System (2.2) is square if the vertex V lies inside Ω and thus $\mathcal{E}_V^{\mathrm{b}} = \emptyset$, see Figures 1–2. If $V \in \partial\Omega$, then $\mathcal{E}_V^{\mathrm{b}} \neq \emptyset$,

see Figure 3, and the homogeneous Dirichlet boundary condition gives $\Lambda^{\mathrm{b}} = 0$. Thus, keeping only $\Lambda_V^{\mathrm{int}} := \{\Lambda_\sigma\}_{\sigma \in \mathcal{E}_V^{\mathrm{int}}}$ and $\Lambda_V^{\mathrm{ext}} := \{\Lambda_\sigma\}_{\sigma \in \mathcal{E}_V^{\mathrm{ext}}}$ as the unknowns, we can rewrite (2.2) as a block system which is always square:

$$\begin{pmatrix} \mathbb{Z}_V^{\mathrm{int}} & \mathbb{Z}_V^{\mathrm{ext}} \\ \mathbb{N}_V^{\mathrm{int}} & \mathbb{N}_V^{\mathrm{ext}} \end{pmatrix} \begin{pmatrix} \Lambda_V^{\mathrm{int}} \\ \Lambda_V^{\mathrm{ext}} \end{pmatrix} = \begin{pmatrix} E_V \\ P_V \end{pmatrix}. \tag{2.3}$$

Indeed, it follows from elementary properties of the mesh $\mathcal{T}_h$ that $|\mathcal{T}_V| = |\mathcal{E}_V^{\mathrm{ext}}|$, which gives $|\mathcal{E}_V^{\mathrm{int}}| + |\mathcal{T}_V| = |\mathcal{E}_V^{\mathrm{int}}| + |\mathcal{E}_V^{\mathrm{ext}}|$. Moreover, ordering the elements in the patch $\mathcal{T}_V$ with their adjacent external faces from $\mathcal{E}_V^{\mathrm{ext}}$, the matrix $\mathbb{N}_V^{\mathrm{ext}}$ is diagonal.

Suppose now that the diagonal entries of the matrix $\mathbb{N}_V^{\mathrm{ext}}$ are nonzero and that the matrix of (2.3) is nonsingular. Then (2.3) can be reduced to the following *Schur-complement system*:

$$\mathbb{M}_V \Lambda_V^{\mathrm{int}} = E_V - \mathbb{J}_V P_V \tag{2.4}$$

with

$$\mathbb{M}_V := \mathbb{Z}_V^{\mathrm{int}} - \mathbb{Z}_V^{\mathrm{ext}} (\mathbb{N}_V^{\mathrm{ext}})^{-1} \mathbb{N}_V^{\mathrm{int}}, \tag{2.5a}$$

$$\mathbb{J}_V := \mathbb{Z}_V^{\mathrm{ext}} (\mathbb{N}_V^{\mathrm{ext}})^{-1}. \tag{2.5b}$$

We obtain in particular from (2.4)

$$\Lambda_V^{\mathrm{int}} = (\mathbb{M}_V)^{-1}(E_V - \mathbb{J}_V P_V), \tag{2.6}$$

i.e., a *local expression* of the *original unknowns* Λ_V^{int} from the *new unknowns* P_V. These local expressions, together with the relation (2.1), will enable us to reduce system (1.1), (2.1) to (1.3). In order to state such a result in a proper way, we need to introduce some more notation.

2.5   Weights and extension operators

Each face $\sigma \in \mathcal{E}_h^{\mathrm{i}}$ belongs to d sets $\mathcal{E}_V^{\mathrm{int}}$ (recall that d is the space dimension). We associate a *weight* $w_{V,\sigma}$ to each vertex $V \in \mathcal{V}_h$ and each face $\sigma \in \mathcal{E}_V^{\mathrm{int}}$ and require

$$\sum_{V;\, \sigma \in \mathcal{E}_V^{\mathrm{int}}} w_{V,\sigma} = 1 \qquad \forall \sigma \in \mathcal{E}_h^{\mathrm{i}}. \tag{2.7}$$

The condition (2.7) is sufficient for all our theoretical developments and gives a great flexibility; in the numerical experiments in Section 5 below we, however, only consider the simplest case $w_{V,\sigma} = 1/d$.

Let $V \in \mathcal{V}_h$. Let us define a mapping $\Upsilon_V : \mathbb{R}^{|\mathcal{E}_V^{\mathrm{int}}|} \to \mathbb{R}^{|\mathcal{E}_h^{\mathrm{i}}|}$, extending a vector $\Lambda_V^{\mathrm{int}} = \{\Lambda_\sigma\}_{\sigma \in \mathcal{E}_V^{\mathrm{int}}}$ of values associated with the faces from $\mathcal{E}_V^{\mathrm{int}}$ to a vector of values associated with all interior faces $\mathcal{E}_h^{\mathrm{i}}$ by

$$[\Upsilon_V(\Lambda_V^{\mathrm{int}})]_\sigma := \begin{cases} \Lambda_\sigma & \text{if } \sigma \in \mathcal{E}_V^{\mathrm{int}} \\ 0 & \text{if } \sigma \notin \mathcal{E}_V^{\mathrm{int}} \end{cases}. \tag{2.8}$$

Let $\mathbb{W}_V$ be a diagonal matrix of size $|\mathcal{E}_V^{\mathrm{int}}| \times |\mathcal{E}_V^{\mathrm{int}}|$, with the diagonal entries given by the weights $w_{V,\sigma}$. Let $\Lambda^{\mathrm{i}} \in \mathbb{R}^{|\mathcal{E}_h^{\mathrm{i}}|}$ be an arbitrary vector and denote $\Lambda_V^{\mathrm{int}} = \{\Lambda_\sigma\}_{\sigma \in \mathcal{E}_V^{\mathrm{int}}}$ for a given vertex $V \in \mathcal{V}_h$. Due to (2.7), we have

$$\Lambda^{\mathrm{i}} = \sum_{V \in \mathcal{V}_h} \Upsilon_V(\mathbb{W}_V \Lambda_V^{\mathrm{int}}). \tag{2.9}$$

Let us now introduce a mapping $\Upsilon_V : \mathbb{R}^{|\mathcal{E}_V^{\mathrm{int}}| \times |\mathcal{E}_V^{\mathrm{int}}|} \to \mathbb{R}^{|\mathcal{E}_h^{\mathrm{i}}| \times |\mathcal{E}_h^{\mathrm{i}}|}$ (with the same name as the previous one, as one can easily distinguish them), extending a local matrix $\mathbb{M}_V$ to a full-size one by zeros by

$$[\Upsilon_V(\mathbb{M}_V)]_{\sigma,\gamma} := \left\{ \begin{array}{ll} (\mathbb{M}_V)_{\sigma,\gamma} & \text{if } \sigma \in \mathcal{E}_V^{\mathrm{int}} \text{ and } \gamma \in \mathcal{E}_V^{\mathrm{int}} \\ 0 & \text{if } \sigma \notin \mathcal{E}_V^{\mathrm{int}} \text{ or } \gamma \notin \mathcal{E}_V^{\mathrm{int}} \end{array} \right. .$$

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

$$[\Theta_V(\mathbb{J}_V)]_{\sigma,K} := \left\{ \begin{array}{ll} (\mathbb{J}_V)_{\sigma,K} & \text{if } \sigma \in \mathcal{E}_V^{\mathrm{int}} \text{ and } K \in \mathcal{T}_V \\ 0 & \text{if } \sigma \notin \mathcal{E}_V^{\mathrm{int}} \text{ or } K \notin \mathcal{T}_V \end{array} \right. .$$

2.6 The reduced system

Now we are in a position to formulate our reduced system.

Lemma 2.1 (Reduction of (1.1) to one unknown per element)**.** *Consider problem* (1.1) *and augment it by* (2.1)*. For each vertex $V \in \mathcal{V}_h$, define the local problems by* (2.3)*. Let the matrices $\mathbb{N}_V^{\mathrm{ext}}$ of* (2.3) *and $\mathbb{M}_V$ of* (2.5a) *be nonsingular. Define $\mathbb{J}_V$ by* (2.5b) *and $\widetilde{\mathbb{M}}^{\mathrm{inv}} \in \mathbb{R}^{|\mathcal{E}_h^{\mathrm{i}}| \times |\mathcal{E}_h^{\mathrm{i}}|}$ and $\overline{\mathbb{M}}^{\mathrm{inv}} \in \mathbb{R}^{|\mathcal{E}_h^{\mathrm{i}}| \times |\mathcal{T}_h|}$ by*

$$\widetilde{\mathbb{M}}^{\mathrm{inv}} := \sum_{V \in \mathcal{V}_h} \Upsilon_V(\mathbb{W}_V(\mathbb{M}_V)^{-1}), \quad \overline{\mathbb{M}}^{\mathrm{inv}} := \sum_{V \in \mathcal{V}_h} \Theta_V(\mathbb{W}_V(\mathbb{M}_V)^{-1}\mathbb{J}_V). \tag{2.10}$$

Denote $\mathbb{N} = (\mathbb{N}^{\mathrm{i}}, \mathbb{N}^{\mathrm{b}})$, as induced by $\Lambda = (\Lambda^{\mathrm{i}}, \Lambda^{\mathrm{b}})^t$. Then (1.1)*,* (2.1) *can be reduced to*

$$(\mathbb{N}^{\mathrm{i}} \overline{\mathbb{M}}^{\mathrm{inv}} + \mathbb{I})P = \mathbb{N}^{\mathrm{i}} \widetilde{\mathbb{M}}^{\mathrm{inv}} E, \tag{2.11}$$

i.e., a system of the form (1.3) *with $\mathbb{S} \in \mathbb{R}^{|\mathcal{T}_h| \times |\mathcal{T}_h|}$ given by $\mathbb{S} := \mathbb{N}^{\mathrm{i}} \overline{\mathbb{M}}^{\mathrm{inv}} + \mathbb{I}$ and $H \in \mathbb{R}^{|\mathcal{T}_h|}$ given by $H := \mathbb{N}^{\mathrm{i}} \widetilde{\mathbb{M}}^{\mathrm{inv}} E$. The original unknowns Λ^{i} then satisfy* (2.6)*.*

Proof. Using the notation for the extension operators, we obtain from (2.6)

$$\Upsilon_V(\mathbb{W}_V \Lambda_V^{\mathrm{int}}) = \Upsilon_V(\mathbb{W}_V(\mathbb{M}_V)^{-1})E - \Theta_V(\mathbb{W}_V(\mathbb{M}_V)^{-1}\mathbb{J}_V)P.$$

Summing over all $V \in \mathcal{V}_h$ and using (2.9), we deduce

$$\Lambda^{\mathrm{i}} = \widetilde{\mathbb{M}}^{\mathrm{inv}} E - \overline{\mathbb{M}}^{\mathrm{inv}} P \tag{2.12}$$

with the matrices $\widetilde{\mathbb{M}}^{\mathrm{inv}}$ and $\overline{\mathbb{M}}^{\mathrm{inv}}$ specified in (2.10). We have from (2.1) and from the second block row of (1.1)

$$\mathbb{N}\Lambda = (\mathbb{N}^{\mathrm{i}}, \mathbb{N}^{\mathrm{b}}) \left(\begin{array}{c} \Lambda^{\mathrm{i}} \\ \Lambda^{\mathrm{b}} \end{array} \right) = \mathbb{N}^{\mathrm{i}} \Lambda^{\mathrm{i}} = P. \tag{2.13}$$

Plugging (2.12) into (2.13), we arrive at (2.11). $\qquad \square$
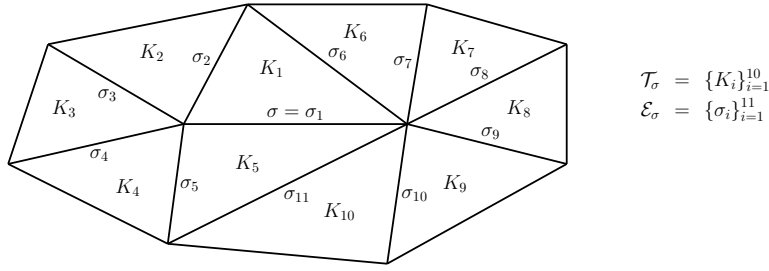
Figure 4: An example of the patch $\mathcal{T}_\sigma$ and of the set $\mathcal{E}_\sigma$ for $\sigma \in \mathcal{E}_h^{\mathrm{i}}$ in the interior of Ω

$$\mathcal{T}_\sigma = \{K_i\}_{i=1}^{10}$$
$$\mathcal{E}_\sigma = \{\sigma_i\}_{i=1}^{11}$$

$$\mathcal{T}_\sigma = \{K_i\}_{i=1}^{5}$$
$$\mathcal{E}_\sigma = \{\sigma_i\}_{i=1}^{5}$$
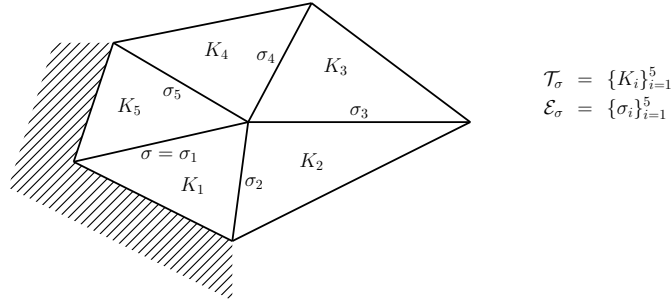
Figure 5: An example of the patch $\mathcal{T}_\sigma$ and of the set $\mathcal{E}_\sigma$ for $\sigma \in \mathcal{E}_h^{\mathrm{i}}$ having a node on the boundary of Ω

2.7 Equivalence of the reduced system with the original one

For the moment, we do not know whether system (2.11) is suitable for solving the original problem (1.1). We show here that under certain assumptions, (2.11) is well-posed and equivalent to (1.1).

Let $\sigma \in \mathcal{E}_h^{\mathrm{i}}$ be fixed and let $\mathcal{T}_\sigma$ be given by all elements which share a vertex with σ and $\mathcal{E}_\sigma$ by the interior faces of this patch, see Figures 4 and 5. For a fixed $\sigma \in \mathcal{E}_h^{\mathrm{i}}$, consider the lines from the first block row of (1.1) associated with the faces from $\mathcal{E}_\sigma$ and the lines from (2.1) associated with the elements from $\mathcal{T}_\sigma$, giving rise to

$$\begin{pmatrix} \mathbb{Z}_\Sigma \\ \mathbb{N}_\Sigma \end{pmatrix} \Lambda_\Sigma = \begin{pmatrix} E_\Sigma \\ P_\Sigma \end{pmatrix}, \tag{2.14}$$

with Λ_Σ being the vector of unknowns Λ_γ associated with all faces γ of the patch $\mathcal{T}_\sigma$. The matrix $\mathbb{Z}_\Sigma$ has $|\mathcal{E}_\sigma|$ rows, and $\mathbb{N}_\Sigma$ has $|\mathcal{T}_\sigma|$ rows. Note that the number of rows in (2.14) is always greater than or equal to the number of columns. In Figure 4, $|\mathcal{E}_\sigma| = 11$, $|\mathcal{T}_\sigma| = 10$, and the dimension of Λ_Σ is $11 + 8$, whereas in Figure 5, we have a square system of size 10.

In order to study the wellposedness of the global system (2.11), we will use the following assumption on the local systems (2.3) and (2.14):

Assumption 2.2 (Local well-posedness properties)**.** *The mesh $\mathcal{T}_h$, the matrices $\mathbb{Z}^{\mathrm{i}}$, $\mathbb{Z}^{\mathrm{b}}$, the matrix $\mathbb{N}$, and the corresponding matrices $\mathbb{N}_V^{\mathrm{ext}}$ in (2.3), $\mathbb{M}_V$ in (2.4), and $\begin{pmatrix} \mathbb{Z}_\Sigma \\ \mathbb{N}_\Sigma \end{pmatrix}$ in (2.14) are such that:*

$$\mathbb{N}_V^{\mathrm{ext}} \ \text{is nonsingular} \quad \forall V \in \mathcal{V}_h; \tag{2.15a}$$

$$\mathbb{M}_V \ \text{is nonsingular} \quad \forall V \in \mathcal{V}_h; \tag{2.15b}$$

$$\begin{pmatrix} \mathbb{Z}_\Sigma \\ \mathbb{N}_\Sigma \end{pmatrix} \ \text{has a full rank} \quad \forall \sigma \in \mathcal{E}_h^{\mathrm{i}}. \tag{2.15c}$$

Under Assumption 2.2, we obtain our main results:

Theorem 2.3 (Equivalence of (1.1), (2.1) and of (2.11), (2.12)). *Let Assumption 2.2 hold. Let (Λ, P) be the solution of (1.1), (2.1). Then P is a solution of (2.11), Λ^{i} of (2.12), and $\Lambda^{\mathrm{b}} = 0$. Conversely, let P be a solution of (2.11) and let Λ^{i} be given by (2.12). Set $\Lambda^{\mathrm{b}} = 0$. Then the couple (Λ, P) is the solution of (1.1), (2.1).*

Proof. Let (Λ, P) be the solution of (1.1), (2.1). The fact that P is also a solution of (2.11) and Λ^{i} the solution of (2.12) follows by Lemma 2.1, using assumptions (2.15a) and (2.15b). $\Lambda^{\mathrm{b}} = 0$ trivially follows from the second block line of (1.1).

We now prove the converse. Let P be a solution of (2.11). Fix $V \in \mathcal{V}_h$ and consider the corresponding system (2.3). Then define Λ_V^{int} by the Schur complement procedure (2.6). This is possible, thanks to assumptions (2.15a) and (2.15b). The crucial question, however, is whether the values for Λ_σ, $\sigma \in \mathcal{E}_h^{\mathrm{i}}$, obtained from systems (2.6) for different V coincide. Suppose that there exists an interior face $\sigma \in \mathcal{E}_h^{\mathrm{i}}$ and vertices $V \in \mathcal{V}_h$ and $V' \in \mathcal{V}_h$ such that $\sigma \in \mathcal{E}_V^{\mathrm{int}} \cap \mathcal{E}_{V'}^{\mathrm{int}}$ and such that the value Λ_σ obtained from (2.6) corresponding to V and from (2.6) corresponding to V' are different. Then this leads to a contradiction. Indeed, both corresponding systems (2.3) are subsystems of (2.14). As there are at least as many rows as columns in (2.14) and by assumption (2.15c), all obtained values Λ_σ must coincide. Completing them with $\Lambda_\sigma = 0$ for all $\sigma \in \mathcal{E}_h^{\mathrm{b}}$, we have (2.2). As each face $\sigma \in \mathcal{E}_h^{\mathrm{i}}$ belongs to at least one set $\mathcal{E}_V^{\mathrm{int}}$, running with (2.2) through all $V \in \mathcal{V}_h$, we obtain all lines of (1.1), (2.1). $\square$

The following is an immediate consequence of Theorem 2.3 and of the well-posedness of (1.1), (2.1):

Corollary 2.4 (Well-posedness of (2.11)). *Let Assumption 2.2 hold. Then there exists a unique solution P of (2.11).*

Finally, the following theorem shows that problem (2.11) is computationally appealing:

Theorem 2.5 (Stencil of the system matrix of (2.11)). *Let Assumption 2.2 hold. Let $K \in \mathcal{T}_h$. Then, on a row of the system matrix $\mathbb{S} := \mathbb{N}^{\mathrm{i}} \overline{\mathbb{M}}^{\mathrm{inv}} + \mathbb{I}$ of (2.11) associated with this element K, only possible nonzero entries are on columns associated with elements L of $\mathcal{T}_h$ such that K and L share a common vertex.*

Proof. It follows from (2.6) and (2.9) that, for $\sigma \in \mathcal{E}_h^{\mathrm{i}}$, Λ_σ can depend on P_K whenever there is $V \in \mathcal{V}_h$ such that $\sigma \in \mathcal{E}_V^{\mathrm{int}}$ and $K \in \mathcal{T}_V$. The assertion follows from the fact that (2.11) has been obtained upon plugging (2.12) into (2.13). $\square$

3 Realization for Crouzeix–Raviart discretization of linear elliptic diffusion problems and geometrical interpretation

We apply here the developments of the previous section to the discretization of a model linear elliptic diffusion problem by the Crouzeix–Raviart nonconforming finite element method. We also give a geometrical interpretation and show how the properties of the final matrix $\mathbb{S}$ can be influenced as a function of the local geometry and of the diffusion tensor $\underline{\mathbf{S}}$.

3.1 Model linear elliptic diffusion problem

We consider in this section the pure diffusion model problem

$$-\nabla\cdot(\underline{\mathbf{S}}\nabla p) = g \qquad \text{in } \Omega, \tag{3.1a}$$

$$p = 0 \qquad \text{on } \partial\Omega. \tag{3.1b}$$

For simplicity of our exposition, we only consider the homogeneous Dirichlet boundary condition (3.1b) and assume that $\underline{\mathbf{S}}$, a symmetric, bounded, and uniformly positive definite diffusion–dispersion tensor, and g, the source term, are piecewise constant on $\mathcal{T}_h$.

3.2 The Crouzeix–Raviart nonconforming finite element method

Associate with each $\sigma \in \mathcal{E}_h$ the basis function ψ_σ which is piecewise affine on $\mathcal{T}_h$ and satisfies

$$\psi_\sigma(\mathbf{x}_\sigma) = 1, \quad \psi_\sigma(\mathbf{x}_\gamma) = 0 \quad \forall \gamma \in \mathcal{E}_h, \gamma \neq \sigma. \tag{3.2}$$

The Crouzeix–Raviart nonconforming space Ψ_h (see [6]) is defined as $\Psi_h := \operatorname{span}\{\psi_\sigma; \sigma \in \mathcal{E}_h\}$. Then the discrete weak problem formulation reads: find $\lambda_h \in \Psi_h$ such that $\lambda_h(\mathbf{x}_\sigma) = 0$ for all $\sigma \in \mathcal{E}_h^{\mathrm{b}}$ and such that

$$(\underline{\mathbf{S}}\nabla\lambda_h, \nabla\psi_h) = (g, \psi_h) \qquad \forall \psi_h \in \Psi_h \text{ such that } \psi_h(\mathbf{x}_\sigma) = 0 \quad \forall \sigma \in \mathcal{E}_h^{\mathrm{b}}. \tag{3.3}$$

Herein, $(\cdot, \cdot)$ stands for the L^2-scalar product on Ω and ∇ for the broken gradient operator such that for a function v that is smooth within each mesh element, $\nabla v \in [L^2(\Omega)]^d$ is defined as $(\nabla v)|_K := \nabla(v|_K)$ for all $K \in \mathcal{T}_h$. Using the notion of the basis functions ψ_σ of (3.2), we have $\lambda_h = \sum_{\sigma\in\mathcal{E}_h} \Lambda_\sigma \psi_\sigma$. Thus (3.3) gives rise to (1.1) with the algebraic vector of unknowns $\Lambda^{\mathrm{i}} := \{\Lambda_\sigma\}_{\sigma\in\mathcal{E}_h^{\mathrm{i}}}$, $\Lambda^{\mathrm{b}} := \{\Lambda_\sigma\}_{\sigma\in\mathcal{E}_h^{\mathrm{b}}}$, and

$$\mathbb{Z}_{\gamma,\sigma}^{\mathrm{i}} = (\underline{\mathbf{S}}\nabla\psi_\sigma, \nabla\psi_\gamma) \qquad \gamma, \sigma \in \mathcal{E}_h^{\mathrm{i}}, \tag{3.4a}$$

$$\mathbb{Z}_{\gamma,\sigma}^{\mathrm{b}} = (\underline{\mathbf{S}}\nabla\psi_\sigma, \nabla\psi_\gamma) \qquad \gamma \in \mathcal{E}_h^{\mathrm{i}}, \sigma \in \mathcal{E}_h^{\mathrm{b}}, \tag{3.4b}$$

$$E_\gamma = (g, \psi_\gamma) \qquad \gamma \in \mathcal{E}_h^{\mathrm{i}}. \tag{3.4c}$$

We recall that the Crouzeix–Raviart nonconforming finite element method is equivalent to the lowest-order Raviart–Thomas mixed finite element method [10], see, e.g., Arnold and Brezzi [3], Marini [9], Arbogast and Chen [2], and [15, 17], and the references therein.

3.3 Geometrical interpretation

We now give a geometrical interpretation of the approach of Section 2 for the Crouzeix–Raviart nonconforming finite element method (3.3).

We start by the interpretation of the values P_K defined by (2.1). Let $K \in \mathcal{T}_h$. Remark that

$$\lambda_h|_K = \sum_{\sigma\in\mathcal{E}_K} \Lambda_\sigma \psi_\sigma, \tag{3.5}$$

where, we recall, ψ_σ are the basis functions specified by (3.2). Here and in the sequel, we understand by $|_K$ the polynomial on the simplex K extended to the whole space $\mathbb{R}^d$. Denote by $\mathbb{N}_K$ the row of (2.1) associated with the element K. Then this row gives $\mathbb{N}_K\Lambda_K = P_K$, with $\Lambda_K := \{\Lambda_\sigma\}_{\sigma\in\mathcal{E}_K}$. Combining this observation with (3.5), we have the following geometrical interpretation of the values P_K, $K \in \mathcal{T}_h$, under assumption (3.6):
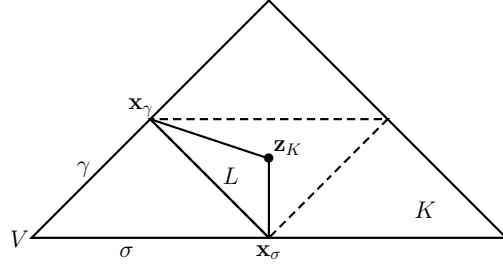
9

Figure 6: Triangle $K \in \mathcal{T}_h$ and subtriangle L given by edge midpoints $\mathbf{x}_\sigma$ and $\mathbf{x}_\gamma$ and the point $\mathbf{z}_K$

Lemma 3.1 (Geometrical interpretation of the values P_K). *Let $\mathbb{N}$ in (2.1) be scaled such that*

$$\sum_{\sigma \in \mathcal{E}_K} \mathbb{N}_{K,\sigma} = 1 \qquad \forall K \in \mathcal{T}_h. \tag{3.6}$$

Then, for any simplex $K \in \mathcal{T}_h$, there is a uniquely defined point $\mathbf{z}_K \in \mathbb{R}^d$ (not necessarily inside K) such that

$$P_K = \lambda_h|_K(\mathbf{z}_K). \tag{3.7}$$

Conversely, associate with any $K \in \mathcal{T}_h$ a point $\mathbf{z}_K \in \mathbb{R}^d$. Then, there is a uniquely defined matrix $\mathbb{N}$ such that (2.1) and (3.6) hold.

In Section 2, we have made assumption (2.15a), requiring that all diagonal entries of the matrix $\mathbb{N}_V^{\text{ext}}$ are nonzero. This assumption combined with (3.6) is nothing but assuming that the points $\mathbf{z}_K$ are such that any d of the $(d+1)$ face barycenters $\mathbf{x}_\sigma$ of K and the point $\mathbf{z}_K$ do not lie in the same hyperplane. In two space dimensions, this means that any two of the three edge midpoints $\mathbf{x}_\sigma$ and the point $\mathbf{z}_K$ do not lie on the same line, i.e., that $\mathbf{z}_K$ does not lie on the boundary of the dashed triangle in Figure 6.

We now give (geometrical) formulas for the local matrices and vectors $\mathbb{M}_V$, E_V, and $\mathbb{J}_V$ of (2.4). Let $K \in \mathcal{T}_h$ and let V be any of the vertices of K. Recall that $\mathcal{E}_{V,K}$ denotes the set of faces of K which have V as vertex. Let a new simplex L (subsimplex of K) be given by the face barycenters $\mathbf{x}_\sigma$, $\sigma \in \mathcal{E}_{V,K}$, and by the point $\mathbf{z}_K$, see Figure 6 (here the points are denoted by $\mathbf{z}_K$, $\mathbf{x}_\sigma$, and $\mathbf{x}_\gamma$). Let φ_σ, $\sigma \in \mathcal{E}_{V,K}$, be the affine function which takes the value 1 in $\mathbf{x}_\sigma$, value 0 in $\mathbf{x}_\gamma$, $\gamma \in \mathcal{E}_{V,K}$, $\gamma \neq \sigma$, and value 0 in $\mathbf{z}_K$. Similarly, let φ_K take the value 1 in $\mathbf{z}_K$ and value 0 in $\mathbf{x}_\sigma$, $\sigma \in \mathcal{E}_{V,K}$. The functions φ_σ, $\sigma \in \mathcal{E}_{V,K}$, and φ_K are the Lagrange basis functions of the first-order polynomials on the simplex L. We extend them onto K to form basis functions of the first-order polynomials on K; we illustrate their gradients in Figure 7. By this basis transformation and by (3.5), (3.7), we get

$$\lambda_h|_K = \sum_{\sigma \in \mathcal{E}_{V,K}} \Lambda_\sigma \varphi_\sigma + P_K \varphi_K. \tag{3.8}$$

Let $V \in \mathcal{V}_h$. Using (3.4), (3.7), and the condition $\Lambda^{\mathrm{b}} = 0$, we see that (2.3) corresponds to

$$(\underline{\mathbf{S}} \nabla \lambda_h, \nabla \psi_\gamma) = (g, \psi_\gamma) \qquad \forall \gamma \in \mathcal{E}_V^{\text{int}},$$
$$P_K = \lambda_h|_K(\mathbf{z}_K) \qquad \forall K \in \mathcal{T}_V.$$

Employing (3.8), we obtain (2.4) with the following formulas for the local matrices and vectors:

10

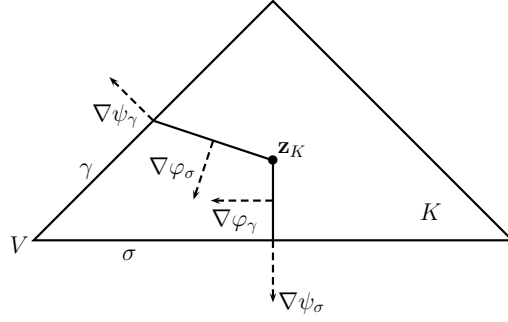Figure 7: Basis functions gradients in an element $K \in \mathcal{T}_h$

Lemma 3.2 (Form of the local matrices and vectors $\mathbb{M}_V$, E_V, and $\mathbb{J}_V$). *Let $\mathbb{Z}^{\mathrm{i}}$, $\mathbb{Z}^{\mathrm{b}}$, and E_γ in (1.1) be given by (3.4). Let $\mathbb{N}$ in (2.1) satisfy (3.6). Then, for all $V \in \mathcal{V}_h$, there holds*

$$(\mathbb{M}_V)_{\sigma,\gamma} := \sum_{K \in \mathcal{T}_V;\, \sigma \in \mathcal{E}_K} (\underline{\mathbf{S}}\nabla\varphi_\gamma, \nabla\psi_\sigma)_K, \tag{3.10a}$$

$$(E_V)_\sigma := \sum_{K \in \mathcal{T}_V;\, \sigma \in \mathcal{E}_K} (g_K, \psi_\sigma)_K, \tag{3.10b}$$

$$(\mathbb{J}_V)_{\sigma,K} := -(\underline{\mathbf{S}}\nabla\varphi_K, \nabla\psi_\sigma)_K. \tag{3.10c}$$

3.4 Sufficient conditions for Assumption 2.2

The geometrical interpretation and the form of the local matrices of Section 3.3 helps us to give a sufficient condition to satisfy Assumption 2.2:

Theorem 3.3 (A sufficient condition for Assumption 2.2). *Let $\mathbb{Z}^{\mathrm{i}}$, $\mathbb{Z}^{\mathrm{b}}$, and E_γ in (1.1) be given by (3.4). Let $\mathbb{N}$ in (2.1) satisfy (3.6) and let $\mathbb{N}_{K,\sigma} \neq 0$ for all $K \in \mathcal{T}_h$ and $\sigma \in \mathcal{E}_K$. Let the matrices $\mathbb{E}_{V,K} \in \mathbb{R}^{d \times d}$ given by (with the notation of Figure 7)*

$$(\mathbb{E}_{V,K})_{\sigma,\gamma} := (\underline{\mathbf{S}}\nabla\varphi_\gamma, \nabla\psi_\sigma)_K \tag{3.11}$$

be positive definite for all $K \in \mathcal{T}_h$ and for all vertices V of K. Then Assumption 2.2 holds true.

Proof. Let $V \in \mathcal{V}_h$. From the assumption that $\mathbb{N}_{K,\sigma} \neq 0$, we immediately have (2.15a), as the matrix $\mathbb{N}_V^{\mathrm{ext}}$ is diagonal and formed by these nonzero values.

We next show (2.15b), or, more precisely, that the positive definiteness of all $\mathbb{E}_{V,K}$, $K \in \mathcal{T}_V$, in fact implies the positive definiteness of $\mathbb{M}_V$. Let $X \in \mathbb{R}^{|\mathcal{E}_V^{\mathrm{int}}|}$, $X \neq 0$. Set $\varphi := \sum_{\sigma \in \mathcal{E}_V^{\mathrm{int}}} X_\sigma \varphi_\sigma$ and $\psi := \sum_{\sigma \in \mathcal{E}_V^{\mathrm{int}}} X_\sigma \psi_\sigma$. For each $K \in \mathcal{T}_V$, define a mapping $\Pi_{V,K} : \mathbb{R}^{|\mathcal{E}_V^{\mathrm{int}}|} \to \mathbb{R}^{|\mathcal{E}_V^{\mathrm{int}} \cap \mathcal{E}_K^{\mathrm{i}}|}$, restricting a vector of values associated with the faces from $\mathcal{E}_V^{\mathrm{int}}$ to a vector of values associated with the faces from $\mathcal{E}_V^{\mathrm{int}} \cap \mathcal{E}_K^{\mathrm{i}}$. Then, for a vertex not lying on the boundary of Ω,

$$X^t \mathbb{M}_V X = \sum_{K \in \mathcal{T}_V} (\underline{\mathbf{S}}\nabla\varphi, \nabla\psi)_K = \sum_{K \in \mathcal{T}_V} [\Pi_{V,K}(X)]^t \mathbb{E}_{V,K} \Pi_{V,K}(X) > 0, \tag{3.12}$$

due to the fact that all $\mathbb{E}_{V,K}$ are positive definite and X is nonzero. For boundary vertices, we proceed similarly, with only submatrices of $\mathbb{E}_{V,K}$.

11

We finally show (2.15c). We can write (2.14) in the form (cf. (2.3))

$$\begin{pmatrix} \mathbb{Z}_\Sigma^{\text{int}} & \mathbb{Z}_\Sigma^{\text{ext,b}} \\ \mathbb{N}_\Sigma^{\text{int}} & \mathbb{N}_\Sigma^{\text{ext,b}} \end{pmatrix} \begin{pmatrix} \Lambda_\Sigma^{\text{int}} \\ \Lambda_\Sigma^{\text{ext,b}} \end{pmatrix} = \begin{pmatrix} E_\Sigma \\ P_\Sigma \end{pmatrix}, \tag{3.13}$$

where $\Lambda_\Sigma^{\text{int}} := \{\Lambda_\gamma\}_{\gamma \in \mathcal{E}_\sigma}$ and $\Lambda_\Sigma^{\text{ext,b}}$ is the vector of unknowns Λ_γ associated with all faces γ lying on the boundary of the patch $\mathcal{T}_\sigma$. A crucial property is that each element $K \in \mathcal{T}_\sigma$ having a face on the boundary of $\mathcal{T}_\sigma$ has only one such face that we associate to it. This is in line with the fact that the matrix of (3.13) has at least as many rows as columns. Removing the rows of (3.13) corresponding to elements $K \in \mathcal{T}_\sigma$ not having a face on the boundary of $\mathcal{T}_\sigma$, (3.13) reduces to a square system

$$\begin{pmatrix} \mathbb{Z}_\Sigma^{\text{int}} & \mathbb{Z}_\Sigma^{\text{ext,b}} \\ \mathbb{N}_\Sigma^{'\text{int}} & \mathbb{N}_\Sigma^{'\text{ext,b}} \end{pmatrix} \begin{pmatrix} \Lambda_\Sigma^{\text{int}} \\ \Lambda_\Sigma^{\text{ext,b}} \end{pmatrix} = \begin{pmatrix} E_\Sigma \\ P_\Sigma' \end{pmatrix}. \tag{3.14}$$

After reordering the rows of the second block line of (3.14), $\mathbb{N}_\Sigma^{'\text{ext,b}}$ gets a nonsingular diagonal matrix. Then a Schur-complement form of (3.14) can be given (cf. (2.4)–(2.5b)). The entries of this Schur-complement matrix either take the form of (3.4a)–(3.4b) (when in the interior of $\mathcal{T}_\sigma$), or the form (3.10a). Then proceeding as in (3.12), positive definiteness of the Schur-complement matrix of (3.14) can be shown. Consequently, the matrix of (3.14) is nonsingular and thus the matrix of (3.13) has a full rank. $\qquad\square$

In two space dimensions, we have the following simple criterion:

Lemma 3.4 (A simple elementwise positive definiteness criterion in two space dimensions). *Let* $d = 2$. *The matrices* $\mathbb{E}_{V,K}$ *of Theorem 3.3 are positive definite if and only if for all elements* $K \in \mathcal{T}_h$ *and all vertices* V *of* K *(with the notation of Figure 7),*

$$\underline{\mathbf{S}}_K \nabla \varphi_\sigma \cdot \nabla \psi_\sigma > 0, \quad \underline{\mathbf{S}}_K \nabla \varphi_\gamma \cdot \nabla \psi_\gamma > 0, \tag{3.15a}$$

$$|\underline{\mathbf{S}}_K \nabla \varphi_\sigma \cdot \nabla \psi_\gamma + \underline{\mathbf{S}}_K \nabla \varphi_\gamma \cdot \nabla \psi_\sigma|^2 < 4(\underline{\mathbf{S}}_K \nabla \varphi_\sigma \cdot \nabla \psi_\sigma)(\underline{\mathbf{S}}_K \nabla \varphi_\gamma \cdot \nabla \psi_\gamma). \tag{3.15b}$$

Proof. A matrix $\mathbb{M}$ is positive definite if and only if its symmetric part $\frac{1}{2}(\mathbb{M} + \mathbb{M}^t)$ is positive definite and a symmetric matrix in $\mathbb{R}^2$ is positive definite if and only if its diagonal entries and its determinant are positive; (3.15a)–(3.15b) is nothing but applying this criterion to the matrices $\mathbb{E}_{V,K}$. $\qquad\square$

Figure 8 illustrates the sets of points $\mathbf{z}_K$ (filled region) for different examples of elements K where the local criterion (3.15a)–(3.15b) is satisfied (and thus our approach is guaranteed to work); the triangles connecting the edge midpoints are given by the dashed lines and the barycenters by the stars.

3.5 Choice of the evaluation point

From the developments above, we see that the freedom in the choice of the matrix $\mathbb{N}$ of (2.1) is expressed via the freedom of the choice of the evaluation point $\mathbf{z}_K$ associated with $K \in \mathcal{T}_h$; then the well-posedness of (2.11) and properties of the system matrix of (2.11) depend on the choice of the points $\mathbf{z}_K$. We distinguish three classes of specific points $\mathbf{z}_K$, $K \in \mathcal{T}_h$:

1. $\mathbf{z}_K$ is the barycenter of K;
2. $\mathbf{z}_K$ is the barycenter of the region where the matrices $\mathbb{E}_{V,K}$ of (3.11) for all vertices V of K are positive definite (equivalently, where (3.15a)–(3.15b) holds for $d = 2$) (barycenter of the filled region in Figure 8);
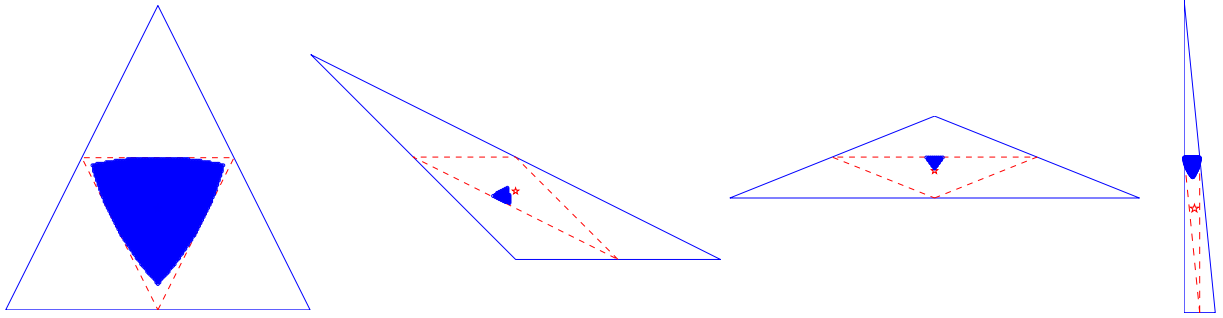
Figure 8: Examples of differently shaped triangles and of the corresponding regions where the local criterion (3.15a)–(3.15b) is satisfied

3. $\mathbf{z}_K$ is the $\underline{\mathbf{S}}$-circumcenter of K (if $d = 2$).

Remark 3.5 ($\underline{\mathbf{S}}$-circumcenter). *In case 3 above, the so-called $\underline{\mathbf{S}}$-circumcenter is the point characterized by the relations $\underline{\mathbf{S}}_K \nabla \varphi_\gamma \cdot \nabla \psi_\sigma = 0$ and $\underline{\mathbf{S}}_K \nabla \varphi_\sigma \cdot \nabla \psi_\gamma = 0$, with the notation of Figure 7. Such a concept is known, for instance, in finite volume methods, cf. Aavatsmark et al. [1]. Therein, the terminology "$\underline{\mathbf{S}}$-orthogonal grids" is used. It is important to note that, in this case, the* matrix $\mathbb{M}_V$ *of* (2.5a) *is* diagonal. *Consequently, its inverse is trivial and, moreover, the stencil of* $\mathbb{S}$ *further reduces from that stated in Theorem 2.5 to the neighbors of a given mesh element K only.*

More comments on these various choices, as well as on the relations to mixed finite element and finite volume methods, can be found in [17] and in the references therein.

4 Realization for Crouzeix–Raviart discretization of nonlinear parabolic advection–diffusion–reaction problems

Let $\beta(\cdot)$, $\underline{\mathbf{S}}(\cdot)$, $f(\cdot)$, and $r(\cdot)$ be given nonlinear functions and $\mathbf{w}$ a given vector field. Let a final simulation time $T > 0$ be given. We show here briefly that the approach of Section 2 can also be applied to Crouzeix–Raviart discretization of nonlinear parabolic advection–diffusion–reaction problems of the form

$$\partial_t \beta(p) - \nabla \cdot (\underline{\mathbf{S}}(p)\nabla p - f(p)\mathbf{w}) + r(p) = g \quad \text{in } \Omega \times (0, T), \tag{4.1a}$$

$$p = 0 \quad \text{on } \partial\Omega \times (0, T), \tag{4.1b}$$

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

Consider a strictly increasing sequence of discrete times $\{t^n\}_{0 \le n \le N}$ such that $t^0 = 0$ and $t^N = T$ and denote $\tau^n := t^n - t^{n-1}$, $1 \le n \le N$. Let $\lambda_h^0 \in \Psi_h$ be the approximation of the initial condition p_0. The Crouzeix–Raviart nonconforming finite element method combined with the backward Euler time stepping for problem (4.1a)–(4.1c) reads: for all $n \ge 1$, given λ_h^{n-1}, find $\lambda_h^n \in \Psi_h$ such that

$$\left(\frac{\beta(\lambda_h^n) - \beta(\lambda_h^{n-1})}{\tau^n}, \psi_h \right) + (\underline{\mathbf{S}}(\lambda_h^n)\nabla\lambda_h^n, \nabla\psi_h) + (\nabla \cdot (f(\lambda_h^n)\mathbf{w}), \psi_h)$$

$$+ (r(\lambda_h^n), \psi_h) = (g, \psi_h) \quad \forall \psi_h \in \Psi_h. \tag{4.2}$$

Typically, mass lumping and/or numerical quadrature is used in (4.2) for the temporal and reaction terms and upwind-weighting stabilization for the advection term. Such a procedure is closely

13

related to combined finite volume–nonconforming finite element method for (4.1a)–(4.1c), cf. [7] and the references therein. Finally, we obtain on each time step n a system of nonlinear algebraic equations of size $|\mathcal{E}_h| \times |\mathcal{E}_h|$. This system is typically linearized by, e.g., the Newton method; on a time step n and a linearization step k, a system of linear algebraic equations in the matrix form

$$\begin{pmatrix} \mathbb{Z}^{\mathrm{i},n,k} & \mathbb{Z}^{\mathrm{b},n,k} \\ 0 & \mathbb{I} \end{pmatrix} \begin{pmatrix} \Lambda^{\mathrm{i},n,k} \\ \Lambda^{\mathrm{b},n,k} \end{pmatrix} = \begin{pmatrix} E^{n,k} \\ 0 \end{pmatrix} \tag{4.3}$$

is obtained. Note that the matrices $\mathbb{Z}^{\mathrm{i},n,k}$, $\mathbb{Z}^{\mathrm{b},n,k}$ of (4.3) have the same sparsity pattern as the matrices $\mathbb{Z}^{\mathrm{i}}$, $\mathbb{Z}^{\mathrm{b}}$ of (3.4a)–(3.4b). Thus the procedure of Section 2 applies here as well. Consider a matrix $\mathbb{N} \in \mathbb{R}^{|\mathcal{T}_h| \times |\mathcal{E}_h|}$ as in Section 2.2 and set $\Lambda^{n,k} = (\Lambda^{\mathrm{i},n,k}, \Lambda^{\mathrm{b},n,k})^t$. We first complement (4.3) with

$$\mathbb{N}\Lambda^{n,k} = P^{n,k}, \tag{4.4}$$

in an analogy of (1.1), (2.1). We then form the local systems

$$\begin{pmatrix} \mathbb{Z}_V^{n,k,\mathrm{int}} & \mathbb{Z}_V^{n,k,\mathrm{ext}} \\ \mathbb{N}_V^{\mathrm{int}} & \mathbb{N}_V^{\mathrm{ext}} \end{pmatrix} \begin{pmatrix} \Lambda_V^{n,k,\mathrm{int}} \\ \Lambda_V^{n,k,\mathrm{ext}} \end{pmatrix} = \begin{pmatrix} E_V^{n,k} \\ P_V^{n,k} \end{pmatrix} \tag{4.5}$$

for all $V \in \mathcal{V}_h$, cf. (2.3). Proceeding from (4.5) in the way of Section 2, we reduce the linear system (4.3) on each time step n and linearization step k to an equivalent of (1.3) in the form

$$\mathbb{S}^{n,k} P^{n,k} = H^{n,k}. \tag{4.6}$$

In particular, the well-posedness of (4.6) is guaranteed under Assumption 2.2. Once we solve (4.6) for $P^{n,k}$, we obtain $\Lambda^{n,k}$ (the solution of (4.3)) by an equivalent of (2.12) and continue the linearization/time stepping in order to get next linear system (4.3).

5 Numerical experiments

We present in this section the results of several numerical experiments, illustrating the theoretical developments of the paper.

In Section 5.1, we concentrate on the model problem (3.1a)–(3.1b) and the developments of Section 3. We first in Section 5.1.1 focus on the influence of the choice of the evaluation point $\mathbf{z}_K$. Then, in Sections 5.1.2–5.1.4 respectively, we systematically compare various reduced formulations to the original nonconforming method for anisotropic meshes, inhomogeneous diffusion–dispersion tensor, and inhomogeneous and anisotropic diffusion–dispersion tensor, respectively. Section 5.2 is then devoted to the numerical illustration of the application of our approach to nonlinear parabolic (advection–)diffusion–reaction problems, illustrating the developments of Section 4.

For consistency, we use the same terminology as in [14, 17, 16]:

1. **MFEB**: approach of the present paper, $\mathbf{z}_K$ is the barycenter of K;
2. **MFEC**: approach of the present paper, $\mathbf{z}_K$ is the $\underline{\mathbf{S}}$-circumcenter of K;
3. **MFEO**: approach of the present paper, $\mathbf{z}_K$ is the barycenter of the region where the matrices $\mathbb{E}_{V,K}$ of (3.11) for all vertices V of K are positive definite;
4. **CMFE**: approach of [14];
5. **FV**: approach of [19, 4, 18], corresponding to a variant of the finite volume method;
6. **NCFE**: the original formulation (1.2);

Abbreviation	Meaning
Meth.	method, one of the equivalent nonconforming finite element formulations
Un.	number of unknowns (matrix size)
Mat.	matrix
St.	stencil (the maximum number of nonzero entries on each matrix row)
Nonz.	total number of matrix nonzero entries
CN	2-norm condition number
CNS	2-norm condition number after diagonal scaling
DS	direct linear solver
(P)IS	(preconditioned) iterative linear solver
Iter.	number of iterations of an iterative linear solver
Prec.	CPU time of incomplete factorizations for preconditioning
SPD	symmetric positive definite
SID	symmetric indefinite
NPD	nonsymmetric positive definite
NNS	nonsymmetric negative stable
NID	nonsymmetric indefinite

Table 1: Abbreviations used in Tables 2–11

7. **FV–NCFE**: combined finite volume–nonconforming finite element method (FV–NCFE) of [7] for unsteady nonlinear advection–diffusion–reaction problems.

Recall that the first five approaches have one unknown per element $K \in \mathcal{T}_h$, whereas the two last ones have one unknown per interior face $\sigma \in \mathcal{E}_h^{\mathrm{i}}$. Recall also all these approaches are *equivalent*, in the sense that the same solution Λ^{i} is always recovered (up to rounding errors). We refer to [19, 4, 18, 14] for details on the links between the different methods; a unified treatment and additional numerical experiments are presented in [17].

We present our results in Tables 2–11 below; Table 1 summarizes the different abbreviations used therein. Recall that a real matrix $\mathbb{S} \in \mathbb{R}^{M \times M}$ is positive definite if $P^t \mathbb{S} P > 0$ for all $P \in \mathbb{R}^M$, $P \neq 0$, and negative stable when all its eigenvalues have positive real parts (this is in particular the case for positive definite matrices). The 2-norm condition number of a matrix $\mathbb{S}$ is defined by $\|\mathbb{S}\|_2 \|\mathbb{S}^{-1}\|_2$. We also consider the 2-norm condition number after diagonal scaling, by which we mean the minimal of the two 2-norm condition numbers of the two matrices $\mathrm{diag}(\mathbb{S}))^{-1} \mathbb{S}$, $|\mathrm{diag}(\mathbb{S})|^{-1/2} \mathbb{S} |\mathrm{diag}(\mathbb{S})|^{-1/2}$. Here $|\mathrm{diag}(\mathbb{S})|^{-1/2}$ is obtained by taking the diagonal of the matrix $\mathbb{S}$, the absolute value of the entries, and the reciprocal of their square roots.

We also study the computational cost. We first test the Matlab "\" direct solver. A direct solver may not be usable for very large systems or may not be suitable for parabolic or nonlinear problems. Thus also the behavior of iterative solvers is very important. We test two simple iterative methods. If the matrix is symmetric and positive definite, we use the conjugate gradient method [8]. For nonsymmetric matrices, we employ the bi-conjugate gradient stabilized method [13]. Unpreconditioned iterative linear solvers may be rather slow but usually illustrate very well the matrix properties and especially the matrix condition number. To accelerate their convergence, we use incomplete Cholesky or incomplete LU factorizations with a specified drop tolerance as preconditioners, cf. [12]. 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

Meth.	Un.	Mat.	St.	Nonz.	b = 1		b = 0.1		b = 0.025	
					CN	CNS	CN	CNS	CN	CNS
MFEB	32	NPD	12	280	19	19	19	19	19	19
MFEO	32	NPD	13	298	19	18	15	15	14	14
NCFE	40	SPD	5	136	29	25	206	25	3090	25

Table 2: Matrix properties of the different equivalent reformulations of the nonconforming finite element method, coefficients (5.1), mesh A

minimal. We always use a zero start vector and stop the iterative process as soon as the relative residual $\|H - \mathbb{S}\tilde{P}\|_2/\|H\|_2$, where $\tilde{P}$ is the approximate solution to the system $\mathbb{S}P = H$, decreases below 1e-8.

All computations were performed in double precision on a notebook with Intel Core2 Duo 2.6 GHz processor and MS Windows Vista operating system. Machine precision was in the power of 1e-16. All linear system solutions were done with the help of MATLAB 7.0.4. In all tested cases, the criterion (3.15a)–(3.15b) was satisfied.

5.1 Linear elliptic problems

We first consider the Crouzeix–Raviart nonconforming finite element method (3.3) for the model problem (3.1a)–(3.1b).

5.1.1 Influencing the matrix properties by the choice of the evaluation point

We set here $\Omega := (0, b) \times (0, 1)$, where we test three different values of the parameter b: $b = 1$, $b = 0.1$, and $b = 0.025$. The associated meshes get more and more anisotropic while decreasing the value of b. We consider (3.1a)–(3.1b) with

$$\underline{\mathbf{S}} = \mathbb{I}, \tag{5.1}$$

$g = -2e^x e^y$, and a Dirichlet boundary condition given by the exact solution $p(x, y) = e^x e^y$. In the left part of Figure 9, we plot the evaluation points $\mathbf{z}_K$ given as the barycenters (MFEB) and as the optimal evaluation points $\mathbf{z}_K$ (MFEO approach) in the case $b = 0.025$. Note the alignment of these last points in almost horizontal lines (remark that the x and y axes have not the same scale in this figure).

In Table 2, we compare the original NCFE method with the MFEB and MFEO reformulations. We can see that the matrix condition number of the NCFE method is strongly influenced by the anisotropy of the mesh. In contrast, the MFEB reformulation shows a stable behavior. The choice of the optimized evaluation point in MFEO proves here as superior over MFEB, as the condition number even decreases with increasing the anisotropy of the mesh. The situation is different if we apply a diagonal scaling to the NCFE method. Then the condition number is also stable with respect to the mesh anisotropy. Thus our static condensation acts as a built-in diagonal preconditioner.

5.1.2 Anisotropic meshes

Let us now consider two domains Ω given respectively by the corners $[0, 0], [0.1, 0], [-0.1, 1], [-0.2, 1]$ and $[0, 0], [0.1, 0], [0.3, 1], [0.2, 1]$. Figure 9, middle and right, shows the corresponding generic mesh
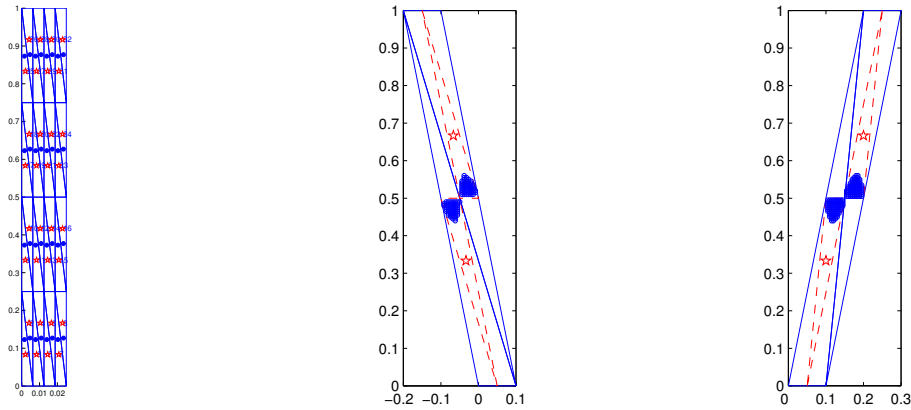
Figure 9: Mesh A with barycenters (stars) and optimal evaluation points (bullets) (left); generic elements of meshes B (middle) and C (right) with barycenters and the regions where the local criterion (3.15a)–(3.15b) is satisfied

							DS	IS		PIS		
Meth.	Un.	Mat.	St.	Nonz.	CN	CNS	CPU	CPU	Iter.	CPU	Prec.	Iter.
MFEB	8192	NNS	13	104458	6214	6214	0.13	3.04	422.0	0.68	0.09	28.0
MFEC	8192	SID	4	32512	794	808	0.05	2.57	760.5	0.16	0.03	8.0
MFEO	8192	NPD	13	99280	4206	3808	0.09	2.14	289.0	0.40	0.17	10.0
CMFE	8192	NPD	13	104458	3469	2675	0.13	1.88	256.5	0.30	0.06	11.5
FV	8192	SID	4	32512	794	808	0.05	2.54	754.5	0.16	0.04	8.0
NCFE	12160	SPD	5	60292	10164	7113	0.08	2.08	728.0	0.34	0.19	13.0

Table 3: Matrix properties and computational cost of the different equivalent formulations of the nonconforming finite element method, coefficients (5.1), mesh B

elements, together with the regions where the local criterion (3.15a)–(3.15b) is satisfied. As before, we use $\underline{\mathbf{S}}$ given by (5.1) and consider the same source function g and inhomogeneous Dirichlet boundary condition, leading to the exact solution $p(x,y) = e^x e^y$. Tables 3 and 4 present the results for the sixth-level uniform refinement of the meshes B and C, respectively.

The MFEC approach leads in this case to a symmetric matrix. Moreover, it turns out that this matrix coincides, up to a constant scaling factor, with that of the FV method. This matrix turns out to be very well conditioned, especially in the case of the mesh B. Consequently, the MFEC and FV methods give the smallest CPU times for both the direct and the preconditioned iterative solvers. The matrices produced by the MFEB approach turn out to have the highest condition number of the MFEB, MFEC, MFEO, and CMFE approaches. Consequently, the MFEB approach leads to the highest CPU times using the iterative solver (both with and without preconditioning). As before, the choice of the optimal evaluation point in the MFEO approach shows to be advantageous. In fact, the iterative solver can be even faster than that for the MFEC and FV approaches. Interestingly enough, the MFEB approach does not lead, as the only one, to positive definite matrices; also in this respect, the MFEO brings an improvement over MFEB. Finally, the MFEO approach can also lead to a decrease of the number of nonzero entries, see Table 3, whence also

17

							DS	IS		PIS		
Meth.	Un.	Mat.	St.	Nonz.	CN	CNS	CPU	CPU	Iter.	CPU	Prec.	Iter.
MFEB	8192	NNS	13	104458	5526	5193	0.14	2.84	380.5	0.69	0.10	28.5
MFEC	8192	SID	4	32512	1584	1589	0.05	2.29	651.5	0.16	0.05	6.5
MFEO	8192	NPD	13	104458	3597	3591	0.13	2.37	318.0	0.45	0.13	13.0
CMFE	8192	NPD	13	104332	4426	2534	0.13	2.26	300.5	0.34	0.07	12.5
FV	8192	SID	4	32512	1584	1589	0.05	2.07	607.0	0.17	0.04	8.5
NCFE	12160	SPD	5	60292	9768	6637	0.08	1.97	710.0	0.30	0.14	14.0

Table 4: Matrix properties and computational cost of the different equivalent formulations of the nonconforming finite element method, coefficients (5.1), mesh C

the direct solver performance in this case is improved. NCFE leads here to the worst-conditioned matrices, although the gap to the other methods is not very important. This definitely influences the CPU times of the iterative solvers. Altogether, speed-ups up to a factor of 2 can be achieved by the equivalent reformulations of the original NCFE method.

5.1.3 Inhomogeneous diffusion tensor

We consider here $\Omega = (-1, 1) \times (-1, 1)$, divided into four subdomains Ω_i corresponding to the axis quadrants (in the counterclockwise direction). The focus is on an inhomogeneous diffusion tensor given by

$$\underline{\mathbf{S}}|_{\Omega_i} = s\mathbb{I}, \, i \in 1, 3, \quad \underline{\mathbf{S}}|_{\Omega_i} = \mathbb{I}, \, i \in 2, 4. \tag{5.2}$$

Inhomogeneous Dirichlet boundary conditions are imposed so that the weak solution has a singularity at the origin. We refer to [11] for more details. We test the different methods on the fourth level of uniform refinement of the mesh D of the left part of Figure 10. We report the results for two cases, $s = 100$ in Table 5 and $s = 10000$ in Table 6.

The most important feature is that the condition number of the MFEB, MFEC, and MFEO approaches are almost completely insensitive to the jump of the diffusion tensor, similarly to the behavior of the condition number with respect to the anisotropy observed in Section 5.1.1. Consequently, the CPU times of the iterative solvers are up to two orders of magnitude faster compared to CMFE, FV, and NCFE; the MFEC formulation, thanks to its sparsity pattern, behaves particularly well. Using the diagonal scaling, once again as in Section 5.1.1, brings the condition number of all approaches to the same level. Consequently, preconditioned iterative methods behave similarly to the homogeneous isotropic case, which also seems to be true for the direct solver. Here the MFEC and FV approaches are superior compared to the other ones. Overall, speed-ups over the NCFE in the range from 2 to 30 in the extreme case can be achieved.

5.1.4 Inhomogeneous and anisotropic diffusion tensor

We finally consider a test case inspired from [18]. Let $\Omega = (0, 1) \times (0, 1)$. We focus on an inhomogeneous and anisotropic diffusion tensor given by

$$\underline{\mathbf{S}}_K = \begin{pmatrix} \cos(\theta_K) & -\sin(\theta_K) \\ \sin(\theta_K) & \cos(\theta_K) \end{pmatrix} \begin{pmatrix} s_K & 0 \\ 0 & \nu s_K \end{pmatrix} \begin{pmatrix} \cos(\theta_K) & \sin(\theta_K) \\ -\sin(\theta_K) & \cos(\theta_K) \end{pmatrix} \text{ for } K \in \mathcal{T}_h,$$
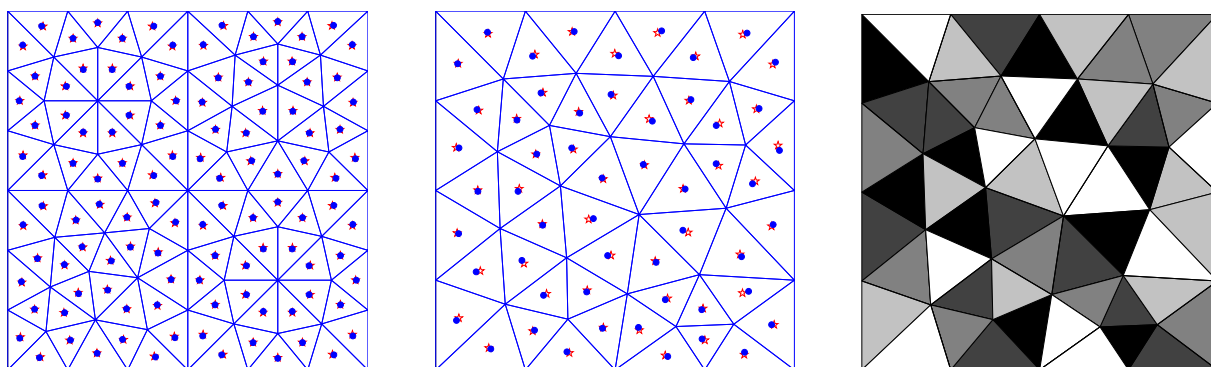
18

Figure 10: Mesh D (left) and mesh E (middle) with barycenters (pentagrams) and optimal evaluation points (circles); mesh E (right) with the coloring indicating the different cases of (5.3)

							DS	IS		PIS		
Meth.	Un.	Mat.	St.	Nonz.	CN	CNS	CPU	CPU	Iter.	CPU	Prec.	Iter.
MFEB	28672	NNS	15	363141	21128	21036	0.69	13.18	407.5	1.90	0.50	14.0
MFEC	28672	NNS	4	114304	26202	22813	0.19	5.31	439.5	1.12	0.39	11.5
MFEO	28672	NNS	15	363141	21150	20784	0.70	14.24	442.0	2.05	0.44	15.5
CMFE	28672	NPD	15	363928	389347	15004	0.70	44.45	1356.5	2.00	0.47	15.0
FV	28672	SPD	4	114304	957466	17117	0.18	18.87	3167.0	0.97	0.44	17.0
NCFE	42816	SPD	5	213312	613276	20023	0.36	34.19	2822.0	1.89	0.82	19.0

Table 5: Matrix properties and computational cost of the different equivalent formulations of the nonconforming finite element method, coefficients (5.2), mesh D, $s = 100$

where we set

$$s_K \in \{10, 5, 1, 0.5, 0.1\}, \quad \theta_K \in \left\{\frac{\pi}{5}, \frac{3\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{5}, \frac{\pi}{3}\right\}, \quad \nu = 0.2. \tag{5.3}$$

							DS	IS		PIS		
Meth.	Un.	Mat.	St.	Nonz.	CN	CNS	CPU	CPU	Iter.	CPU	Prec.	Iter.
MFEB	28672	NNS	15	363141	21367	21266	0.69	13.22	407.5	2.01	1.10	8.0
MFEC	28672	NNS	4	114304	26859	23108	0.20	5.21	432.5	1.42	0.73	10.0
MFEO	28672	NNS	15	363141	21404	21010	0.70	13.20	405.0	2.14	1.14	8.5
CMFE	28672	NPD	15	363928	3.84e+7	15207	0.72	575.68	17544.5	2.00	0.63	13.0
FV	28672	SPD	4	114304	9.43e+7	17439	0.18	82.48	14042.0	0.90	0.36	17.0
NCFE	42816	SPD	5	213312	6.04e+7	20326	0.36	158.13	13091.0	1.77	0.66	20.0

Table 6: Matrix properties and computational cost of the different equivalent formulations of the nonconforming finite element method, coefficients (5.2), mesh D, $s = 10000$

19

| | | | | | | | DS | IS | | PIS | | |
Meth.	Un.	Mat.	St.	Nonz.	CN	CNS	CPU	CPU	Iter.	CPU	Prec.	Iter.
MFEB	13824	NID	14	177652	78637	49183	0.26	8.91	586.5	1.01	0.56	7.5
MFEC	13824	NID	4	55040	4439957	1345919	0.09	—	—	0.42	0.20	6.0
MFEO	13824	NID	14	177652	65729	47337	0.27	8.97	590.5	1.02	0.64	5.5
CMFE	13824	NID	14	177652	72329	26573	0.27	14.04	923.5	1.01	0.53	8.0
FV	13824	SID	4	55040	2149267	520164	0.09	—	—	0.41	0.20	5.5
NCFE	20608	SPD	5	102528	186537	56933	0.18	11.04	2338.0	1.20	0.65	18.0

Table 7: Matrix properties and computational cost of the different equivalent formulations of the nonconforming finite element method, coefficients (5.3), mesh E

Inhomogeneous Dirichlet boundary conditions are imposed by the function $p(x, y) = 0.1y + 0.9$. A sink term $g = -0.001$ is prescribed on two elements of the initial mesh. Here no analytical solution is given. We perform the computations on the fourth-level uniform refinement of the mesh E, see the middle part of Figure 10. The different grey shades in Figure 10, right, correspond, with increasing intensity, to the different choices s_K and θ_K in (5.3). Note that herein the optimal evaluation points differ more significantly from the barycenters than in the mesh D of Section 5.1.3. We report the results in Table 7.

The condition numbers of the MFEC and FV approaches are in this case highly increased compared to the other approaches. Preconditioned iterative solvers or the direct solver allow to overcome this difficulty, but the unpreconditioned iterative solvers do not converge in 50000 iterations. This nonconvergence is obviously also linked to the fact that the NCFE formulation is the only one giving positive definite matrices; all other approaches lead to indefinite matrices. Still the present alternative formulations, as well as the FV one, allow to solve the linear system 1.2- to 3-times faster.

5.2 Nonlinear parabolic problems

We now advance to the nonlinear parabolic setting of Section 4. We consider two test cases taken from [14].

5.2.1 A reaction–diffusion problem

We first consider the nonlinear reaction–diffusion problem

$$\frac{\partial(p + p^\alpha)}{\partial t} - \nabla \cdot (\underline{\mathbf{S}} \nabla p) + 3p + \alpha p^\alpha = 0 \qquad \text{in } \Omega \times (0, T) \tag{5.4}$$

with $\Omega = (0, 2) \times (0, 1)$, $T = 1$, $\alpha = 0.5$, and either

$$\underline{\mathbf{S}} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ in } \Omega \tag{5.5}$$

or

$$\underline{\mathbf{S}} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ for } x < 1, \quad \underline{\mathbf{S}} = \begin{pmatrix} 0.75 & 0.25 \\ 0.25 & 10 \end{pmatrix} \text{ for } x > 1. \tag{5.6}$$

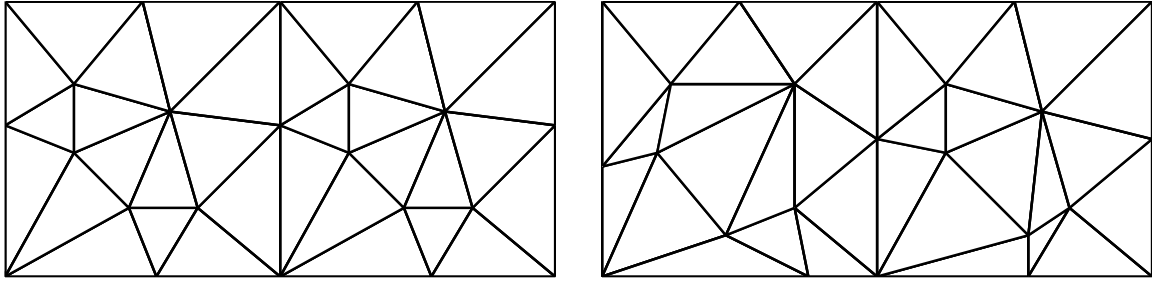Dirichlet boundary and initial conditions are given by the exact solution $p(x, y, t) = e^x e^y e^{-t}/e^3$.

20

Figure 11: Mesh F (left) and mesh G (right)

							DS	IS		PIS		
Meth.	Un.	Mat.	St.	Nonz.	CN	CNS	CPU	CPU	Iter.	CPU	Prec.	Iter.
MFEB	32768	NPD	14	405782	1664	1674	0.80	7.34	202.5	1.73	0.52	12.5
MFEC	32768	NPD	4	130688	2041	2034	0.26	3.60	227.0	0.71	0.43	3.0
FV–NCFE	48960	SPD	5	244032	3421	2119	0.42	6.93	484.0	2.22	0.46	35.0

Table 8: Matrix properties and computational cost of the different equivalent formulations of the FV–NCFE method, reaction–diffusion problem (5.4), first time and linearization steps, coefficients (5.5), mesh F

For the discretization of (5.4), we employ the combined finite volume–nonconforming finite element method (FV–NCFE) of [7] (recall that it represents the nonconforming finite element method with mass lumping and numerical quadrature) with Newton linearization. On each time step n and linearization step k, we obtain a linear algebraic system of the form (4.3) that we reformulate as (4.6) as described in Section 4. We consider the evaluation points $\mathbf{z}_K$ either as barycenters of the elements (MFEB) or as $\underline{\mathbf{S}}$-circumcenters (MFEC). The MFEC approach still leads here to a very narrow 4-point stencil. We note that the FV approach of [19, 4, 18] cannot directly be used here.

In Tables 8 and 9, we compare the properties of the system matrices $\mathbb{Z}^{\mathrm{i},n,k}$ and $\mathbb{S}^{n,k}$ and the computational costs of (4.3) and (4.6) for the first time step and the first Newton linearization step. Table 8 does so for the coefficients (5.5) and the mesh F of Figure 11, whereas Table 9 for the coefficients (5.6) and the mesh G of Figure 11. As in the linear elliptic cases, the MFEB and MFEC approaches give matrices with quite low condition number, which can hardly be improved by diagonal scaling for the isotropic diffusion–dispersion tensor $\underline{\mathbf{S}}$. The MFEC approach allows a speed-up by a factor of roughly two for the direct solver and three for the preconditioned iterative solver. The MFEB approach allows an important speed-up for the preconditioned iterative solver but not for the direct one or the unpreconditioned iterative one.

5.2.2 An advection–diffusion–reaction problem

We finally consider the nonlinear advection–diffusion–reaction problem

$$\frac{\partial(p + p^\alpha)}{\partial t} - \nabla \cdot (\underline{\mathbf{S}}\nabla p) + \nabla \cdot (p\mathbf{w}) + \alpha p^\alpha = 0 \qquad \text{in } \Omega \times (0, T) \tag{5.7}$$

21

Meth.	Un.	Mat.	St.	Nonz.	CN	CNS	DS CPU	IS CPU	Iter.	PIS CPU	Prec.	Iter.
MFEB	32768	NNS	14	414304	21101	9279	0.80	26.43	726.5	2.31	1.02	10.5
MFEC	32768	NID	4	130688	110496	112764	0.26	—	—	0.95	0.57	4.0
FV–NCFE	48960	SPD	5	244032	46527	9590	0.42	22.84	1686.0	3.02	1.51	22.0

Table 9: Matrix properties and computational cost of the different equivalent formulations of the FV–NCFE method, reaction–diffusion problem (5.4), first time and linearization steps, coefficients (5.6), mesh G

with $\Omega = (0,2) \times (0,1)$, $T = 1$, $\alpha = 0.5$, and either

$$\underline{\mathbf{S}} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ in } \Omega\,, \quad \mathbf{w} = (3,0) \text{ in } \Omega \tag{5.8}$$

or

$$\underline{\mathbf{S}} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ for } x < 1\,, \quad \underline{\mathbf{S}} = \begin{pmatrix} 8 & -7 \\ -7 & 20 \end{pmatrix} \text{ for } x > 1\,,$$
$$\mathbf{w} = (3,0) \text{ for } x < 1\,, \quad \mathbf{w} = (3,12) \text{ for } x > 1\,. \tag{5.9}$$

Dirichlet boundary and initial conditions are as before given by the exact solution $p(x,y,t) = e^x e^y e^{-t}/e^3$.

For the discretization of (5.7), we again employ the combined finite volume–nonconforming finite element method (FV–NCFE) of [7] (local Péclet upstream weighting is used for the stabilization of the advective term). We use Newton linearization and proceed as explained in Section 4. Only the evaluation points $\mathbf{z}_K$ given by barycenters (MFEB) are considered; the MFEC leads here to much wider stencil (equal to that of MFEB) because of the presence of the advective term.

In Tables 10 and 11, we compare the properties of the system matrices $\mathbb{Z}^{i,n,k}$ and $\mathbb{S}^{n,k}$ and the computational costs of (4.3) and (4.6) for the first time step and the first Newton linearization step. Table 10 does so for the coefficients (5.8) and the mesh F of Figure 11, whereas Table 11 for the coefficients (5.9) and the mesh G of Figure 11. Here again, the MFEB approach gives matrices with quite low condition number, which can hardly be improved by diagonal scaling. The CPU times of the direct solver are similar as in the previous cases. Concerning the iterative solvers, one notable difference is that the FV–NCFE method now leads to nonsymmetric matrices because of the advective term; thus the Bi-CGStab solver is used in place of the CG one. In the preconditioned case and for coefficients (5.9) also in the unpreconditioned one, roughly 1.5 times faster CPU times can be achieved for the MFEB formulation with respect to the original one.

6   Concluding remarks

We have introduced a new local static condensation from face to element unknowns and we have applied it to the Crouzeix–Raviart nonconforming finite element method.

It appears that the matrix condition numbers of some of the equivalent reformulations of the nonconforming finite element method developed (MFEO in particular) are insensitive to the anisotropy of the mesh and to the inhomogeneity of the diffusion tensor. This seems to be a built-in property, whereas a similar result for the original matrix of the nonconforming finite element method can only be achieved upon employing a diagonal scaling.

							DS	IS		PIS		
Meth.	Un.	Mat.	St.	Nonz.	CN	CNS	CPU	CPU	Iter.	CPU	Prec.	Iter.
MFEB	32768	NPD	14	422894	1697	1708	0.89	9.40	251.5	1.83	0.79	9.5
FV–NCFE	48960	NPD	5	244032	3483	2158	0.44	8.74	315.0	2.84	1.50	10.0

Table 10: Matrix properties and computational cost of the different equivalent formulations of the FV–NCFE method, advection–diffusion–reaction problem (5.7), first time and linearization steps, coefficients (5.8), mesh F

							DS	IS		PIS		
Meth.	Un.	Mat.	St.	Nonz.	CN	CNS	CPU	CPU	Iter.	CPU	Prec.	Iter.
MFEB	32768	NNS	14	417218	26317	20610	0.89	29.37	794.5	2.27	1.32	8.5
FV–NCFE	48960	NPD	5	244032	104836	18160	0.44	42.42	1471.5	3.35	1.62	12.5

Table 11: Matrix properties and computational cost of the different equivalent formulations of the FV–NCFE method, advection–diffusion–reaction problem (5.7), first time and linearization steps, coefficients (5.9), mesh G

Amongst the different equivalent reformulations of the nonconforming finite element method, the MFEC approach proposed in this paper and the FV approach of [19, 4, 18] are of special interest and due to their sparsity structure especially appealing for the use with direct solvers. However, they are not applicable for advection–diffusion problems and only on very special meshes in three-space dimensions; however, the MFEC can in contrast to FV easily be applied to unsteady nonlinear reaction–diffusion problems. They also lead to quite badly conditioned matrices for anisotropic diffusion tensors. Alternalively, the MFEB approach is very robust throughout all different cases tested, including anisotropic tensors.

The static condensation principle developed in this paper is quite general and is not restricted to a particular model problem or discretization scheme. In particular, its application to the different equivalent reformulations of the nonconforming finite element method give an appealing alternative to the classical formulation, enable to influence the final matrix properties, lead to (quite important) computational savings for only a marginally more expensive setup which may involve inversion of local linear systems, and offer a nontraditional viewpoint on the nonconforming finite element method.

A A general static condensation principle

The purpose of this Appendix is to generalize the approach of Section 2, allowing in particular further reduction of the number of unknowns and relaxing assumption (2.15a). We do the presentation in an abstract framework, not (necessarily) related to the sets such as the mesh elements, faces, and vertices; as an example, this approach allows to reduce the system (1.1) to one unknown per a set of elements (typically a couple of simplices forming a quadrilateral, a parallelepiped, or a general polygon/polyhedron). As such, it gives rise to a discretization scheme leading to *one unknown per element* on *arbitrary polygonal/polyhedral meshes*, see also the discussion and references

in [17].

A.1 Setting

Let $\mathcal{T}_h$ be a simplicial mesh of the domain Ω as specified in Section 2.1. Let $\mathbb{Z}^i \in \mathbb{R}^{|\mathcal{E}_h^i| \times |\mathcal{E}_h^i|}$ be a given matrix and $E \in \mathbb{R}^{|\mathcal{E}_h^i|}$ a given vector. In contrast to Section 2.2, we do not need to assume here anything about the emplacement of the nonzero entries of the matrix $\mathbb{Z}^i$, we merely suppose that (1.2) is well-posed. Let $\bar{\mathcal{T}}_h$ be a given set; $\bar{\mathcal{T}}_h$ can be the set of mesh elements $\mathcal{T}_h$ as in Section 2, but this is not necessary. Typically, the elements of $\bar{\mathcal{T}}_h$ are sets of the simplices from $\mathcal{T}_h$. We define one new unknown P_K for each $K \in \bar{\mathcal{T}}_h$ and let P be the corresponding algebraic vector, $P = \{P_K\}_{K \in \bar{\mathcal{T}}_h}$. We show how (1.2) can be equivalently reduced to (1.3) with $\mathbb{S} \in \mathbb{R}^{|\bar{\mathcal{T}}_h| \times |\bar{\mathcal{T}}_h|}$.

Let $\mathbb{N}^i$ be a $|\bar{\mathcal{T}}_h| \times |\mathcal{E}_h^i|$ matrix, *completely arbitrary*. Consider the following analogy of (1.1), (2.1): find $\Lambda^i \in \mathbb{R}^{|\mathcal{E}_h^i|}$ and $P \in \mathbb{R}^{|\bar{\mathcal{T}}_h|}$ such that (1.2) and

$$\mathbb{N}^i \Lambda^i = P \tag{A.1}$$

hold together. Remark that (1.2), (A.1) is well-posed.

A.2 Definition of the local problems

Consider now a new set $\bar{\mathcal{V}}_h$; $\bar{\mathcal{V}}_h$ can be the set of the mesh vertices $\mathcal{V}_h$ as in Section 2, but this is not necessary. With each $V \in \bar{\mathcal{V}}_h$, we associate a set of faces from $\mathcal{E}_h^i$ denoted by $\bar{\mathcal{E}}_V^{\text{int}}$ and a set of elements of $\bar{\mathcal{T}}_h$ denoted by $\bar{\mathcal{T}}_V$. Again, these sets can be the same as $\mathcal{E}_V^{\text{int}}$ and $\mathcal{T}_V$ in Section 2.1, see Figures 1–3, but this is not required. We only suppose that each $K \in \bar{\mathcal{T}}_h$ belongs to at least one set $\bar{\mathcal{T}}_V$ and that each $\sigma \in \mathcal{E}_h^i$ belongs to at least one set $\bar{\mathcal{E}}_V^{\text{int}}$.

Let $V \in \bar{\mathcal{V}}_h$. We now consider the lines from (1.2) associated with the faces from $\bar{\mathcal{E}}_V^{\text{int}}$ and the lines from (A.1) associated with the elements from $\bar{\mathcal{T}}_V$. This gives rise to the following *local linear system*, an analogy of (2.2):

$$\left(\begin{array}{c} \mathbb{Z}_V \\ \mathbb{N}_V \end{array} \right) \Lambda_V = \left(\begin{array}{c} E_V \\ P_V \end{array} \right). \tag{A.2}$$

Denote by $\bar{\mathcal{E}}_V$ the faces of $\mathcal{E}_h^i$ corresponding to the unknowns Λ_V. In order to proceed further, we need to assume at this point that $|\bar{\mathcal{E}}_V| = |\bar{\mathcal{E}}_V^{\text{int}}| + |\bar{\mathcal{T}}_V|$, so that the system matrix of (A.2) is square; we also suppose that it is nonsingular. Then (A.2) enables to *express locally* the unknowns Λ_V from P_V, considered as parameters. In contrast to (2.4), (A.2) is of bigger size, which eventually will lead to a wider stencil of the final matrix $\mathbb{S}$. On the other hand, we do not need here to elaborate on the block structure (2.3) and to ensure that the matrices $\mathbb{N}_V^{\text{ext}}$ in (2.3) are square and diagonal.

A.3 The reduced system

We now proceed similarly to Section 2.6. Keep the developments and notation (2.7)–(2.9), where we merely replace $\mathcal{E}_V^{\text{int}}$ by $\bar{\mathcal{E}}_V$. As for the matrix mapping Υ_V, change it into $\Upsilon_V : \mathbb{R}^{|\bar{\mathcal{E}}_V| \times (|\bar{\mathcal{E}}_V^{\text{int}}| + |\bar{\mathcal{T}}_V|)} \to \mathbb{R}^{|\mathcal{E}_h^i| \times (|\mathcal{E}_h^i| + |\bar{\mathcal{T}}_h|)}$, extending a local matrix $\mathbb{M}_V$ to a full-size one by zeros by

$$[\Upsilon_V(\mathbb{M}_V)]_{\sigma, \gamma K} := \left\{ \begin{array}{ll} (\mathbb{M}_V)_{\sigma, \gamma K} & \text{if } \sigma \in \bar{\mathcal{E}}_V \text{ and } \gamma K \in \bar{\mathcal{E}}_V^{\text{int}} \cup \bar{\mathcal{T}}_V \\ 0 & \text{if } \sigma \notin \bar{\mathcal{E}}_V \text{ or } \gamma K \notin \bar{\mathcal{E}}_V^{\text{int}} \cup \bar{\mathcal{T}}_V \end{array} \right. .$$

Inverting the system matrix in (A.2), introducing the weights matrices $\mathbb{W}_V$, summing over all $V \in \bar{\mathcal{V}}_h$, setting

$$\mathbb{M}^{\text{inv}} := \sum_{V \in \bar{\mathcal{V}}_h} \Upsilon_V \left(\mathbb{W}_V \left(\begin{array}{c} \mathbb{Z}_V \\ \mathbb{N}_V \end{array} \right)^{-1} \right), \tag{A.3}$$

and using the analogy of (2.9), we obtain

$$\Lambda^{i} = \mathbb{M}^{\mathrm{inv}} \begin{pmatrix} E \\ P \end{pmatrix}. \tag{A.4}$$

Using the block form of $\mathbb{M}^{\mathrm{inv}}$, $\mathbb{M}^{\mathrm{inv}} = (\mathbb{M}_{E}^{\mathrm{inv}}, \mathbb{M}_{P}^{\mathrm{inv}})$, and plugging (A.4) into (A.1), we arrive at

$$(\mathbb{N}^{i}\mathbb{M}_{P}^{\mathrm{inv}} - \mathbb{I})P = -\mathbb{N}^{i}\mathbb{M}_{E}^{\mathrm{inv}}E, \tag{A.5}$$

i.e., at a system of the form (1.3) with $\mathbb{S} := \mathbb{N}^{i}\mathbb{M}_{P}^{\mathrm{inv}} - \mathbb{I}$ and $H := -\mathbb{N}^{i}\mathbb{M}_{E}^{\mathrm{inv}}E$. Once we solve (A.5) for P, we can obtain Λ^{i} locally from (A.2) or from (A.4).

A.4 Equivalence of the reduced system with the original one

As in Section 2.7, we show here that, under a weakening of Assumption 2.2, (A.5) is well-posed and equivalent to (1.2).

Let $\sigma \in \mathcal{E}_{h}^{i}$ be fixed and denote by $\bar{\mathcal{E}}_{\sigma}$ the set of all faces $\gamma \in \mathcal{E}_{h}^{i}$ appearing in the various $\bar{\mathcal{E}}_{V}$, $\bar{\mathcal{E}}_{\sigma} := \cup_{V; \sigma \in \bar{\mathcal{E}}_{V}} \bar{\mathcal{E}}_{V}$. Denote also by $\bar{\mathcal{T}}_{\sigma}$ the set of all elements $K \in \bar{\mathcal{T}}_{h}$ appearing in the corresponding $\bar{\mathcal{T}}_{V}$, $\bar{\mathcal{T}}_{\sigma} := \cup_{V; \sigma \in \bar{\mathcal{E}}_{V}} \bar{\mathcal{T}}_{V}$. Similarly to (A.2), for a fixed $\sigma \in \mathcal{E}_{h}^{i}$, consider the lines from (1.2) associated with the faces from $\bar{\mathcal{E}}_{\sigma}$ and the lines from (A.1) associated with the elements from $\bar{\mathcal{T}}_{\sigma}$, giving

$$\begin{pmatrix} \mathbb{Z}_{\Sigma} \\ \mathbb{N}_{\Sigma} \end{pmatrix} \Lambda_{\Sigma} = \begin{pmatrix} E_{\Sigma} \\ P_{\Sigma} \end{pmatrix} \tag{A.6}$$

with obvious notation. Remark that (A.6) is similar to (2.14) but is larger (as the sets $\bar{\mathcal{E}}_{\sigma}$ and $\bar{\mathcal{T}}_{\sigma}$ are bigger here). Remark also that (A.2) form subsystems of (A.6).

Assumption A.1 (Local well-posedness properties). *The mesh $\mathcal{T}_{h}$, the matrix $\mathbb{Z}^{i}$, the matrix $\mathbb{N}^{i}$, and the sets $\bar{\mathcal{E}}_{V}^{\mathrm{int}}$, $\bar{\mathcal{E}}_{V}$, and $\bar{\mathcal{T}}_{V}$ in (A.2) and the sets $\bar{\mathcal{E}}_{\sigma}$ and $\bar{\mathcal{T}}_{\sigma}$ in (A.6) are such that:*

$$\text{each } K \in \bar{\mathcal{T}}_{h} \text{ belongs to at least one set } \bar{\mathcal{T}}_{V}; \tag{A.7a}$$

$$\text{each } \sigma \in \mathcal{E}_{h}^{i} \text{ belongs to at least one set } \bar{\mathcal{E}}_{V}^{\mathrm{int}}; \tag{A.7b}$$

$$\begin{pmatrix} \mathbb{Z}_{V} \\ \mathbb{N}_{V} \end{pmatrix} \text{ is square and nonsingular} \quad \forall V \in \bar{\mathcal{V}}_{h}; \tag{A.7c}$$

$$\begin{pmatrix} \mathbb{Z}_{\Sigma} \\ \mathbb{N}_{\Sigma} \end{pmatrix} \text{ has at least as many rows as columns and a full rank} \quad \forall \sigma \in \mathcal{E}_{h}^{i}. \tag{A.7d}$$

The following three results are equivalents of Theorems 2.3, 2.5 and of Corollary 2.4. They can be proven analogously as in Section 2.7.

Theorem A.2 (Equivalence of (1.2), (A.1) and of (A.5), (A.4)). *Let Assumption A.1 hold. Let (Λ^{i}, P) be the solution of (1.2), (A.1). Then P is also a solution of (A.5) and Λ^{i} the solution of (A.4). Conversely, let P be a solution of (A.5) and let Λ^{i} be given by (A.4). Then the couple (Λ^{i}, P) is the solution of (1.2), (A.1).*

Corollary A.3 (Well-posedness of (A.5)). *Let Assumption A.1 hold. Then there exists one and only one solution P to (A.5).*

Theorem A.4 (Stencil of the system matrix of (A.5)). *Let Assumption A.1 hold. Let $K \in \bar{\mathcal{T}}_{h}$. Consider the row of $\mathbb{N}^{i}$ associated with K and denote by $\bar{\mathcal{E}}_{K}$ the faces $\sigma \in \mathcal{E}_{h}^{i}$ corresponding to nonzero entries on this row of $\mathbb{N}^{i}$. Then, on a row of the system matrix $\mathbb{S} = \mathbb{N}^{i}\mathbb{M}_{P}^{\mathrm{inv}} - \mathbb{I}$ of (A.5) associated with this K, only possible nonzero entries are on columns associated with elements of $\bar{\mathcal{T}}_{h}$ from the set $\cup_{\sigma \in \bar{\mathcal{E}}_{K}} \cup_{V; \sigma \in \bar{\mathcal{E}}_{V}} \bar{\mathcal{T}}_{V}$. Thus, in particular, when $\mathbb{Z}^{i}$ and $\mathbb{N}^{i}$ are sparse, then $\mathbb{S}$ is also sparse.*

We conclude by the two following remarks:

Remark A.5 (Approaches of Section 2 and of Appendix A)**.** *Comparing Theorem 2.5 with Theorem A.4, we see that the disadvantage of the approach of Appendix A is that the system matrix $\mathbb{S}$ has, a priori, a much wider stencil than that of Section 2. Note, however, that the approach of Appendix A is much more general and that the indicated stencils may further reduce in particular circumstances, cf. Remark 3.5.*

Remark A.6 (Further generalization of Assumption A.1)**.** *It appears that the matrix $\left(\begin{smallmatrix} \mathbb{Z}_V \\ \mathbb{N}_V \end{smallmatrix} \right)$ in (A.7c) needs not to be square and nonsingular. When (A.7d) holds true, there is still a way to equivalently reduce (1.2) to (1.3). Such an approach is studied in [16].*

References

[1] AAVATSMARK, I., BARKVE, T., BØE, Ø., AND MANNSETH, T. Discretization on unstructured grids for inhomogeneous, anisotropic media. I. Derivation of the methods. *SIAM J. Sci. Comput. 19*, 5 (1998), 1700–1716.

[2] ARBOGAST, T., AND CHEN, Z. On the implementation of mixed methods as nonconforming methods for second-order elliptic problems. *Math. Comp. 64*, 211 (1995), 943–972.

[3] ARNOLD, D. N., AND BREZZI, F. Mixed and nonconforming finite element methods: implementation, postprocessing and error estimates. *RAIRO Modél. Math. Anal. Numér. 19*, 1 (1985), 7–32.

[4] CHAVENT, G., YOUNÈS, A., AND ACKERER, P. On the finite volume reformulation of the mixed finite element method for elliptic and parabolic PDE on triangles. *Comput. Methods Appl. Mech. Engrg. 192*, 5-6 (2003), 655–682.

[5] CIARLET, P. G. *The Finite Element Method for Elliptic Problems*, vol. 4 of *Studies in Mathematics and its Applications*. North-Holland, Amsterdam, 1978.

[6] CROUZEIX, M., AND RAVIART, P.-A. Conforming and nonconforming finite element methods for solving the stationary Stokes equations. I. *Rev. Française Automat. Informat. Recherche Opérationnelle Sér. Rouge 7*, R-3 (1973), 33–75.

[7] EYMARD, R., HILHORST, D., AND VOHRALÍK, M. A combined finite volume–nonconforming/mixed-hybrid finite element scheme for degenerate parabolic problems. *Numer. Math. 105*, 1 (2006), 73–131.

[8] HESTENES, M. R., AND STIEFEL, E. Methods of conjugate gradients for solving linear systems. *J. Research Nat. Bur. Standards 49* (1952), 409–436 (1953).

[9] MARINI, L. D. An inexpensive method for the evaluation of the solution of the lowest order Raviart–Thomas mixed method. *SIAM J. Numer. Anal. 22*, 3 (1985), 493–496.

[10] RAVIART, P.-A., AND THOMAS, J.-M. A mixed finite element method for 2nd order elliptic problems. In *Mathematical aspects of finite element methods (Proc. Conf., Consiglio Naz. delle Ricerche (C.N.R.), Rome, 1975)*. Springer, Berlin, 1977, pp. 292–315. Lecture Notes in Math., Vol. 606.

[11] RIVIÈRE, B., WHEELER, M. F., AND BANAS, K. Part II. Discontinuous Galerkin method applied to single phase flow in porous media. *Comput. Geosci. 4*, 4 (2000), 337–349.

[12] SAAD, Y. *Iterative methods for sparse linear systems*, second ed. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2003.

[13] VAN DER VORST, H. A. Bi-CGSTAB: a fast and smoothly converging variant of Bi-CG for the solution of nonsymmetric linear systems. *SIAM J. Sci. Statist. Comput. 13*, 2 (1992), 631–644.

[14] VOHRALÍK, M. Equivalence between lowest-order mixed finite element and multi-point finite volume methods on simplicial meshes. *M2AN Math. Model. Numer. Anal. 40*, 2 (2006), 367–391.

[15] VOHRALÍK, M., MARYŠKA, J., AND SEVERÝN, O. Mixed and nonconforming finite element methods on a system of polygons. *Appl. Numer. Math. 57*, 2 (2007), 176–193.

[16] VOHRALÍK, M., AND WOHLMUTH, B. I. All order mixed finite element methods with one unknown per element. In preparation, 2012.

[17] VOHRALÍK, M., AND WOHLMUTH, B. I. Mixed finite element methods: implementation with one unknown per element, local flux expressions, positivity, polygonal meshes, and relations to other methods. *Math. Models Methods Appl. Sci.* (2012). Accepted for publication.

[18] YOUNÉS, A., ACKERER, P., AND CHAVENT, G. From mixed finite elements to finite volumes for elliptic PDEs in two and three dimensions. *Internat. J. Numer. Methods Engrg. 59*, 3 (2004), 365–388.

[19] YOUNÉS, A., MOSE, R., ACKERER, P., AND CHAVENT, G. A new formulation of the mixed finite element method for solving elliptic and parabolic PDE with triangular elements. *J. Comput. Phys. 149*, 1 (1999), 148–167.