A posteriori error estimates and adaptivity for locally conservative methods. Inexpensive implementation and evaluation, polytopal meshes, iterative linearization and algebraic solvers, and applications to complex porous media flows

Martin Vohralík, Soleiman Yousef

A posteriori error estimates and adaptivity for locally conservative methods

Inexpensive implementation and evaluation, polytopal meshes, iterative linearization and algebraic solvers, and applications to complex porous media flows

Martin Vohralík[†]   Soleiman Yousef[‡]

May 29, 2025

Abstract

A posteriori estimates give bounds on the error between the unknown solution of a partial differential equation and its numerical approximation. We present here the methodology based on H^1-conforming potential and $\boldsymbol{H}(\mathrm{div})$-conforming equilibrated flux reconstructions, where the error bounds are guaranteed and fully computable. We consider any lowest-order locally conservative method of the finite volume type and treat general polytopal meshes. We start by a pure diffusion problem and first address the discretization error. We then progressively pass to more complicated model problems and also take into account the errors arising in iterative linearization of nonlinear problems and in algebraic resolution of systems of linear algebraic equations. We focus on the ease of implementation and evaluation of the estimates. In particular, the evaluation of our estimates is explicit and inexpensive, since it merely consists in some local matrix-vector multiplications. Here, on each mesh element, the matrices are either directly inherited from the given numerical method, or easily constructed from the element geometry, while the vectors are the algebraic unknowns of the flux and potential approximations on the given element. Our last problem is a real-life unsteady nonlinear coupled degenerate advection–diffusion–reaction system describing a complex multiphase multicomponent flow in porous media. Here, on each step of the time-marching scheme, on each step of the iterative linearization procedure, and on each step of the linear algebraic solver, the estimate gives a guaranteed upper bound on the total intrinsic error, still takes the simple matrix-vector multiplication form, and distinguishes the different error components. It leads to an easy-to-implement and fast-to-run adaptive algorithm with guaranteed overall precision, adaptive stopping criteria for nonlinear and linear solvers, and adaptive space and time mesh refinements and derefinements. Progressively along the theoretical exposition, numerical experiments on academic benchmarks as well as on real-life problems in two and three space dimensions illustrate the performance of the derived methodology. The presentation is largely self-standing, developing all the details and recalling all necessary basic notions.

Keywords: Partial differential equation, numerical approximation, locally conservative method, finite volume method, mixed finite element method, mimetic finite difference method, mixed virtual element method, polytopal mesh, iterative linearization, iterative linear algebraic solver, a posteriori error estimate, error components, balancing, adaptive stopping criteria, adaptive mesh refinement, mass balance recovery, porous media, Darcy flow, multiphase multicomponent flow.

Contents

[†]Project-team SERENA, Inria Paris, 48 rue Barrault, 75647 Paris, France & CERMICS, Ecole nationale des ponts et chaussées, IP Paris, 77455 Marne la Vallée, France (martin.vohralik@inria.fr).
[‡]IFP Energies nouvelles, 1 & 4 av. Bois Préau, 92852 Rueil-Malmaison, France (soleiman.yousef@ifpen.fr).

1 Introduction

The purpose of this contribution is to present the theory of *a posteriori error estimates* and *adaptivity* in numerical approximation of partial differential equations (PDEs) by locally conservative methods of the *finite volume* type. We focus on estimates that *certify* the *error* and can be *easily coded*, *cheaply evaluated*, and *efficiently used*, even in complex *practical simulations*. We consider *polytopal meshes*, i.e., meshes formed by general polygonal or polyhedral elements, appealing in various applications. We cover

the overall chain of computational practice, including *iterative linearization* of nonlinear problems and *algebraic resolution* of systems of linear algebraic equations.

An *optimal* a posteriori estimate can be described by a set of requested properties, including namely the bound on the error between the unknown solution of the partial differential equation and the available numerical approximation that is *guaranteed* and *fully computable* from the approximate solution. We introduce this concept in Section 2 at an abstract level. We start from the notion of a PDE, discuss some intrinsic physical and mathematical properties of the weak solutions of PDEs, the notion of a total error and its components, and of adaptivity which aims at balancing these error components. One particular property of the methodology we develop here is that it leads to (exact or approximate) mass balance recovery at any step of numerical resolution, namely during the iteration of the nonlinear and linear solvers.

This contribution is made as self-standing as possible. For this purpose, in Section 3, we set up the notation, recall the intrinsic function spaces related to the considered partial differential equations, namely the infinite-dimensional *Sobolev spaces* $H_0^1(\Omega)$ and $\boldsymbol{H}(\mathrm{div}, \Omega)$, and recall their finite-dimensional subspaces composed of *piecewise polynomials* with respect to a simplicial mesh. We also describe there the polytopal meshes we consider and introduce the basic principle of the finite volume and related lowest-order locally conservative methods.

The heart of our exposition starts in Section 4, where we treat in detail the model steady linear diffusion problem with homogeneous and isotropic diffusion tensor (the *Poisson equation*). Only the basic cell-centered finite volume discretization on a simplicial mesh and only the *spatial discretization* error are investigated here, to expose the construction principles as clearly as possible. We namely motivate and present the methodology based on $H^1(\Omega)$-conforming potential and $\boldsymbol{H}(\mathrm{div}, \Omega)$-conforming equilibrated flux reconstructions, which yields a guaranteed and fully computable error upper bound. These reconstructions are practically obtained here in piecewise polynomial spaces. This is perfectly fine for academic purposes and has an asymptotically optimal evaluation cost, since it is in particular explicit and no local problems need to be solved. A cheaper implementation and evaluation of a posteriori error estimates is, however, possible and suitable for more complicated model or real-life problems, which we treat later. Numerical experiments illustrate these basic developments.

In Section 5, we extend the above analysis to tensor-valued steady linear diffusion problems (the *singlephase steady linear Darcy flow*), general polytopal meshes, and any lowest-order locally conservative method such as mixed finite elements, mixed and hybrid finite volumes, mimetic finite differences, mixed virtual finite elements or hybrid high-order methods. In particular, though we still rely on the notions of potential and equilibrated flux reconstructions, there become *virtual* here, since they are not constructed in practice anymore. Similarly, though we rely on a notion of a simplicial submesh of a general polytopal mesh, this is also virtual, where the physical construction and computer implementation of a simplicial submesh is avoided. In particular, the evaluation of our estimates merely consists in some local matrix-vector multiplications, where, on each mesh element, the matrices are either directly inherited from the given numerical method, or easily constructed from the element geometry, while the vectors are the algebraic unknowns of the flux and potential approximations on the given element. This gives an easy and practically accessible application to polytopal meshes of the general methodology of H^1-conforming potential reconstruction and $\boldsymbol{H}(\mathrm{div}, \Omega)$-conforming flux reconstruction. Numerical experiments focus on general polygonal meshes and illustrate the performance of this methodology.

In Section 6, we extend the previous developments to encompass two cornerstones of numerical approximation in applications: the *nonlinear* and *linear solvers*. Our problem here is the *singlephase steady nonlinear Darcy flow*, where an *iterative linearization* is applied and where the *algebraic resolution* of the system of arising linear algebraic equations is addressed. We still derive a guaranteed upper bound on the total error. Moreover, we identify the three arising *error components*, related to discretization, linearization, and algebraic solver, and design *adaptive stopping criteria* for the iterative solvers. Remarkably, the estimators still take the form of a *simple matrix-vector multiplication*, with the same local matrices as in the linear case. Numerical illustrations confirm the theoretical findings.

In the last part of this contribution, Section 7, we consider an unsteady nonlinear coupled degenerate advection–diffusion–reaction problem, the *multiphase compositional unsteady Darcy flow*. We apply the entire methodology developed above to obtain estimates that are still guaranteed and fully computable for the intrinsic error measure, apply to any lowest-order locally conservative method on a polytopal mesh, and take a simple matrix-vector multiplication form. Moreover, they are valid on each stage of the overall solution algorithm: on *each time* step n, each *linearization* step k, and each *linear solver step* i. They also allow to distinguish different error components and design *adaptive stopping criteria* for the solvers, as well as *adaptive choice* of *space* and *time meshes*. A comprehensive numerical study on

realistic porous media problems is included.

The present a posteriori error estimates can be readily implemented into numerical codes on general polytopal meshes, with a minimal overhead. These estimates allow for a very fast evaluation and, according to the main results summarized in Theorems 4.18, 5.18, 6.2, and 7.6, give a guaranteed control over the intrinsic error committed in the numerical approximation. Additionally, all the different error components (time and space discretizations, linearization, algebraic) are identified, leading to fully adaptive algorithms with all adaptive stopping criteria for linear and nonlinear solvers, adaptive time step management, and adaptive mesh refinement. Numerical experiments on real-life problems confirm important *computational speed-ups* that can be achieved with our methodology, in addition to the *certification of the computed output.*

Our presentation proceeds along the lines of Vohralík [249], Ern and Vohralík [124], and Vohralík and Yousef [255], using some central concepts for general polytopal meshes from Vohralík and Wohlmuth [254]. Ample bibliographic references age given in the dedicated sections below.

2 A posteriori error estimation and adaptivity – the principles

In this section, we set up the main ideas. We proceed informally; notation is in detail set in Section 3.

2.1 Partial differential equations

Let $\Omega \subset \mathbb{R}^d$, $d \geq 1$, be a polytopal domain where our problem is set. Consider an abstract (steady but possibly nonlinear) partial differential equation in the form: for a source term f, find a real-valued function $p : \Omega \to \mathbb{R}$ such that

$$\mathcal{A}(p) = f \qquad \text{in } \Omega, \tag{2.1a}$$

$$p = 0 \qquad \text{on } \partial\Omega. \tag{2.1b}$$

The prominent example that we will consider below is the Laplace operator

$$\mathcal{A}(p) = -\nabla{\cdot}(\nabla p) = -\Delta p = -\sum_{i=1}^{d} \partial^2_{\boldsymbol{x}_i} p. \tag{2.2}$$

In general, $\mathcal{A}(\cdot)$ is a (nonlinear) operator including *partial derivatives* of the unknown solution p.

More generally, for a final time $t_{\mathrm{F}} > 0$, a source term f, and an initial condition p_0, consider an abstract (unsteady nonlinear) partial differential equation in the form: find a real-valued function $p : \Omega \times (0, t_{\mathrm{F}}) \to \mathbb{R}$ such that

$$\mathcal{A}(p) = f \qquad \text{in } \Omega \times (0, t_{\mathrm{F}}), \tag{2.3a}$$

$$p = 0 \qquad \text{on } \partial\Omega \times (0, t_{\mathrm{F}}), \tag{2.3b}$$

$$p(\cdot, 0) = p_0 \qquad \text{in } \Omega. \tag{2.3c}$$

Here, $\mathcal{A}(\cdot)$ is a (nonlinear) operator including partial derivatives of p and namely a partial derivative with respect to the time, for example

$$\mathcal{A}(p) = \partial_t p - \nabla{\cdot}(\underline{\boldsymbol{K}}(|\nabla p|)\nabla p), \tag{2.4}$$

where $|\nabla p|$ is the Euclidean size of the vector ∇p and $\underline{\boldsymbol{K}}(\cdot)$ is a scalar- or matrix-valued *nonlinear function* of a real variable.

2.2 Physical and mathematical properties of the weak solutions of PDEs

In order to properly describe the solution p of (2.1) or (2.3), adequate mathematical *function spaces* are necessary. We develop this task in details in Section 3.2 below, introducing namely the Sobolev spaces $H_0^1(\Omega)$ and $\boldsymbol{H}(\mathrm{div}, \Omega)$. These spaces also reflect the *physical principles* behind problems (2.1) and (2.3). For instance, for the Laplace operator (2.2) in (2.1), it turns out that the solution p satisfies

$$p \in H_0^1(\Omega), \quad \boldsymbol{u} := -\nabla p \in \boldsymbol{H}(\mathrm{div}, \Omega), \quad \nabla{\cdot}\boldsymbol{u} = f. \tag{2.5}$$

This reflects that the scalar-valued field p (pressure, potential, primal variable) is *continuous* in the appropriate trace sense (cannot jump across interfaces), the normal trace of the vector-valued field $\boldsymbol{u} :=$

$-\nabla p$ (Darcy velocity, flux or flow field, dual variable) is *normal-component continuous* in the appropriate trace sense (what flows out from a subdomain through an interface flows in through this interface in the neighboring subdomain), the Darcy velocity is related to the pressure by the *constitutive law* $\boldsymbol{u} := -\nabla p$, and the Darcy velocity is in *equilibrium* with the load f as expressed by $\nabla \cdot \boldsymbol{u} = f$.

2.3 Numerical approximation: spatial discretization, temporal discretization, iterative linearization, and iterative linear algebraic resolution

It is in general not possible to find analytically the exact potential p and the exact flux $\boldsymbol{u}$ of problems like (2.1) or (2.3). Then, we want to find an approximation obtained by a mathematical algorithm implemented on a computer. Typically, we will take a subdivision of Ω into computational cells, a spatial mesh $\mathcal{T}_h$ of Ω, with a characteristic size h, and use some *spatial discretization* scheme. For (2.3), we additionally consider a subdivision of the time interval $(0, t_{\mathrm{F}})$ into time steps, a temporal mesh of $(0, t_{\mathrm{F}})$, with a characteristic size τ, and proceed to a *temporal discretization*. When a nonlinear function as $\underline{\boldsymbol{K}}(\cdot)$ in (2.4) is present, we typically need an *iterative linearization*, to, on each iteration step k, approximate a nonlinear operator by a linear one. Finally, for steady problems or implicit discretizations of unsteady problems, we need an *iterative linear algebraic solver* for an approximate solution of a (large sparse) system of linear algebraic equations, with an iteration index i. In this text, we focus on discretizations yielding *numerical approximations* to the exact flux $\boldsymbol{u}$ (rather than to the exact potential p) that we denote by $\boldsymbol{u}_h^{k,i}$ or $\boldsymbol{u}_{h\tau}^{k,i}$.

2.4 Total error and its components

In principle, the numerical approximation is not exact, so that $\boldsymbol{u}_h^{k,i} \neq \boldsymbol{u}$ for (2.1) and $\boldsymbol{u}_{h\tau}^{k,i} \neq \boldsymbol{u}$ for (2.3). Evaluating the difference $\boldsymbol{u} - \boldsymbol{u}_h^{k,i}$ or $\boldsymbol{u} - \boldsymbol{u}_{h\tau}^{k,i}$ in some norm then gives the notion of the *total error*. This is in our exposition composed of the spatial discretization error, temporal discretization error, iterative linearization error, and iterative linear algebraic solver error.[1]

2.4.1 Intrinsic error measure

In the choice of the norm to measure the distances $\boldsymbol{u} - \boldsymbol{u}_h^{k,i}$ or $\boldsymbol{u} - \boldsymbol{u}_{h\tau}^{k,i}$, we in this contribution stick to problem-induced norms, written as

$$|||\boldsymbol{u} - \boldsymbol{u}_h^{k,i}||| \quad \text{or} \quad |||\boldsymbol{u} - \boldsymbol{u}_{h\tau}^{k,i}|||. \tag{2.6}$$

These *intrinsic error measures* arise from the problem considered, the mathematical function spaces adopted, and are defined so as to yield the most direct link to the sum of the *residual* and *nonconformity* in $\boldsymbol{u}_h^{k,i}$ or $\boldsymbol{u}_{h\tau}^{k,i}$, cf. Theorem 4.9 below. In the Laplace case (2.2), our choice of the intrinsic error measure in particular gives

$$|||\boldsymbol{u} - \boldsymbol{u}_h^{k,i}||| = ||\boldsymbol{u} - \boldsymbol{u}_h^{k,i}||, \tag{2.7}$$

where $\|\cdot\|$ is the $\boldsymbol{L}^2(\Omega)$ norm on Ω. This $L^2(\Omega)$ flux error $\|\boldsymbol{u} - \boldsymbol{u}_h^{k,i}\|$ is a physically crucially important quantity in the applications.

2.4.2 Spatial discretization error

The *spatial discretization error* is the error arising from the discretization on the spatial mesh $\mathcal{T}_h$ of Ω. For steady problems (2.1) and supposing that we have obtained the approximate flux $\boldsymbol{u}_h$ as specified "on paper", i.e., in particular solving "exactly" the associated (non)linear systems (so that there is no linearization index k and no algebraic solver index i), this is the only error component. In the Laplace case (2.2), the spatial discretization error writes

$$\|\boldsymbol{u} - \boldsymbol{u}_h\|. \tag{2.8}$$

In the setting of locally conservative methods that we consider, the approximate flux $\boldsymbol{u}_h$ will satisfy

$$\boldsymbol{u}_h \in \boldsymbol{H}(\mathrm{div}, \Omega) \quad \text{with} \quad \nabla \cdot \boldsymbol{u}_h = f \tag{2.9}$$

---

[1]In this text, we suppose exact computer implementation, i.e., we neglect rounding errors coming from approximate computer work with real numbers via floating point arithmetic. Such errors can also be taken into account, as well as additional errors such as those from (iterative) regularization or model selection.

for a piecewise constant source term f. Property (2.9) is precisely the "local conservation" from "locally conservative" methods. Additionally, we will be able to find an approximate potential $\tilde{p}_h$ such that

$$-\nabla \tilde{p}_h|_K = \boldsymbol{u}_h|_K$$

for all elements K of the mesh $\mathcal{T}_h$, so that the constitutive law is satisfied. In comparison with the continuous level (2.5), we see that the only deficiency of the discrete level is that the approximate potential $\tilde{p}_h$ is *nonconforming*, does not belong to the correct space,

$$\tilde{p}_h \notin H_0^1(\Omega), \tag{2.10}$$

in contrast to the situation at the continuous level, where $p \in H_0^1(\Omega)$ as per (2.5).

2.4.3 Iterative linear algebraic solver error

In order to obtain the approximate solution $\boldsymbol{u}_h$ as specified by some spatial discretization scheme, one typically needs to solve a (large, sparse) system of linear algebraic equations. In practice, we typically proceed iteratively, yielding an approximation $\boldsymbol{u}_h^i$ to $\boldsymbol{u}_h$ on each algebraic solver step $i \geq 0$. Then the *iterative linear algebraic solver error* is

$$\|\boldsymbol{u}_h - \boldsymbol{u}_h^i\|. \tag{2.11}$$

2.4.4 Iterative linearization error

In presence of nonlinearities, one typically employs an iterative approximation of a nonlinear operator $\mathcal{A}(\cdot)$ by linear operators $\mathcal{A}^k(\cdot)$ (on the continuous or on the discrete level). For steady problems, this yields an approximation $\boldsymbol{u}_h^k$ to $\boldsymbol{u}_h$ on each iterative linearization step $k \geq 0$. The corresponding *iterative linearization error* can then be structurally expressed as

$$\|\boldsymbol{u}_h - \boldsymbol{u}_h^k\|. \tag{2.12}$$

2.4.5 Temporal discretization error

The *temporal discretization error* is the error arising from the discretization on the temporal mesh of the time interval $(0, t_{\mathrm{F}})$. Naturally, it only appears for unsteady problems of the form (2.3), when some temporal discretization has been applied.

2.4.6 Error components

Wrapping up the previous sections, the total error is typically composed of several error components. For problems of the form (2.3)–(2.4) namely, we still have the potential and the Darcy velocity

$$p \in H_0^1(\Omega), \quad \boldsymbol{u} := -\underline{\boldsymbol{K}}(|\nabla p|)\nabla p \in \boldsymbol{H}(\mathrm{div}, \Omega), \quad \partial_t p + \nabla \cdot \boldsymbol{u} = f, \tag{2.13}$$

as in (2.5). Then, for a numerical approximation $\boldsymbol{u}_{h\tau}^{k,i}$, we will have

total error	$\|\|\|\boldsymbol{u} - \boldsymbol{u}_{h\tau}^{k,i}\|\|\|,$	(2.14a)
temporal discretization error	$\|\|\|\boldsymbol{u} - \boldsymbol{u}_\tau\|\|\|,$	(2.14b)
spatial discretization error	$\|\|\|\boldsymbol{u}_\tau - \boldsymbol{u}_{h\tau}\|\|\|,$	(2.14c)
iterative linearization error	$\|\|\|\boldsymbol{u}_{h\tau} - \boldsymbol{u}_{h\tau}^k\|\|\|,$	(2.14d)
iterative linear algebraic solver error	$\|\|\|\boldsymbol{u}_{h\tau}^k - \boldsymbol{u}_{h\tau}^{k,i}\|\|\|,$	(2.14e)

in extension of (2.8), (2.12), and (2.11). Details are developed in Sections 6 and 7.

2.5 A posteriori error estimate

An a posteriori error estimate is real number $\eta(\boldsymbol{u}_h^{k,i})$ or $\eta(\boldsymbol{u}_{h\tau}^{k,i})$ that bounds the error between the unknown exact flux $\boldsymbol{u}$ of the partial differential equation like (2.1) or (2.3) and the available numerical approximation $\boldsymbol{u}_h^{k,i}$ or $\boldsymbol{u}_{h\tau}^{k,i}$ of $\boldsymbol{u}$. Importantly, $\eta(\boldsymbol{u}_h^{k,i})$ or $\eta(\boldsymbol{u}_{h\tau}^{k,i})$ has to be *fully computable* from the approximate solution $\boldsymbol{u}_h^{k,i}$ or $\boldsymbol{u}_{h\tau}^{k,i}$. The following are some highly desirable properties.

2.5.1 Guaranteed error upper bound (reliability)

We say that an a posteriori error estimate gives a *guaranteed error upper bound* when the intrinsic error measure (2.6) is bounded by it as

$$\||\boldsymbol{u} - \boldsymbol{u}_h^{k,i}\|| \leq \eta(\boldsymbol{u}_h^{k,i}) \quad \text{or} \quad \||\boldsymbol{u} - \boldsymbol{u}_{h\tau}^{k,i}\|| \leq \eta(\boldsymbol{u}_{h\tau}^{k,i}). \tag{2.15}$$

This allows to *certify* the *error*. This may appear as a miracle, since the left-hand side of (2.15) is unknown, whereas the right-hand side of (2.15) is known, but is indeed possible, as we will see in the developments below. The key will be to follow at the discrete level the physical and mathematical properties of the weak solutions of PDEs as expressed by (2.5) and (2.13). In the setting of Section 2.4.2, namely, from (2.5) and since $\boldsymbol{u}_h \in \boldsymbol{H}(\mathrm{div}, \Omega)$ with $\nabla \cdot \boldsymbol{u}_h = f$, we easily show that

$$\|\boldsymbol{u} - \boldsymbol{u}_h\| \leq \eta(\boldsymbol{u}_h) := \|\boldsymbol{u}_h + \nabla \zeta_h\| \tag{2.16}$$

for an arbitrary

$$\zeta_h \in H_0^1(\Omega); \tag{2.17}$$

we will see the details in Theorem 4.18 below. Here, ζ_h is discrete, piecewise polynomial, and obtained by some *local* modification from $\tilde{p}_h$ (postprocessing and averaging); we call it a *potential reconstruction*. Similarly, when $\boldsymbol{u}_h^{k,i} \notin \boldsymbol{H}(\mathrm{div}, \Omega)$ or $\nabla \cdot \boldsymbol{u}_h^{k,i} \neq f$ (namely in presence of linear and nonlinear solvers), an *equilibrated flux reconstruction* $\boldsymbol{\sigma}_h^{k,i}$ such that

$$\boldsymbol{\sigma}_h^{k,i} \in \boldsymbol{H}(\mathrm{div}, \Omega) \quad \text{and} \quad \nabla \cdot \boldsymbol{\sigma}_h^{k,i} = f \tag{2.18}$$

(we will also consider merely $\nabla \cdot \boldsymbol{\sigma}_h^{k,i} = f + \rho_h^{k,i}$ where $\rho_h^{k,i}$ is a remainder made as small as necessary) will be the key together with the potential reconstruction (2.17).

2.5.2 Error lower bound (efficiency)

Relation (2.15) is only one-sided, so that it might potentially happen that the left-hand side of (2.15) is zero (small), whereas the right-hand side is nonzero (big). A mathematical equivalence between the error and the a posteriori error estimator is ensured when, in addition to (2.15), there also holds

$$\eta(\boldsymbol{u}_h^{k,i}) \leq C\||\boldsymbol{u} - \boldsymbol{u}_h^{k,i}\|| \quad \text{or} \quad \eta(\boldsymbol{u}_{h\tau}^{k,i}) \leq C\||\boldsymbol{u} - \boldsymbol{u}_{h\tau}^{k,i}\||, \tag{2.19}$$

which is called (global) *efficiency*. Here, C is a generic constant. Actually, (2.19) may only hold when on the right-hand side, there are some additional terms without structural importance, i.e., of vanishing importance with decreasing the mesh size h and the time step τ (so-called "data oscillation" terms).

Strengthening (2.19) to hold locally in space (and in time) is actually often possible, and highly desirable. This then means that when we *predict* the *presence of error* in some *mesh element* or *time step*, then it is *indeed localized there*, or in a small neighborhood. In the example (2.16), in particular, we easily have

$$\underbrace{\|\boldsymbol{u}_h + \nabla \zeta_h\|^2}_{[\eta(\boldsymbol{u}_h)]^2} = \sum_{K \in \mathcal{T}_h} \underbrace{\|\boldsymbol{u}_h + \nabla \zeta_h\|_K^2}_{[\eta_K(\boldsymbol{u}_h)]^2}, \tag{2.20}$$

and it is possible to strengthen (2.19) to

$$\eta_K(\boldsymbol{u}_h) \leq C\|\boldsymbol{u} - \boldsymbol{u}_h\|_{\omega_K}, \tag{2.21}$$

where ω_K are the simplices of the mesh $\mathcal{T}_h$ sharing a vertex with the simplex K. This is called *local efficiency*.

2.5.3 Robustness with respect to the spatial and temporal domains as well as physical and numerical parameters

There may be various ways how to simultaneously achieve (2.15) and (2.19) (and also (2.21)). Then, a crucial indicator of the quality of the derived estimate $\eta(\boldsymbol{u}_h^{k,i})$ or $\eta(\boldsymbol{u}_{h\tau}^{k,i})$ is the nature and behavior of the "generic" constant C from (2.19) (and (2.21)) (recall that there is no unknown constant in (2.15)). Ideally, C is truly generic, only depending on the space dimension d and the shape regularity (smallest angle) of the spatial meshes $\mathcal{T}_h$ (say taking a value $2 - 10$ in practical applications when $1 \leq d \leq 3$). In any case, it is requested that C is independent of the spatial domain Ω (namely its size) and the temporal domain $(0, t_\mathrm{F})$ (the final simulation time t_F) and of any numerical parameters, namely the mesh size h and the time step τ. It is also desirable to have C independent of the physical parameters such us medium properties (the smallest and highest (eigen)values of diffusion tensor $\underline{\boldsymbol{K}}$ or specifications of the nonlinearity in $\underline{\boldsymbol{K}}(\cdot)$). We then speak of *robustness*.

2.5.4 Asymptotic exactness

One characterization of the quality of an a posteriori error estimate is the so-called effectivity index given by

$$I_{\mathrm{eff}} := \frac{\eta(\boldsymbol{u}_h^{k,i})}{\||\boldsymbol{u} - \boldsymbol{u}_h^{k,i}\||} \quad \text{or} \quad I_{\mathrm{eff}} := \frac{\eta(\boldsymbol{u}_{h\tau}^{k,i})}{\||\boldsymbol{u} - \boldsymbol{u}_{h\tau}^{k,i}\||}. \tag{2.22}$$

Recall that in our setting, from (2.15), I_{eff} is greater than or equal to 1. If I_{eff} goes to the optimal value of 1 with decreasing mesh size h and time step τ, we speak of *asymptotic exactness*. We will see such cases below.

2.5.5 Inexpensive evaluation

There exist very sharp estimates $\eta(\boldsymbol{u}_h^{k,i})$ or $\eta(\boldsymbol{u}_{h\tau}^{k,i})$ in (2.15), which, however, are expensive to compute and evaluate, comparatively as expensive as to compute the numerical approximations $\boldsymbol{u}_h^{k,i}$ or $\boldsymbol{u}_{h\tau}^{k,i}$ themselves. Indeed, from (2.16) and (2.5), as sharp as requested upper bound can be obtained upon a solution of some global problem with increasing size, see Theorem 4.9 below for more details. Such classes of estimates are, however, excluded from our considerations, as not suitable in practice (they may serve to certify the error but it is hard to think of their use in adaptivity because of their cost). Some other classes of estimates request a solution of some small, local, mutually independent problems in patches of mesh elements. These are termed implicit and can be labelled "inexpensive". In this contribution, however, we do not consider such estimates either, and we only focus on estimates whose evaluation from the approximate solution $\boldsymbol{u}_h^{k,i}$ or $\boldsymbol{u}_{h\tau}^{k,i}$ is yet cheaper, *explicit*. For example, the potential reconstruction ζ_h from (2.17) is obtained from $\boldsymbol{u}_h$ (passing through $\tilde{p}_h$) by some local postprocessing and averaging. On general polytopal meshes $\mathcal{T}_H$, we will then develop estimates whose evaluation merely consists in *multiplications* of *local vectors* by *local matrices*, where the local matrices are pre-processed from the scheme or cell geometry at hand and the local vectors are the degrees of freedom representing the available values of the local fluxes and potentials. These estimate are then as *inexpensive* as possible.

2.5.6 Estimating the error components

We have seen in Section 2.4.6 that the total error in a numerical approximation of, say (2.3)–(2.4), has several components, namely due to spatial discretization, temporal discretization, iterative linearization, and iterative linear algebraic resolution. Congruently, an a posteriori error estimator should *identify* and estimate these *error components*, developing (2.15) structurally into

$$\||\boldsymbol{u} - \boldsymbol{u}_h^{k,i}\|| \le \eta(\boldsymbol{u}_h^{k,i}) \le \eta_{\mathrm{sp}}(\boldsymbol{u}_h^{k,i}) + \eta_{\mathrm{lin}}(\boldsymbol{u}_h^{k,i}) + \eta_{\mathrm{alg}}(\boldsymbol{u}_h^{k,i}) \tag{2.23a}$$
$$\text{or}$$
$$\||\boldsymbol{u} - \boldsymbol{u}_{h\tau}^{k,i}\|| \le \eta(\boldsymbol{u}_{h\tau}^{k,i}) \le \eta_{\mathrm{sp}}(\boldsymbol{u}_{h\tau}^{k,i}) + \eta_{\mathrm{tm}}(\boldsymbol{u}_{h\tau}^{k,i}) + \eta_{\mathrm{lin}}(\boldsymbol{u}_{h\tau}^{k,i}) + \eta_{\mathrm{alg}}(\boldsymbol{u}_{h\tau}^{k,i}), \tag{2.23b}$$

in congruence with the example of error components in (2.14).

2.6 Adaptivity: balancing the error component estimates

Having approximately identified the total error and its components as per (2.23), one can think of how to use this information. We are in particular going to develop *adaptivity* consisting in *balancing* of the estimated *error components*. For (2.23b), for example, this consists in making all the estimators $\eta_{\mathrm{sp}}(\boldsymbol{u}_{h\tau}^{k,i})$, $\eta_{\mathrm{tm}}(\boldsymbol{u}_{h\tau}^{k,i})$, $\eta_{\mathrm{lin}}(\boldsymbol{u}_{h\tau}^{k,i})$, and $\eta_{\mathrm{alg}}(\boldsymbol{u}_{h\tau}^{k,i})$ of comparable size, uniformly throughout the entire simulation. Indeed, shall for example the iterative linear algebraic solver error estimate $\eta_{\mathrm{alg}}(\boldsymbol{u}_{h\tau}^{k,i})$ be orders of magnitude smaller than the spatial or temporal discretization error estimates, we may be wasting our computational resources (unfortunately, this is common in practice). On a more common note, say for a steady problem, we will develop

$$\left[\eta_{\mathrm{sp}}(\boldsymbol{u}_h^{k,i})\right]^2 = \sum_{K \in \mathcal{T}_h} \left[\eta_{K,\mathrm{sp}}(\boldsymbol{u}_h^{k,i})\right]^2,$$

in extension of (2.20). We then design algorithms ensuring that the elementwise a posteriori error estimators $\eta_{K,\mathrm{sp}}(\boldsymbol{u}_{h\tau}^{k,i})$ have all comparable values. This is termed *mesh adaptivity* (adaptive mesh refinement).

2.7 Mass balance recovery at any step of numerical resolution

A locally conservative method produces an approximation to the Darcy velocity $\boldsymbol{u}$ from (2.5) in a way that

$$\boldsymbol{u}_h \in \boldsymbol{H}(\mathrm{div}, \Omega) \quad \text{with} \quad \nabla \cdot \boldsymbol{u}_h = f$$

for a piecewise constant source term f, see (2.9). Unfortunately, during an iterative linearization or during iterative linear algebraic resolution, this may be lost in that

$$\boldsymbol{u}_h^{k,i} \notin \boldsymbol{H}(\mathrm{div}, \Omega) \quad \text{and/or} \quad \nabla \cdot \boldsymbol{u}_h^{k,i} \neq f.$$

One of the specificities of our approach is that in the construction of our a posteriori error estimates giving a guaranteed error upper bound and distinguishing the total error components, we actually recover a locally conservative flux field (Darcy velocity) $\boldsymbol{\sigma}_h^{k,i}$ such that

$$\boldsymbol{\sigma}_h^{k,i} \in \boldsymbol{H}(\mathrm{div}, \Omega) \quad \text{and} \quad \nabla \cdot \boldsymbol{\sigma}_h^{k,i} = f,$$

see (2.18). In practice, we may go for a cheaper variant where $\nabla \cdot \boldsymbol{\sigma}_h^{k,i} = f + \rho_h^{k,i}$ with $\rho_h^{k,i}$ is a remainder made as small as necessary.

2.8 Bibliographic resources

Focusing on the spatial discretization error, it has been rather soon understood how to structurally obtain a guaranteed error upper bound as in (2.16): this follows from the Prager–Synge equality [216], relying on the physical and mathematical properties of the weak solutions of PDEs as discussed in Section 2.2. In particular, in finite element methods, where one obtains an approximate solution $p_h \in H_0^1(\Omega)$, one was in quest of an equilibrated flux reconstruction $\boldsymbol{\sigma}_h \in \boldsymbol{H}(\mathrm{div}, \Omega)$ with $\nabla \cdot \boldsymbol{\sigma}_h = f$ as in (2.18). A *local* construction, is however, a bit involved in finite elements, so that its establishment was rather labyrinthine, see Ladevèze [174], Ladevèze and Leguillon [175], Bossavit [56], Destuynder and Métivet [97], Larson and Niklasson [177], Luce and Wohlmuth [183], Vohralík [248], Braess and Schöberl [57], Cottereau *et al.* [96], Ern and Vohralík [124, 125], and the references therein.

In contrast, in methods locally conservative by construction, we already have $\boldsymbol{u}_h \in \boldsymbol{H}(\mathrm{div}, \Omega)$ with $\nabla \cdot \boldsymbol{u}_h = f$: directly from mixed finite elements and by local postprocessing for finite volume-type methods, as detailed below. And since a potential reconstruction $\zeta_h \in H_0^1(\Omega)$ as in (2.17) is rather easy to design, a posteriori error estimates of the form (2.16) were put in place in Achdou *et al.* [3], Ainsworth [12, 13], Kim [162], and Vohralík [247, 249], see also the references therein. In discontinuous Galerkin and related methods, $\boldsymbol{u}_h \in \boldsymbol{H}(\mathrm{div}, \Omega)$ with $\nabla \cdot \boldsymbol{u}_h = f$ is not directly available but can be readily reconstructed, see Bastian and Rivière [34], Kim [162], and Ern *et al.* [119]. More references specifically for lowest-order locally conservative methods/finite volumes are discussed in Sections 4.12 and 5.11 below.

Unsteady problems were analyzed in Picasso [214], Verfürth [242], Bergam *et al.* [47], Ern and Vohralík [123], and Ern *et al.* [120], see also the references therein. Iterative linear algebraic solver error and iterative linearization error were in particular addressed in Becker *et al.* [39], Jiránek *et al.* [154], Arioli *et al.* [25], Ern and Vohralík [124], Heid and Wihler [147], Gantner *et al.* [136], Haberl *et al.* [142], Mitra and Vohralík [190], and the references therein. In-depth presentation of a posteriori error estimates and adaptivity can then be found in the monographs by Synge [236], Hlaváček *et al.* [152], Zeidler [263], Ainsworth and Oden [14], Babuška and Strouboulis [29], Bangerth and Rannacher [31], Han [144], Ladevèze and Pelle [176], Repin [219], Nochetto *et al.* [198], Deuflhard and Weiser [99], and Verfürth [245], see also Chamoin and Legoll [76], Bartels and Kaltenbach [33], Mghazli [186, 187], and Smears [232].

3 Sobolev spaces and their piecewise polynomial subspaces, meshes, and finite volume methods

In this section, we describe the infinite-dimensional Sobolev spaces $H_0^1(\Omega)$ and $\boldsymbol{H}(\mathrm{div}, \Omega)$ and their finite-dimensional subspaces formed by piecewise polynomials. We also set up the notation, define a simplicial mesh, and describe the considered polytopal meshes. We finally introduce the basic principle of the finite volume and related lowest-order locally conservative methods.

3.1 Basic notation

We introduce here the basic notation used throughout the manuscript.

3.1.1 Domain Ω

We let $\Omega \subset \mathbb{R}^d$ be an open bounded connected set in a form of an interval or polygon or polyhedron for respectively $d = 1, 2, 3$ space dimensions, or a polytope in general for $d \geq 1$, with a Lipschitz-continuous boundary $\partial\Omega$. The notation ω is reserved for a subdomain $\omega \subset \Omega$.

3.1.2 Lebesgue spaces

For $\omega \subset \Omega$, we let $L^2(\omega)$ stand for square-integrable scalar-valued functions $v, w : \omega \to \mathbb{R}$ with the scalar product

$$(v, w)_\omega := \int_\omega v(\boldsymbol{x}) w(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}. \tag{3.1a}$$

Square-integrable vector-valued functions $\boldsymbol{v}, \boldsymbol{w} : \omega \to \mathbb{R}^d$ are then collected in $\boldsymbol{L}^2(\omega) := [L^2(\omega)]^d$, with the scalar product

$$(\boldsymbol{v}, \boldsymbol{w})_\omega := \int_\omega \boldsymbol{v}(\boldsymbol{x}) \cdot \boldsymbol{w}(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}. \tag{3.1b}$$

The associated norms are denoted by $\|\cdot\|_\omega$. When $\omega = \Omega$, we drop the subscripts.

3.1.3 Measures, sizes, and cardinalities

We write $|\omega|$ for the Lebesgue measure of a domain ω, $|\boldsymbol{v}|$ for the Euclidean size of a vector $\boldsymbol{v} \in \mathbb{R}^d$, and $|\mathcal{S}|$ for the cardinality of a set $\mathcal{S}$. We denote by $\boldsymbol{n}_\omega$ the exterior unit normal of a domain ω.

3.2 Infinite-dimensional Sobolev spaces $H_0^1(\Omega)$ and $\boldsymbol{H}(\mathrm{div}, \Omega)$

In the mathematical sense, the solutions of partial differential equations live in specific infinite-dimensional spaces that admit generalized (weak) partial derivatives and divergences. This is in particular the Sobolev space $H_0^1(\Omega)$ for scalar-valued functions and the Sobolev space $\boldsymbol{H}(\mathrm{div}, \Omega)$ for vector-valued functions. We recapitulate some basics about these spaces here. We follow Allaire [15], Ern and Guermond [118], and Vohralík [251, 252].

3.2.1 Sobolev space $H_0^1(\Omega)$

We start by the notion of a weak partial derivative. Let $\mathcal{D}(\Omega)$ be the space of functions from $C^\infty(\Omega)$ with a compact support in Ω.

Definition 3.1 (Weak partial derivative). *Let a scalar-valued function $v : \Omega \to \mathbb{R}$ be given. We say that v admits a weak i-th partial derivative, $1 \leq i \leq d$, if*

1. *$v \in L^2(\Omega)$;*
2. *there exists a function $w_i : \Omega \to \mathbb{R}$ such that*
   (a) *$w_i \in L^2(\Omega)$;*
   (b) *$(v, \partial_{\boldsymbol{x}_i}\varphi) = -(w_i, \varphi) \qquad \forall \varphi \in \mathcal{D}(\Omega)$.*

The function w_i is called the weak i-th partial derivative of v. We use the notation $\partial_{\boldsymbol{x}_i} v = w_i$.

We then define a weak gradient:

Definition 3.2 (Weak gradient). *Let a scalar-valued function $v : \Omega \to \mathbb{R}$ be given. We say that v admits a weak gradient if v admits the weak i-th partial derivative for all $1 \leq i \leq d$. We set*

$$\nabla v := (\partial_{\boldsymbol{x}_1} v, \ldots, \partial_{\boldsymbol{x}_d} v)^t. \tag{3.2}$$

Definition 3.3 (The space $H^1(\Omega)$). *The space $H^1(\Omega)$ is the space of all the functions that admit the weak gradient.*

Let us recall from [15, 118] that $H^1(\Omega)$ is a Hilbert space for the scalar product given by $(u, v) + (\nabla u, \nabla v)$.
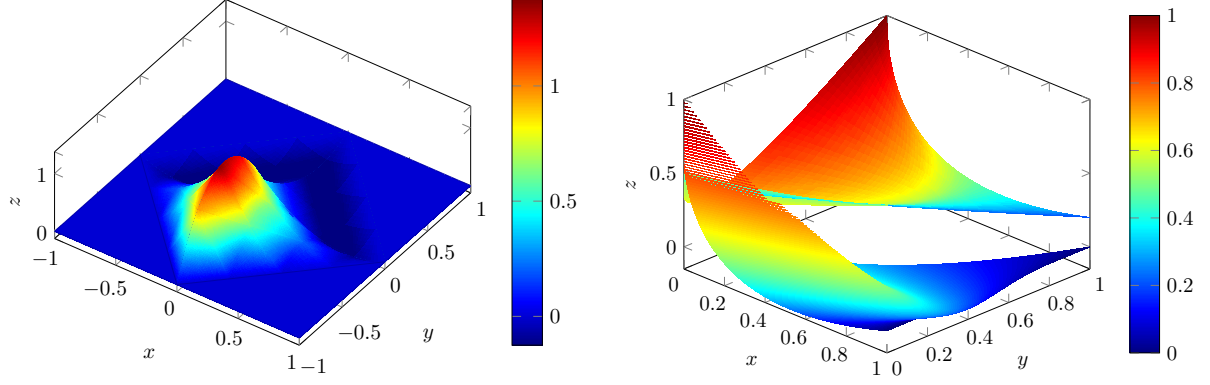
Figure 1: Example of a function belonging to the Sobolev space $H_0^1(\Omega)$ (left; the function actually lies in $C^0(\overline{\Omega})$, but not in $C^1(\overline{\Omega})$). Example of a function not belonging to the Sobolev space $H_0^1(\Omega)$ but belonging to the broken Sobolev space $H^1(\mathcal{T}_h)$ for a mesh $\mathcal{T}_h$ composed of two triangles (right; note that there is no trace-continuity on the common face of the two mesh elements)

Definition 3.4 (The space $H_0^1(\Omega)$). *The space $H_0^1(\Omega)$ is the space of functions $v \in H^1(\Omega)$ such that $v|_{\partial\Omega} = 0$.*

Recalling again from [15, 118], $H_0^1(\Omega)$ is a Hilbert space for the scalar product given by $(\nabla u, \nabla v)$.

Remark 3.5 (Trace and trace continuity). *In the above definition, we have used the notation $v|_{\partial\Omega}$. As the functions from $H^1(\Omega)$ are taken from the Lebesgue space $L^2(\Omega)$, $v|_{\partial\Omega}$ in the sense of a restriction is not defined since $\partial\Omega$ is a set of measure zero. In our notation, by $v|_{\partial\Omega}$, we mean the* trace *of $v \in H^1(\Omega)$ on the boundary of Ω, a key notion for the Sobolev space $H^1(\Omega)$. Crucially, $v|_{\partial\Omega}$ in the sense of traces and $v|_{\partial\Omega}$ as a restriction coincide for any function $v \in H^1(\Omega)$ that is continuous, $v \in H^1(\Omega) \cap C^0(\overline{\Omega})$. Let us also recall that $H^1(\Omega) \subset C^0(\overline{\Omega})$ in one space dimension but $H^1(\Omega) \not\subset C^0(\overline{\Omega})$ in multiple space dimensions. Crucially, functions v in $H^1(\Omega)$, though not necessarily continuous, are continuous in the trace sense. An illustration of a function v belonging to $H^1(\Omega)$ is given in Figure 1, left, and of a function v not belonging to $H^1(\Omega)$ in Figure 1, right. We refer for details to [15, 118, 251].*

3.2.2 Sobolev space $\boldsymbol{H}(\mathrm{div}, \Omega)$

We now turn to weak a divergence.

Definition 3.6 (Weak divergence). *Let a vector-valued function $\boldsymbol{v} : \Omega \to \mathbb{R}^d$ be given. We say that $\boldsymbol{v}$ admits a weak divergence if*

1. *$\boldsymbol{v} \in \boldsymbol{L}^2(\Omega)$;*
2. *there exists a function $w : \Omega \to \mathbb{R}$ such that*
   (a) *$w \in L^2(\Omega)$;*
   (b) *$(\boldsymbol{v}, \nabla\varphi) = -(w, \varphi) \qquad \forall \varphi \in \mathcal{D}(\Omega)$.*

The function w is called the weak divergence *of $\boldsymbol{v}$. We use the notation $\nabla\cdot\boldsymbol{v} = w$.*

Definition 3.7 (The space $\boldsymbol{H}(\mathrm{div}, \Omega)$). *The space $\boldsymbol{H}(\mathrm{div}, \Omega)$ is the space of all the functions that admit the weak divergence.*

Let us recall from [15, 118] that $\boldsymbol{H}(\mathrm{div}, \Omega)$ is a Hilbert space for the scalar product given by $(\boldsymbol{u}, \boldsymbol{v}) + (\nabla\cdot\boldsymbol{u}, \nabla\cdot\boldsymbol{v})$. Similarly to Remark 3.5, there holds:

Remark 3.8 (Normal trace and normal-trace continuity). *On $\partial\Omega$, functions $\boldsymbol{v} \in \boldsymbol{H}(\mathrm{div}, \Omega)$ admit a normal component in appropriate sense, called normal trace and denoted as $(\boldsymbol{v}\cdot\boldsymbol{n})|_{\partial\Omega}$. Namely, if $\boldsymbol{v} \in \boldsymbol{H}(\mathrm{div}, \Omega) \cap [C^0(\overline{\Omega})^d]$, then $\boldsymbol{v}\cdot\boldsymbol{n}$ is the usual normal component of $\boldsymbol{v}$ on the boundary $\partial\Omega$. Functions $\boldsymbol{v}$ in $\boldsymbol{H}(\mathrm{div}, \Omega)$ are not necessarily continuous in each component, not necessarily normal-component continuous, but are normal-trace continuous. An illustration of a function $\boldsymbol{v}$ belonging to $\boldsymbol{H}(\mathrm{div}, \Omega)$ is given in Figure 2, left, and of a function $\boldsymbol{v}$ not belonging to $\boldsymbol{H}(\mathrm{div}, \Omega)$ in Figure 2, right. We refer for details to [15, 118].*
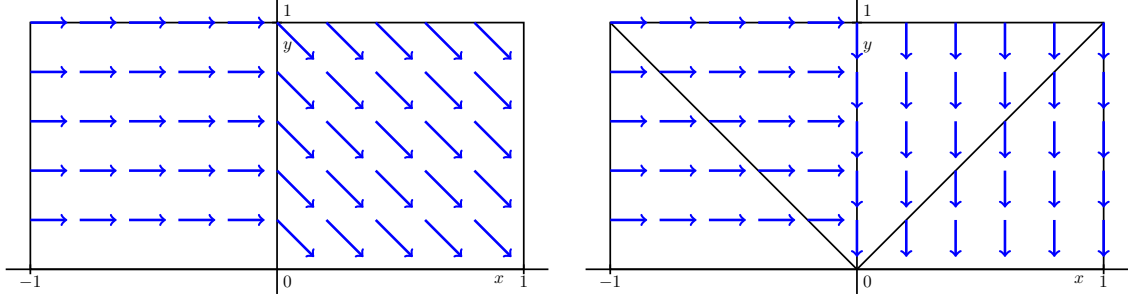
Figure 2: Example of a function belonging to the Sobolev space $\boldsymbol{H}(\mathrm{div}, \Omega)$ (left; on the interface $x = 0$ between the two subdomains $\Omega_1 = (-1, 0) \times (0, 1)$ and $\Omega_2 = (0, 1) \times (0, 1)$, $\boldsymbol{v}|_{\Omega_1} \cdot \boldsymbol{n} = \boldsymbol{v}|_{\Omega_2} \cdot \boldsymbol{n}$, with $\boldsymbol{n} = (1, 0)^{\mathrm{t}}$; $\boldsymbol{v}$ is not continuous for each component (the y component of $\boldsymbol{v}$ is discontinuous, as it passes from value 0 in Ω_1 to a nonzero value in Ω_2) but $\boldsymbol{v}$ is normal-trace continuous). Example of a function not belonging to the Sobolev space $\boldsymbol{H}(\mathrm{div}, \Omega)$ but belonging to the broken Sobolev space $\boldsymbol{H}(\mathrm{div}, \mathcal{T}_h)$ (right, there is no normal trace-continuity across the mesh face at $x = 0$).

The Sobolev spaces $H_0^1(\Omega)$ from Definition 3.4 and $\boldsymbol{H}(\mathrm{div}, \Omega)$ from Definition 3.7 form exactly the right setting for the following Green theorem:

Theorem 3.9 (Green theorem). *Let $v \in H_0^1(\Omega)$ and let $\boldsymbol{w} \in \boldsymbol{H}(\mathrm{div}, \Omega)$. Then*

$$(\boldsymbol{w}, \nabla v) + (\nabla \cdot \boldsymbol{w}, v) = 0. \tag{3.3}$$

In the following, we will also employ the above notation and developments for subdomains $\omega \subset \Omega$.

3.3 Simplicial meshes

In order to form finite-dimensional subspaces of $H_0^1(\Omega)$ and $\boldsymbol{H}(\mathrm{div}, \Omega)$, we will need a simplicial mesh $\mathcal{T}_h$ of Ω. This a finite collection of closed d-simplices (intervals for $d = 1$, triangles for $d = 2$, tetrahedra for $d = 3$) such that $\cup_{K \in \mathcal{T}_h} K = \overline{\Omega}$ and such that the intersection of two different simplices is either empty or their entire l-dimensional face, $0 \leq l \leq (d - 1)$. Here, for example for space dimension $d = 3$, a vertex is a 0-dimensional face, an edge is a 1-dimensional face, and a face is a 2-dimensional face. Henceforth, by face, we mean $(d - 1)$-dimensional face. Figure 3 gives illustrations for $d = 2$ and $d = 3$. We denote by

$$\theta_{\mathcal{T}_h} := \max_{K \in \mathcal{T}_h} \frac{h_K}{\iota_K} \tag{3.4}$$

the shape-regularity parameter of the mesh $\mathcal{T}_h$, where h_K is the diameter of the simplex $K \in \mathcal{T}_h$ and ι_K is the diameter of largest ball inscribed in K. This parameter expresses how much the mesh is "distorted": it is close to 1 for meshes of equilateral simplices and increases in presence of small angles. For each mesh $\mathcal{T}_h$, $\theta_{\mathcal{T}_h}$ is a bounded number. Our theoretical constants such as C from (2.19) and (2.21) will depend on $\theta_{\mathcal{T}_h}$. If we consider a sequence of meshes, then we need $\theta_{\mathcal{T}_h}$ uniformly bounded.

The set of the $((d-1)$-dimensional) faces of $\mathcal{T}_h$ is denoted as $\mathcal{F}_h$. It is divided into interior faces $\mathcal{F}_h^{\mathrm{int}}$ and boundary faces $\mathcal{F}_h^{\mathrm{ext}}$: $\sigma \in \mathcal{F}_h^{\mathrm{int}}$ if there exist $K, L \in \mathcal{T}_h$, $K \neq L$, such that $\sigma = K \cap L$, and $\sigma \in \mathcal{F}_h^{\mathrm{ext}}$ if there exist $K \in \mathcal{T}_h$ such that $\sigma \subset K \cap \partial\Omega$. We will also need the notation $\mathcal{V}_h$ for the vertices of the mesh $\mathcal{T}_h$ and $\mathcal{V}_\sigma$ for the vertices of the face $\sigma \in \mathcal{F}_h$. For an interior face $\sigma \in \mathcal{F}_h^{\mathrm{int}}$, we fix an arbitrary orientation and denote the corresponding unit normal vector by $\boldsymbol{n}_\sigma$. For a boundary face $\sigma \in \mathcal{F}_h^{\mathrm{ext}}$, we make $\boldsymbol{n}_\sigma$ coincide with the exterior unit normal $\boldsymbol{n}_\Omega$ of Ω. Finally, for all $K \in \mathcal{T}_h$, we denote by $\boldsymbol{n}_K$ the unit normal vector to σ pointing out of K and by $\mathcal{F}_K$ the set of all faces σ of K.

3.4 Broken Sobolev spaces $H^1(\mathcal{T}_h)$ and $\boldsymbol{H}(\mathrm{div}, \mathcal{T}_h)$

Before we pass to finite-dimensional subspaces of $H_0^1(\Omega)$ and $\boldsymbol{H}(\mathrm{div}, \Omega)$, it is useful to define the broken Sobolev space $H^1(\mathcal{T}_h)$ as

$$H^1(\mathcal{T}_h) := \{v \in L^2(\Omega); \, v|_K \in H^1(K) \quad \forall K \in \mathcal{T}_h\}, \tag{3.5}$$

and the broken Sobolev space $\boldsymbol{H}(\mathrm{div}, \mathcal{T}_h)$ as

$$\boldsymbol{H}(\mathrm{div}, \mathcal{T}_h) := \{\boldsymbol{v} \in \boldsymbol{L}^2(\Omega); \, \boldsymbol{v}|_K \in \boldsymbol{H}(\mathrm{div}, K) \quad \forall K \in \mathcal{T}_h\}. \tag{3.6}$$

13
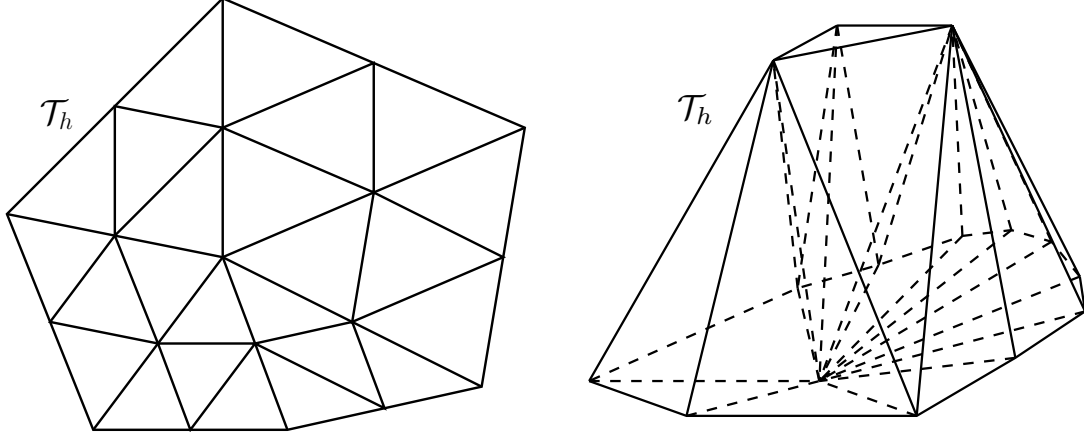
Figure 3: Simplicial mesh $\mathcal{T}_h$ for $d = 2$ (left) and $d = 3$ (right)

An illustration of a function $v \in H^1(\mathcal{T}_h)$ is given in Figure 1, right, and an illustration of a function $\boldsymbol{v} \in \boldsymbol{H}(\mathrm{div}, \mathcal{T}_h)$ is given in Figure 2, right. In this setting, Remarks 3.5 and 3.8 can respectively be developed into, see [100, Lemma 1.23] or [118, Theorem 18.8]:

Lemma 3.10 (Trace continuity over mesh faces). *There holds*

$$v \in H^1(\mathcal{T}_h) \text{ and } (v|_K)|_\sigma = (v|_L)|_\sigma \quad \forall \sigma = K \cap L \in \mathcal{F}_h^{\mathrm{int}} \iff v \in H^1(\Omega), \tag{3.7a}$$

$$v \in H^1(\mathcal{T}_h) \text{ and } (v|_K)|_\sigma = (v|_L)|_\sigma \quad \forall \sigma = K \cap L \in \mathcal{F}_h^{\mathrm{int}}$$
$$\text{and } (v|_K)|_\sigma = 0 \quad \forall \sigma = K \cap \partial\Omega \in \mathcal{F}_h^{\mathrm{ext}} \iff v \in H_0^1(\Omega). \tag{3.7b}$$

In words, functions from the broken Sobolev space $H^1(\mathcal{T}_h)$ belong to the Sobolev space $H^1(\Omega)$ if and only if their traces are uniquely defined on all mesh interior faces, with the same trace from both elements sharing the given interior face. We say that the functions from $H^1(\Omega)$ are trace continuous. Functions from $H_0^1(\Omega)$ then additionally have the traces zero on boundary faces.

Similarly, following [100, Lemma 1.24] or [118, Theorem 18.10], we have:

Lemma 3.11 (Normal trace continuity over mesh faces). *Functions from $\boldsymbol{H}(\mathrm{div}, \mathcal{T}_h)$ belong to $\boldsymbol{H}(\mathrm{div}, \Omega)$ if and only if their normal traces are continuous over mesh interior faces in appropriate normal-trace sense.*

3.5 Finite-dimensional subspaces of $H_0^1(\Omega)$ and $\boldsymbol{H}(\mathrm{div}, \Omega)$

We now introduce standard finite-dimensional subspaces of $H_0^1(\Omega)$ and $\boldsymbol{H}(\mathrm{div}, \Omega)$, again following Allaire [15] and Ern and Guermond [118].

3.5.1 Piecewise polynomial subspaces of $L^2(\Omega)$

Let $K \in \mathcal{T}_h$ be a simplex from the mesh $\mathcal{T}_h$ and let $k \geq 0$ be an integer. We denote by $\mathcal{P}_k(K)$ the space of scalar-valued polynomials on K of total degree at most k. In particular, $\mathcal{P}_0(K)$ stands for constants on K, $\mathcal{P}_1(K)$ for affine functions on K, and $\mathcal{P}_2(K)$ for quadratic functions on K. The spaces for $k = 1, 2$ and $d = 2, 3$ are schematically visualized in Figure 4. In particular, fixing the point values as per Figure 4 uniquely determines a function in $\mathcal{P}_k(K)$. We will often use the space of piecewise k-degree polynomials on the simplicial mesh $\mathcal{T}_h$,

$$\mathcal{P}_k(\mathcal{T}_h) := \{v_h \in L^2(\Omega); v_h|_K \in \mathcal{P}_k(K) \quad \forall K \in \mathcal{T}_h\}. \tag{3.8}$$

3.5.2 Lagrange piecewise polynomial subspaces of $H_0^1(\Omega)$

The space $\mathcal{P}_k(\mathcal{T}_h)$ from (3.8) is a finite-dimensional subspace of the space $H^1(\mathcal{T}_h)$ from (3.5) but not of the Sobolev space $H_0^1(\Omega)$. To create a subspace of $H_0^1(\Omega)$ of the so-called *Lagrange* piecewise polynomials,
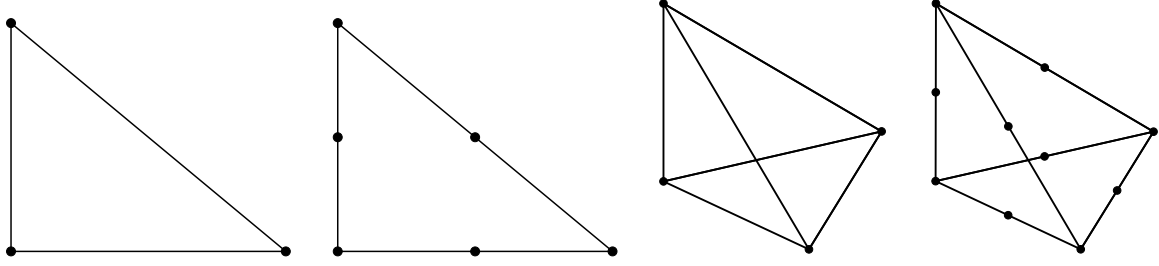
Figure 4: Spaces $\mathcal{P}_k(K)$ and point values uniquely fixing a function $v_h \in \mathcal{P}_k(K)$ (the so-called Lagrange nodes) for a mesh element $K \in \mathcal{T}_h$. Polynomial degrees $k = 1$ and $k = 2$, space dimensions $d = 2$ (left) and $d = 3$ (right)
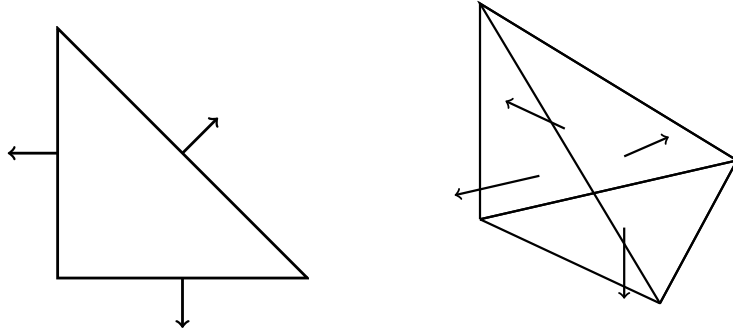
Figure 5: Spaces $\boldsymbol{\mathcal{RT}}_0(K)$ and normal fluxes uniquely fixing a function $v_h \in \boldsymbol{\mathcal{RT}}_0(K)$ for a simplex $K \in \mathcal{T}_h$. Space dimensions $d = 2$ (left) and $d = 3$ (right)

we write, in several equivalent ways,

$$
\begin{aligned}
\mathcal{P}_k(\mathcal{T}_h) \cap H_0^1(\Omega) &= \{v_h \in H_0^1(\Omega);\ v_h|_K \in \mathcal{P}_k(K) \quad \forall K \in \mathcal{T}_h\} \\
&= \{v_h \in \mathcal{P}_k(\mathcal{T}_h);\ (v_h|_K)|_\sigma = (v_h|_L)|_\sigma \quad \forall \sigma = K \cap L \in \mathcal{F}_h^{\mathrm{int}} \\
&\quad\ \text{and } (v_h|_K)|_\sigma = 0 \quad \forall \sigma \subset K \cap \partial\Omega \in \mathcal{F}_h^{\mathrm{ext}}\} \\
&= \{v_h \in \mathcal{P}_k(\mathcal{T}_h) \cap C^0(\overline{\Omega});\ v_h = 0 \text{ on } \partial\Omega\}.
\end{aligned}
\tag{3.9}
$$

Indeed, Lemma 3.10 holds with $H^1(\mathcal{T}_h)$ replaced by $\mathcal{P}_k(\mathcal{T}_h)$ and $H_0^1(\Omega)$ replaced by $\mathcal{P}_k(\mathcal{T}_h) \cap H_0^1(\Omega)$, where the traces on faces σ now equal the usual restrictions to σ. Congruently, fixing the point values in each simplex K of the mesh $\mathcal{T}_h$ as schematically visualized in Figure 4 by a unique bullet value yields the $H_0^1(\Omega)$- or $C^0(\overline{\Omega})$-conformity for piecewise k-degree polynomials. We again refer for details to [15, 118].

3.5.3 Raviart–Thomas piecewise polynomial subspaces of $\boldsymbol{H}(\mathrm{div}, \Omega)$

Recall that $\mathcal{P}_0(K)$ stands for constants on the simplex $K \in \mathcal{T}_h$. Following Whitney [258], Raviart and Thomas [218], and Nédélec [195], see alternatively Boffi *et al.* [53] or Ern and Guermond [118], we define the space of vector-valued polynomials

$$
\begin{aligned}
\boldsymbol{\mathcal{RT}}_0(K) &:= [\mathcal{P}_0(K)]^d + \boldsymbol{x}\mathcal{P}_0(K) \\
&= \{\boldsymbol{v}_h(\boldsymbol{x}) + w_h(\boldsymbol{x})\boldsymbol{x};\ \boldsymbol{v}_h \in [\mathcal{P}_0(K)]^d,\ w_h \in \mathcal{P}_0(K)\} \\
&= \left\{ \begin{pmatrix} a \\ b \end{pmatrix} + c \begin{pmatrix} x \\ y \end{pmatrix} \quad a, b, c \in \mathbb{R},\ \boldsymbol{x} = (x, y)^{\mathrm{t}} \right\} \quad \text{for } d = 2, \\
&= \left\{ \begin{pmatrix} a \\ b \\ c \end{pmatrix} + d \begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad a, b, c, d \in \mathbb{R},\ \boldsymbol{x} = (x, y, z)^{\mathrm{t}} \right\} \quad \text{for } d = 3.
\end{aligned}
\tag{3.10}
$$

This space is slightly bigger than just constant vector-valued polynomials $[\mathcal{P}_0(K)]^d$ but slightly smaller than affine vector-valued polynomials $[\mathcal{P}_1(K)]^d$. Its dimension equals $d + 1$, i.e., the number of faces of
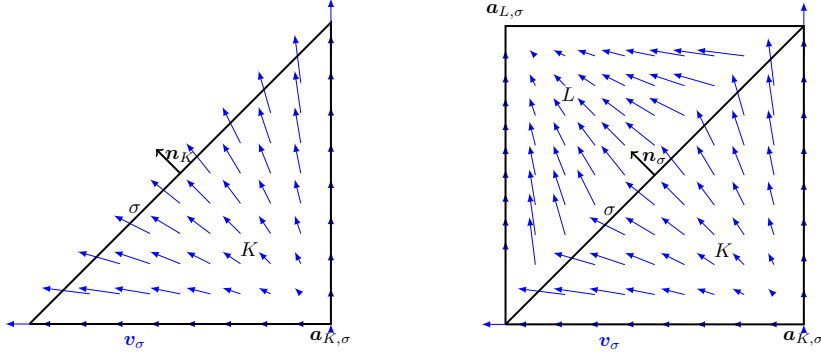
Figure 6: A basis function $\boldsymbol{v}_\sigma$ of $\boldsymbol{\mathcal{RT}}_0(K)$ (supported on the simplex K) with nonzero normal flux only through the face $\sigma \in \mathcal{F}_K$ (left). A basis function $\boldsymbol{v}_\sigma$ of $\boldsymbol{\mathcal{RT}}_0(\mathcal{T}_h) \cap \boldsymbol{H}(\mathrm{div}, \Omega)$ (supported on all (two here) simplices sharing the face σ) with nonzero normal flux only through the face $\sigma \in \mathcal{F}_h$ (right).

the simplex K, which means 3 in two space dimensions and 4 in three space dimensions. Its distinctive properties are

$$\nabla \cdot \boldsymbol{v}_h \in \mathcal{P}_0(K) \quad \text{and} \quad (\boldsymbol{v}_h \cdot \boldsymbol{n}_K)|_\sigma \in \mathcal{P}_0(\sigma) \quad \forall \sigma \in \mathcal{F}_K \qquad \text{for } \boldsymbol{v}_h \in \boldsymbol{\mathcal{RT}}_0(K), \tag{3.11}$$

i.e., the divergence is constant and the normal trace is constant on each face of the simplex K for any function $\boldsymbol{v}_h \in \boldsymbol{\mathcal{RT}}_0(K)$. Let us recall here that for a polynomial $\boldsymbol{v}_h$, the normal trace on a face σ equals the restriction to σ of the normal component $\boldsymbol{v}_h \cdot \boldsymbol{n}_K$. Moreover, a function $\boldsymbol{v}_h$ in $\boldsymbol{\mathcal{RT}}_0(K)$ is uniquely defined by prescribing the face normal fluxes, i.e.,

$$\langle \boldsymbol{v}_h \cdot \boldsymbol{n}_K, 1 \rangle_\sigma := \int_\sigma \boldsymbol{v}_h \cdot \boldsymbol{n}_K \qquad \forall \sigma \in \mathcal{F}_K, \tag{3.12}$$

which is a part of the schematic visualization of the spaces $\boldsymbol{\mathcal{RT}}_0(K)$ for $d = 2, 3$ in Figure 5. We choose the $d + 1$ basis functions $\boldsymbol{v}_\sigma$ as associated with the faces $\sigma \in \mathcal{F}_K$. Setting $\langle \boldsymbol{v}_\sigma \cdot \boldsymbol{n}_K, 1 \rangle_\sigma = 1$ and $\langle \boldsymbol{v}_\sigma \cdot \boldsymbol{n}_K, 1 \rangle_{\sigma'} = 0$ for $\sigma' \neq \sigma$, we obtain the explicit expression

$$\boldsymbol{v}_\sigma(\boldsymbol{x}) = \frac{1}{d|K|}(\boldsymbol{x} - \boldsymbol{a}_{K,\sigma}), \qquad \boldsymbol{x} \in K, \tag{3.13}$$

cf. the illustration in Figure 6, left, where $\boldsymbol{a}_{K,\sigma}$ is the vertex of K opposite to σ.

On the simplicial mesh $\mathcal{T}_h$, we now define

$$\boldsymbol{\mathcal{RT}}_0(\mathcal{T}_h) := \{\boldsymbol{v}_h \in \boldsymbol{L}^2(\Omega); \, \boldsymbol{v}_h|_K \in \boldsymbol{\mathcal{RT}}_0(K) \quad \forall K \in \mathcal{T}_h\}. \tag{3.14}$$

This is a finite-dimensional subspace of the space $\boldsymbol{H}(\mathrm{div}, \mathcal{T}_h)$ from (3.6) but not of the Sobolev space $\boldsymbol{H}(\mathrm{div}, \Omega)$. To create a subspace of $\boldsymbol{H}(\mathrm{div}, \Omega)$, we set

$$\begin{aligned}
\boldsymbol{\mathcal{RT}}_0(\mathcal{T}_h) \cap \boldsymbol{H}(\mathrm{div}, \Omega) &= \{\boldsymbol{v}_h \in \boldsymbol{H}(\mathrm{div}, \Omega); \, \boldsymbol{v}_h|_K \in \boldsymbol{\mathcal{RT}}_0(K) \quad \forall K \in \mathcal{T}_h\} \\
&= \{\boldsymbol{v}_h \in \boldsymbol{\mathcal{RT}}_0(\mathcal{T}_h); \, (\boldsymbol{v}_h|_K \cdot \boldsymbol{n}_K)|_\sigma = -(\boldsymbol{v}_h|_L \cdot \boldsymbol{n}_L)|_\sigma \quad \forall \sigma = K \cap L \in \mathcal{F}_h^{\mathrm{int}}\};
\end{aligned} \tag{3.15}$$

it follows by Lemma 3.11 that the above characterizations coincide. Congruently, fixing the face normal values/integrals in each simplex $K \in \mathcal{T}_h$ as schematically visualized in Figure 5 by the arrows in particular yields the $\boldsymbol{H}(\mathrm{div}, \Omega)$-conformity for piecewise $\boldsymbol{\mathcal{RT}}_0(K)$-polynomials. The space (3.15) of normal-component-continuous piecewise vector-valued polynomials is called the *Raviart–Thomas(–Nédélec)* space. Its dimension is the number of mesh faces in the mesh $\mathcal{T}_h$. There is one basis function $\boldsymbol{v}_\sigma$ for each face $\sigma \in \mathcal{F}_h$, prescribed such that $\langle \boldsymbol{v}_\sigma \cdot \boldsymbol{n}_\sigma, 1 \rangle_\sigma = 1$ and $\langle \boldsymbol{v}_\sigma \cdot \boldsymbol{n}_{\sigma'}, 1 \rangle_{\sigma'} = 0$ for all $\sigma' \in \mathcal{F}_h$ with $\sigma' \neq \sigma$. An illustration is given in Figure 6, right. We again refer for details to [15, 118].

Remark 3.12 (Meshes consisting of rectangular parallelepipeds). *Above, we only consider simplicial meshes since these are our key to treat general polytopal meshes below. In practice, meshes consisting of rectangles, blocks, or in general rectangular parallelepipeds are often used. On rectangles or blocks, in place of (3.8), piecewise polynomial subspaces of $L^2(\Omega)$ are typically chosen as*

$$\mathcal{Q}_k(\mathcal{T}_h) := \{v_h \in L^2(\Omega); \, v_h|_K \in \mathcal{Q}_k(K) \quad \forall K \in \mathcal{T}_h\}, \tag{3.16}$$
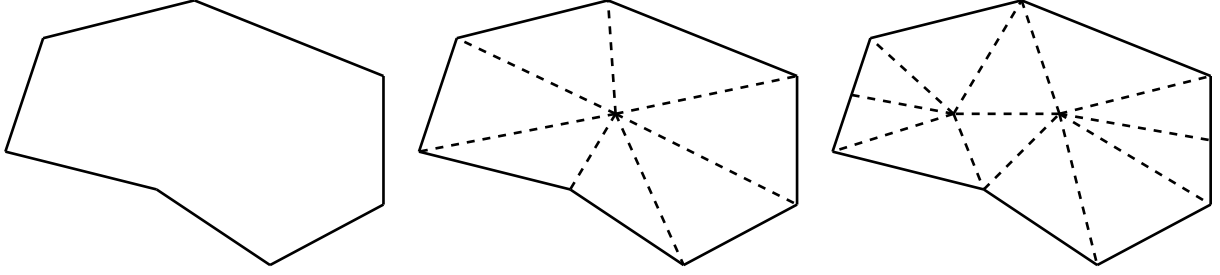
Figure 7: Polygonal element $K \in \mathcal{T}_H$ (left). Virtual simplicial submesh $\mathcal{T}_K$ of K (middle and right). One point inside $K \in \mathcal{T}_H$ shared by all the simplices in $\mathcal{T}_K$ and faces of $\mathcal{T}_h$ do not subdividing the faces of $\mathcal{T}_H$ (middle). General situation (right).

where $\mathcal{Q}_k(K)$ is the space of scalar-valued polynomials on K of degree at most k in each space coordinate. Note that $\mathcal{P}_k(\mathcal{T}_h)$ and $\mathcal{Q}_k(\mathcal{T}_h)$ coincide for the lowest polynomial degree $k = 0$, being simply piecewise constants with respect to the mesh $\mathcal{T}_h$. The Raviart–Thomas spaces $\boldsymbol{\mathcal{RT}}_0(K)$ then take the form

$$\boldsymbol{\mathcal{RT}}_0(K) := \left\{ \begin{pmatrix} a + cx \\ b + dy \end{pmatrix} \quad a, b, c, d \in \mathbb{R}, \ \boldsymbol{x} = (x, y)^{\mathrm{t}} \right\} \quad \textit{for } d = 2, \tag{3.17a}$$

and

$$\boldsymbol{\mathcal{RT}}_0(K) := \left\{ \begin{pmatrix} a + dx \\ b + ey \\ c + fz \end{pmatrix} \quad a, b, c, d, e, f \in \mathbb{R}, \ \boldsymbol{x} = (x, y, z)^{\mathrm{t}} \right\} \quad \textit{for } d = 3, \tag{3.17b}$$

in place of (3.10). Subspaces of $\boldsymbol{H}(\mathrm{div}, \Omega)$ are then still created by (3.14)–(3.15).

In the following, we will sometimes employ the above notation and developments for subdomains $\omega \subset \Omega$ and the corresponding submeshes $\mathcal{T}_\omega$ of ω.

3.6 Polytopal meshes with virtual simplicial submeshes

Recall that $\Omega \subset \mathbb{R}^d$, $d \geq 1$, is an open bounded connected polytope with a Lipschitz-continuous boundary $\partial\Omega$. Meshes with mesh elements formed by polygons (in two space dimensions), polyhedrons (in three space dimensions), or in general polytopes (in d space dimensions) are useful in practice. An example of a general polygonal element ($d = 2$) is given in Figure 7 (left); some examples of polygonal meshes can be found in Figures 8 (left) and 32 and of polyhedral meshes in Figures 34 and 36. These meshes stand in contrast to regular meshes of rectangular parallelepipeds by their generality and, in comparison with meshes formed by simplicial elements, they may contain (much) fewer elements, cf. Figure 8 (right).

3.6.1 Polytopal mesh with a virtual simplicial submesh

We call a polytopal mesh a partition $\mathcal{T}_H$ of the domain Ω into a finite collection of closed d-dimensional polytopes K homotopic to a ball and with Lipschitz-continuous boundary ∂K such that $\cup_{K \in \mathcal{T}_H} K = \overline{\Omega}$. We suppose that the intersection of two different polytopes is either empty or their entire l-dimensional face, $0 \leq l \leq (d-1)$, or a collection of their entire l-dimensional faces, $0 \leq l \leq (d-1)$. The elements K can in particular be nonconvex and non star-shaped. The structural assumption we impose is that there exists a virtual, not to be constructed in practice, simplicial mesh $\mathcal{T}_h$ of Ω in the sense of Section 3.3 such that $\mathcal{T}_K := \mathcal{T}_h|_K$ is a simplicial mesh of the polytope K. An illustration is provided in Figures 7 and 8. Altogether, each element $\kappa \in \mathcal{T}_h$ is a d-simplex, $\cup_{\kappa \in \mathcal{T}_h} \kappa = \overline{\Omega}$, $\cup_{\kappa \in \mathcal{T}_K} \kappa = K$, and $\cup_{K \in \mathcal{T}_H} \mathcal{T}_K = \mathcal{T}_h$.

3.6.2 Mesh faces and mesh vertices

Let $\mathcal{F}_H$ be the set of the $(d-1)$-dimensional faces of $\mathcal{T}_H$. We divide it into interior faces $\mathcal{F}_H^{\mathrm{int}}$ and boundary faces $\mathcal{F}_H^{\mathrm{ext}}$: $\sigma \in \mathcal{F}_H^{\mathrm{int}}$ if there exist $K, L \in \mathcal{T}_H$, $K \neq L$, such that $\sigma \subset K \cap L$, and $\sigma \in \mathcal{F}_H^{\mathrm{ext}}$ if there exist $K \in \mathcal{T}_H$ such that $\sigma \subset K \cap \partial\Omega$. Recall from Section 3.3 that $\mathcal{F}_h$ is the set of the $(d-1)$-dimensional faces of the simplicial mesh $\mathcal{T}_h$. We denote by $\mathcal{F}_{H,h}$ the set of such faces from $\mathcal{F}_h$ that lie in some polytopal face of $\mathcal{F}_H$. Let $\mathcal{F}_K \subset \mathcal{F}_H$ be the set of the $(d-1)$-dimensional faces of the polytope $K \in \mathcal{T}_H$, and let $\mathcal{F}_{K,h}^{\mathrm{ext}}$ collect those faces of $\mathcal{F}_{H,h}$ that lie on the boundary of the element $K \in \mathcal{T}_H$, whereas $\mathcal{F}_{K,h}^{\mathrm{int}}$ those faces of $\mathcal{F}_h$ that lie inside the element $K \in \mathcal{T}_H$. We also set $\mathcal{F}_{K,h} := \mathcal{F}_{K,h}^{\mathrm{ext}} \cup \mathcal{F}_{K,h}^{\mathrm{int}}$ see Figure 9. If the
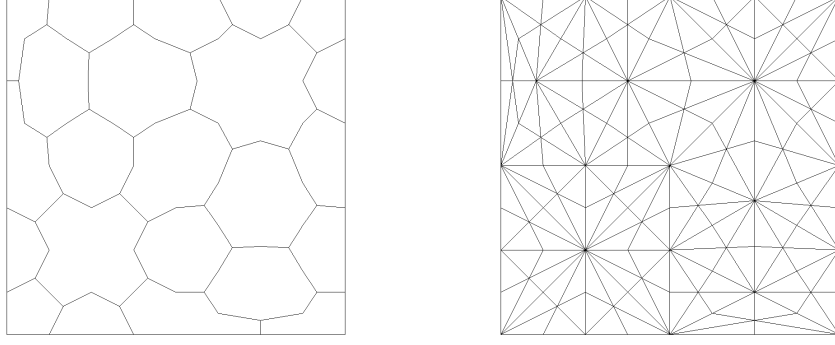
Figure 8: A polygonal mesh $\mathcal{T}_H$ and the corresponding triangular submesh $\mathcal{T}_h$

$$\mathcal{F}_K = \{\sigma_i\}_{i=1}^{7}$$
$$\mathsf{U}_K^{\text{ext}} = \{(\mathsf{U}_K)_{\sigma_i}\}_{i=1}^{7}$$
$$\mathcal{T}_K = \{\kappa_i\}_{i=1}^{7}$$
$$\mathcal{F}_{K,h}^{\text{ext}} = \{\sigma_i\}_{i=1}^{7}$$
$$\mathcal{F}_{K,h}^{\text{int}} = \{\sigma_i\}_{i=8}^{14}$$
$$\mathcal{F}_{K,h} = \mathcal{F}_{K,h}^{\text{ext}} \cup \mathcal{F}_{K,h}^{\text{int}}$$
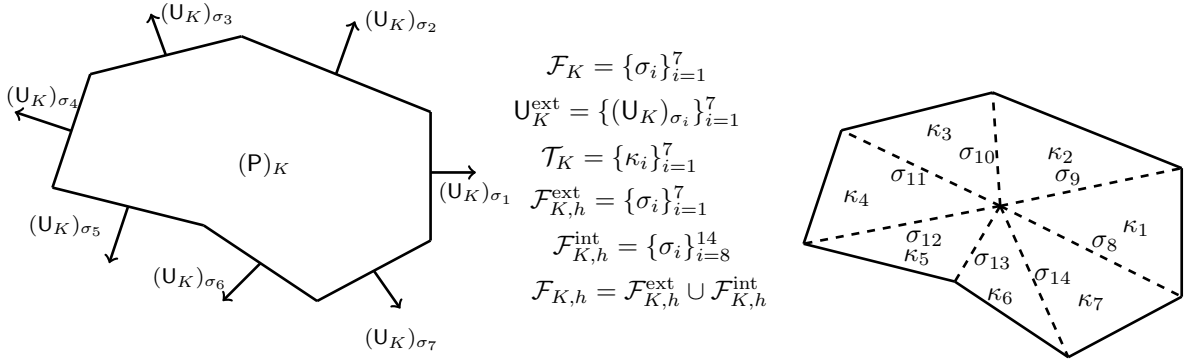
Figure 9: Example of a polygonal element K with its faces $\mathcal{F}_K$, corresponding face fluxes $\mathsf{U}_K^{\text{ext}}$, and pressure head $(\mathsf{P})_K$ (left); virtual simplicial submesh $\mathcal{T}_K$ of K (right)

faces of $\mathcal{T}_h$ do not subdivide the faces of $\mathcal{T}_H$, then $\mathcal{F}_{H,h} = \mathcal{F}_H$ and also $\mathcal{F}_{K,h}^{\text{ext}} = \mathcal{F}_K$. For an interior face $\sigma \in \mathcal{F}_H^{\text{int}}$, we fix an arbitrary orientation and denote the corresponding unit normal vector by $\boldsymbol{n}_\sigma$. For a boundary face $\sigma \in \mathcal{F}_H^{\text{ext}}$, $\boldsymbol{n}_\sigma$ coincides with the exterior unit normal $\boldsymbol{n}_\Omega$ of Ω. Finally, for all $K \in \mathcal{T}_H$, we denote by $\boldsymbol{n}_K$ the unit normal vector to σ pointing out of K. To complete the notation, we let $\mathcal{V}_{K,h}$ to be the set of vertices of all the elements $\kappa \in \mathcal{T}_K$, and, as for simplicial meshes, $\mathcal{V}_\sigma$ the set of vertices of the given face σ.

3.7 Basic principle of finite volume and related lowest-order locally conservative methods

We now introduce the basic principle of the finite volume and related lowest-order locally conservative methods considered in this contribution. In Sections 4–7 below, for each considered partial differential equation, a general lowest-order locally conservative discretization is introduced in detail.

Consider a steady diffusion partial differential equation of the form (2.1)–(2.2) and recall (2.5). A general lowest-order locally conservative method looks for real values $U_{K,\sigma}$ which for each polytopal mesh element $K \in \mathcal{T}_H$ and each of its faces $\sigma \in \mathcal{F}_K$ approximate the normal outflux $\langle \boldsymbol{u}\cdot\boldsymbol{n}_K, 1\rangle_\sigma = \int_\sigma \boldsymbol{u}\cdot\boldsymbol{n}_K$ from K through the face σ such that

$$\sum_{\sigma \in \mathcal{F}_K} U_{K,\sigma} = (f,1)_K \qquad \forall K \in \mathcal{T}_H. \tag{3.18a}$$

Here the sum runs through all the faces of the polytope K and, recalling the notation (3.1a), $(f,1)_K = \int_K f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$ is the integral of the source term. From a neighboring element $L \in \mathcal{T}_H$ sharing the same face σ, we then impose,

$$U_{L,\sigma} = -U_{K,\sigma}. \tag{3.18b}$$

Requirement (3.18b) ensures the normal-trace continuity, as in the Raviart–Thomas space $\boldsymbol{\mathcal{RT}}_0(\mathcal{T}_h) \cap \boldsymbol{H}(\text{div}, \Omega)$ of (3.15). Then (3.18b) expresses the conservation of mass across mesh faces and (3.18a) the

18

equilibrium with the load, respectively the discrete versions of $\boldsymbol{u} \in \boldsymbol{H}(\mathrm{div}, \Omega)$ and $\nabla \cdot \boldsymbol{u} = f$. Mimicking the remaining items from (2.5), namely $p \in H_0^1(\Omega)$ and $\boldsymbol{u} = -\nabla p$, then have to be built in the construction of the discrete normal fluxes $U_{K,\sigma}$ from elementwise approximations of the primal variable p_K.

For an unsteady problem of the form (2.3)–(2.4), we consider an implicit Euler time discretization on discrete times $0 = t^0 < \ldots < t^n < \ldots t^N = t_{\mathrm{F}}$ with time steps $\tau^n := t^n - t^{n-1}$. We then get, for each discrete time t^n, $1 \le n \le N$,

$$\frac{p_K^n - p_K^{n-1}}{\tau^n} + \sum_{\sigma \in \mathcal{F}_K} U_{K,\sigma}^n = (f^n, 1)_K \qquad \forall K \in \mathcal{T}_H \tag{3.19a}$$

and

$$U_{L,\sigma}^n = -U_{K,\sigma}^n, \tag{3.19b}$$

where $f^n := \int_{t^{n-1}}^{t^n} f / \tau^n$. Again, the heart of each method is the construction of the discrete normal fluxes $U_{K,\sigma}^n$ for each polytopal mesh element $K \in \mathcal{T}_H$ and each face $\sigma \in \mathcal{F}_K$ from the elementwise approximations of the primal variable p_K^n.

3.8 Bibliographic resources

Details on the material rapidly exposed in Sections 3.2–3.6 can be found in Adams [5], Thomas [237], Ciarlet [92], Allaire [15], Brenner and Scott [61], Boffi *et al.* [53], Ern and Guermond [118], and Vohralík [251, 252].

Finite volume schemes like (3.18) and (3.19) are very natural and used traditionally in the engineering practice. Amongst first mathematical analyses for diffusion problems, one can cite Manteuffel and White [184], Rose [225], Morton and Süli [191], Faille [132], Herbin [149], and the overview in Eymard *et al.* [126]. For the closely related mixed finite elements, the construction of the spaces in the lowest-order case can be traced back to Whitney [258], whereas higher-order extensions and analysis have been undertaken in Raviart and Thomas [218] and Nédélec [195], see Brezzi and Fortin [63] for an early overview. The ample connections between finite volumes and mixed finite elements are discussed in, e.g., Russell and Wheeler [227], Agouzal *et al.* [7], Baranger *et al.* [32], Younès *et al.* [260, 259], Chavent *et al.* [79], Klausen and Russell [164], and Vohralík [246].

Numerous extensions of the initial approaches together with generalizations to polytopal meshes, including a priori analysis (existence, uniqueness, convergence, a priori error estimates), have been undertaken recently. Let us cite multi-point finite volumes by Aavatsmark *et al.* [1, 2], Edwards [113], and Breil and Maire [58], mimetic finite differences by Brezzi *et al.* [64, 65], mixed and hybrid finite volumes by Eymard *et al.* [108, 128], multipoint flux mixed finite elements by Wheeler and Yotov *et al.* [257, 256], compatible discrete operator schemes by Bonelle and Ern [55], mixed virtual elements by Brezzi *et al.* [62] and Beirão da Veiga *et al.* [40], hybrid high-order methods by Di Pietro, Ern, and Lemaire [102, 101], or mixed finite elements in [172, 173, 230, 88, 254], see also [52, 150, 87, 107]; unifying frameworks can be found in Droniou *et al.* [110, 111], Vohralík and Wohlmuth [254], Cockburn *et al.* [95], Boffi and Di Pietro [54], and the references therein. Amongst recent books, let us cite Jovanović and Süli [155] for finite differences, Beirão da Veiga *et al.* [41] for mimetic finite differences, Droniou *et al.* [109] for a unifying approach, and Nordbotten and Keilegavlen [199] for multi-point finite volumes.

4 The Poisson equation. Simplicial meshes and spatial discretization error

In this section, we start with the simplest partial differential equation for diffusion processes: the Poisson equation. We first recall the definition of a weak solution based on the Sobolev space $H_0^1(\Omega)$ of Section 3.2.1 and show that the associated flux lies in the Sobolev space $\boldsymbol{H}(\mathrm{div}, \Omega)$ of Section 3.2.2. This is very instrumental and will be repeated in all generalizations of Sections 5–7 below. We then define the energy error and explore its structure for a completely generic approximate solution. We subsequently consider a basic cell-centered finite volume discretization incarnating the principles of Section 3.7. We only employ in this section a simplicial mesh as per Section 3.3 and define the approximate solution by lifting the face fluxes into a Raviart–Thomas piecewise polynomial space of Section 3.5.3. Next, we introduce and discuss two central notions of this work: the flux and potential reconstructions. These in particular lead to a posteriori error estimates fulfilling all the building principles of Section 2. Illustrative numerical experiments close this section.

4.1 The Poisson equation

In this section, for the domain Ω as specified in Section 3.1.1, we consider the problem of finding $p : \Omega \to \mathbb{R}$ such that

$$
\begin{array}{rll}
-\nabla{\cdot}(\nabla p) = f & \quad \text{in } \Omega, & \text{(4.1a)} \\
p = 0 & \quad \text{on } \partial\Omega. & \text{(4.1b)}
\end{array}
$$

Here $f \in L^2(\Omega)$ is a source term.

4.2 Weak solution, potential, and flux

The Sobolev space $H_0^1(\Omega)$ from Definition 3.4 gives a proper mathematical setting to define a unique solution p of (4.1). Indeed, there is one and only one $p \in H_0^1(\Omega)$ such that

$$(\nabla p, \nabla v) = (f, v) \qquad \forall v \in H_0^1(\Omega); \tag{4.2}$$

note that this follows trivially from the Riesz representation theorem, since $H_0^1(\Omega)$ is a Hilbert space for the scalar product given by $(\nabla p, \nabla v)$ and since (f, v) is bounded linear form over $H_0^1(\Omega)$. We call (4.2) the weak formulation of (4.1). From the weak solution p, we can define the dual variable

$$\boldsymbol{u} := -\nabla p. \tag{4.3}$$

In relation to the applications below, p may be a pressure and $\boldsymbol{u}$ a Darcy velocity, but we rather generically refer to p as *potential* and to $\boldsymbol{u}$ as *flux*.

4.3 Properties of the exact potential and flux

Using Definitions 3.4 and 3.7 of the spaces $H_0^1(\Omega)$ and $\boldsymbol{H}(\text{div}, \Omega)$, we have the following important result, detailing what we have discovered in (2.5):

Proposition 4.1 (Properties of the weak solution (4.2))**.** *Let p be the solution of* (4.2)*. Let $\boldsymbol{u}$ be given by* (4.3)*. Then*

$$p \in H_0^1(\Omega), \quad \boldsymbol{u} \in \boldsymbol{H}(\text{div}, \Omega), \quad \nabla{\cdot}\boldsymbol{u} = f. \tag{4.4}$$

Proof. The weak solution p belongs to $H_0^1(\Omega)$ by definition. In order to verify that $\boldsymbol{u} \in \boldsymbol{H}(\text{div}, \Omega)$ and $\nabla{\cdot}\boldsymbol{u} = f$, we need to check the three conditions of Definition 3.6. Condition 1 is obvious, as $p \in H_0^1(\Omega)$ and thus $-\nabla p = \boldsymbol{u}$ is square-integrable in view of Definition 3.1, condition 2a. For the function w of condition 2a in Definition 3.6, choose $w := f$ and note that $f \in L^2(\Omega)$ by assumption. Then condition 2b follows immediately from (4.2) and the fact that $\mathcal{D}(\Omega) \subset H_0^1(\Omega)$. $\qquad\square$

Remark 4.2 (Mathematical properties of the weak solution)**.** *It is remarkable that the flux $\boldsymbol{u}$ prescribed from the potential p by* (4.3) *belongs to $\boldsymbol{H}(\text{div}, \Omega)$ and has a divergence in equilibrium with the load f, since this is not explicitly stipulated in the weak formulation* (4.2)*. Congruently, it immediately follows (cf. Proposition 4.4 below) that the three properties* (4.4) *together with the constitutive relation* (4.3) *actually define the weak solution p of* (4.2) *in a unique way. This mathematically describes the intrinsic properties of the exact potential p and flux $\boldsymbol{u}$: p has to be trace continuous and $\boldsymbol{u}$ has to be related by the constitutive relation* (4.3) *to it, normal-trace continuous, and in equilibrium with the load.*

Remark 4.3 (Physical properties of the weak solution)**.** *Scalar-valued physical variables such as the potential (pressure) p are naturally trace continuous, whereas vector-valued physical variables such as the flux (Darcy velocity) $\boldsymbol{u} = -\nabla p$ are naturally normal-trace continuous, cf. Figures 1 and 2. Moreover, $\boldsymbol{u}$ has to be in equilibrium with the load (locally mass conservative) which amounts to $\nabla{\cdot}\boldsymbol{u} = f$. This restates that the three properties* (4.4) *together with the constitutive relation* (4.3) *characterize the exact solution.*

4.4 Approximate solution and spatial discretization error

Let $\boldsymbol{u}_h \in \boldsymbol{L}^2(\Omega)$ be an arbitrary function that we think of a (numerical) approximation of the exact flux $\boldsymbol{u}$ from (4.2)–(4.3). We will give a specific example for the cell-centered finite volume discretizations on a simplicial mesh in Section 4.8 below, but we want to proceed abstractly for the moment, since there is no link to a mesh or to the discrete world of piecewise polynomials at this stage. We will call the $\boldsymbol{L}^2(\Omega)$-distance of $\boldsymbol{u}_h$ to $\boldsymbol{u}$,

$$\|\boldsymbol{u} - \boldsymbol{u}_h\|, \tag{4.5}$$

the *spatial discretization error*.

4.5 Prager–Synge equality

We now state the following important result dating back to Prager and Synge [216]:

Proposition 4.4 (Prager–Synge equality). *Let $p \in H_0^1(\Omega)$ be the weak solution of (4.2) and let $\zeta_h \in H_0^1(\Omega)$ and $\boldsymbol{\sigma}_h \in \boldsymbol{H}(\mathrm{div}, \Omega)$ with $\nabla \cdot \boldsymbol{\sigma}_h = f$ be arbitrary. Then*

$$\|\nabla(p - \zeta_h)\|^2 + \|\nabla p + \boldsymbol{\sigma}_h\|^2 = \|\nabla\zeta_h + \boldsymbol{\sigma}_h\|^2. \tag{4.6}$$

Proof. Adding and subtracting ∇p, we develop

$$\|\nabla\zeta_h + \boldsymbol{\sigma}_h\|^2 = \|\nabla(\zeta_h - p) + \nabla p + \boldsymbol{\sigma}_h\|^2$$
$$= \|\nabla(\zeta_h - p)\|^2 + \|\nabla p + \boldsymbol{\sigma}_h\|^2 + 2(\nabla(\zeta_h - p), \nabla p + \boldsymbol{\sigma}_h).$$

Note from Proposition 4.1 that $\nabla p \in \boldsymbol{H}(\mathrm{div}, \Omega)$ with $\nabla \cdot (\nabla p) = -f$. Thus $(\nabla p + \boldsymbol{\sigma}_h) \in \boldsymbol{H}(\mathrm{div}, \Omega)$ and in particular $\nabla \cdot (\nabla p + \boldsymbol{\sigma}_h) = 0$. Thus, using that $\zeta_h - p \in H_0^1(\Omega)$, the Green theorem (3.3) gives

$$(\underbrace{\nabla p + \boldsymbol{\sigma}_h}_{\in \boldsymbol{H}(\mathrm{div}, \Omega)}, \nabla(\underbrace{\zeta_h - p}_{\in H_0^1(\Omega)})) \overset{(3.3)}{=} -(\underbrace{\nabla \cdot (\nabla p + \boldsymbol{\sigma}_h)}_{=0}, \zeta_h - p) = 0,$$

whence the assertion follows. $\qquad\square$

Remark 4.5 (Prager–Synge equality (4.6)). *We stress that (4.6) is an equality, where on the right-hand side, there are, in practice, two known, discrete objects from the piecewise polynomial subspaces of $H_0^1(\Omega)$ and $\boldsymbol{H}(\mathrm{div}, \Omega)$, whereas the left-hand side are their spatial discretization errors with respect to the weak solution of (4.2). In other words, if we want to trace a simultaneous potential and flux error, then we* can *compute it, there is no (a posteriori) estimate.*

From (4.6), we immediately observe (cf. Theorem 6.1 and Corollary 6.6 in [250] and the references therein):

Corollary 4.6 (Prager–Synge (in)equality). *Let $p \in H_0^1(\Omega)$ be the solution of (4.2), let $\boldsymbol{u} = -\nabla p$ from (4.3), and let $\boldsymbol{u}_h \in \boldsymbol{H}(\mathrm{div}, \Omega)$ with $\nabla \cdot \boldsymbol{u}_h = f$ be arbitrary. Then*

$$\|\boldsymbol{u} - \boldsymbol{u}_h\| = \min_{v \in H_0^1(\Omega)} \|\boldsymbol{u}_h + \nabla v\|, \tag{4.7}$$

so that in particular

$$\|\boldsymbol{u} - \boldsymbol{u}_h\| \le \|\boldsymbol{u}_h + \nabla\zeta_h\| \tag{4.8}$$

for an arbitrary $\zeta_h \in H_0^1(\Omega)$.

Proof. The second assertion (4.8) follows from (4.6) upon taking $\boldsymbol{\sigma}_h = \boldsymbol{u}_h$ and noticing that

$$\|\nabla p + \boldsymbol{u}_h\|^2 = \|\nabla\zeta_h + \boldsymbol{u}_h\|^2 - \|\nabla(p - \zeta_h)\|^2 \le \|\nabla\zeta_h + \boldsymbol{u}_h\|^2$$

for any $\zeta_h \in H_0^1(\Omega)$, using $\boldsymbol{u} = -\nabla p$ from (4.3). This immediately gives (4.7) with infimum in place of minimum and the $\le$ sign in place of the equality. Since we can take $\zeta_h = p$ and since $\boldsymbol{u} = -\nabla p$, the minimum in (4.7) is actually attained and the right-hand side equals the left-hand side. $\qquad\square$

Remark 4.7 (Prager–Synge equality (4.7)). *According to (4.3) and (4.4), to be equal to the exact flux $-\nabla p$, any $\boldsymbol{u}_h \in \boldsymbol{H}(\mathrm{div}, \Omega)$ with $\nabla \cdot \boldsymbol{u}_h = f$ only misses being a minus weak gradient of some function from $H_0^1(\Omega)$. This is the meaning of the characterization (4.7). Notice that (4.7) expresses the distance of $\boldsymbol{u}_h$ to $\nabla H_0^1(\Omega)$.*

Remark 4.8 (Prager–Synge inequality (4.8)). *Having $\boldsymbol{u}_h \in \boldsymbol{H}(\mathrm{div}, \Omega)$ with $\nabla \cdot \boldsymbol{u}_h = f$ at disposal, which we will see in Section 4.8 below is immediate in finite volume methods for the Poisson equation (4.1), when f is piecewise constant, the Prager–Synge inequality (4.8) can be readily used to produce a guaranteed error upper bound also satisfying the other requirements formulated in Section 2.5. This will still be the situation for the singlephase steady linear Darcy flow in Section 5 below, but when we want to take into account iterative algebraic solvers in Section 6 and component fluxes together with nonlinear zero-order terms in Section 7 for nonlinear compositional Darcy flows, it may not be feasible to locally construct $\boldsymbol{u}_h \in \boldsymbol{H}(\mathrm{div}, \Omega)$ with $\nabla \cdot \boldsymbol{u}_h = f$ as an approximate solution. There, we will in general only have $\boldsymbol{u}_h \in \boldsymbol{L}^2(\Omega)$. In this case, we will reconstruct $\boldsymbol{\sigma}_h^{k,i} \in \boldsymbol{H}(\mathrm{div}, \Omega)$ with $\nabla \cdot \boldsymbol{\sigma}_h^{k,i} = f$ or $\nabla \cdot \boldsymbol{\sigma}_h^{k,i} = f + \rho_h^{k,i}$ with $\rho_h^{k,i}$ a remainder made as small as necessary.*

4.6 Equivalence of the spatial discretization error to the flux nonconformity (dual norm of the residual) plus potential nonconformity

Following [213, Theorem 3.1], [125, Theorem 3.3], and the references therein, we now extend Corollary 4.6 to the general case $\boldsymbol{u}_h \in \boldsymbol{L}^2(\Omega)$.

Theorem 4.9 (Error characterization for $\boldsymbol{u}_h \in \boldsymbol{L}^2(\Omega)$)**.** *Let $p \in H_0^1(\Omega)$ be the weak solution of* (4.2), *let $\boldsymbol{u} = -\nabla p$ from* (4.3), *and let $\boldsymbol{u}_h \in \boldsymbol{L}^2(\Omega)$ be arbitrary. Then*

$$\|\boldsymbol{u} - \boldsymbol{u}_h\|^2 = \underbrace{\min_{\substack{\boldsymbol{v} \in \boldsymbol{H}(\mathrm{div},\Omega) \\ \nabla \cdot \boldsymbol{v} = f}} \|\boldsymbol{u}_h - \boldsymbol{v}\|^2}_{\substack{\max_{\varphi \in H_0^1(\Omega);\, \|\nabla\varphi\|=1}\{(f,\varphi)+(\boldsymbol{u}_h,\nabla\varphi)\}^2, \\ \textit{dual norm of the residual of } \boldsymbol{u}_h}} + \underbrace{\min_{v \in H_0^1(\Omega)} \|\boldsymbol{u}_h + \nabla v\|^2}_{\substack{\textit{nonconformity} \\ (\textit{distance of } \boldsymbol{u}_h \textit{ to } \nabla H_0^1(\Omega))}} . \qquad (4.9)$$

Remark 4.10 (Error characterization (4.9))**.** *Relation* (4.9) *extends* (4.7) *from $\boldsymbol{u}_h \in \boldsymbol{H}(\mathrm{div},\Omega)$ with $\nabla \cdot \boldsymbol{u}_h = f$ to any $\boldsymbol{u}_h \in \boldsymbol{L}^2(\Omega)$. The supplementary arising term is naturally the distance of $\boldsymbol{u}_h$ to $\boldsymbol{H}(\mathrm{div},\Omega)$ under the constraint of the divergence being equal to f.*

Remark 4.11 (Dual norm of the residual)**.** *The first term in* (4.9) *has an equivalent expression investigating "how much, for a test function $\varphi \in H_0^1(\Omega)$, $\boldsymbol{u}_h$ misses to fulfill the weak formulation* (4.2) *when put in place of $-\nabla p$", which is related to the residual of* (4.2) *for φ. The overall maximum over all φ from $H_0^1(\Omega)$ with $\|\nabla\varphi\| = 1$ is then the dual norm of the residual.*

Remark 4.12 (Potential nonconformity)**.** *The second term in* (4.9) *expresses the distance of $\boldsymbol{u}_h$ to $\nabla H_0^1(\Omega)$, cf. Remark 4.7.*

Proof of Theorem 4.9. Let us define a function $\zeta \in H_0^1(\Omega)$ by

$$(\nabla \zeta, \nabla v) = -(\boldsymbol{u}_h, \nabla v) \qquad \forall v \in H_0^1(\Omega). \qquad (4.10)$$

There exists one and only one ζ by the Riesz representation theorem; indeed, recall from Section 3.2.1 that the left-hand side of (4.10) is a scalar product on $H_0^1(\Omega)$ and notice that from $|-(\boldsymbol{u}_h, \nabla v)| \leq \|\boldsymbol{u}_h\|\|\nabla v\|$, the right-hand side of (4.10) is a bounded linear form on $H_0^1(\Omega)$ ($\boldsymbol{u}_h$ is here a fixed datum). The function ζ can be seen as the orthogonal projection of the approximate solution $\boldsymbol{u}_h$ onto $\nabla H_0^1(\Omega)$. With the aid of ζ, we can thus write the Pythagorean equality

$$\|\nabla p + \boldsymbol{u}_h\|^2 = \|\nabla(p - \zeta)\|^2 + \|\nabla \zeta + \boldsymbol{u}_h\|^2. \qquad (4.11)$$

Indeed,

$$\|\nabla p + \boldsymbol{u}_h\|^2 = \|\nabla p - \nabla \zeta + \nabla \zeta + \boldsymbol{u}_h\|^2 = \|\nabla(p - \zeta)\|^2 + \|\nabla \zeta + \boldsymbol{u}_h\|^2 + 2(\nabla(p - \zeta), \nabla \zeta + \boldsymbol{u}_h),$$

and the last term in the above expression vanishes in view of the orthogonality (4.10), since $p - \zeta$ can be taken as a test function $v \in H_0^1(\Omega)$ in (4.10). We continue in three steps.

1) Since ζ can be seen as a projection of $\boldsymbol{u}_h$,

$$\|\nabla \zeta + \boldsymbol{u}_h\|^2 = \min_{v \in H_0^1(\Omega)} \|\nabla v + \boldsymbol{u}_h\|^2. \qquad (4.12)$$

Indeed, from (4.11) used for an arbitrary function $v \in H_0^1(\Omega)$ in place of $p \in H_0^1(\Omega)$, we see

$$\|\nabla v + \boldsymbol{u}_h\|^2 = \|\nabla(v - \zeta)\|^2 + \|\nabla \zeta + \boldsymbol{u}_h\|^2.$$

Therefrom, we get

$$\|\nabla \zeta + \boldsymbol{u}_h\|^2 = \|\nabla v + \boldsymbol{u}_h\|^2 - \|\nabla(v - \zeta)\|^2 \leq \|\nabla v + \boldsymbol{u}_h\|^2 \qquad \forall v \in H_0^1(\Omega),$$

and there is an equality for $v = \zeta$. This shows that the second term in (4.11) indeed takes the form needed in (4.9).

2) For the first term in (4.11), we first use that $p - \zeta \in H_0^1(\Omega)$. Thus, the dual norm characterization and (4.10) give

$$\|\nabla(p - \zeta)\| = \max_{\substack{\varphi \in H_0^1(\Omega) \\ \|\nabla\varphi\|=1}} (\nabla(p - \zeta), \nabla \varphi) \overset{(4.10)}{=} \max_{\substack{\varphi \in H_0^1(\Omega) \\ \|\nabla\varphi\|=1}} (\nabla p + \boldsymbol{u}_h, \nabla \varphi). \qquad (4.13)$$

Let now $\varphi \in H_0^1(\Omega)$ with $\|\nabla\varphi\| = 1$ be fixed. Using the characterization (4.2) of the weak solution, we have

$$(\nabla p + \boldsymbol{u}_h, \nabla\varphi) \overset{(4.2)}{=} (f, \varphi) + (\boldsymbol{u}_h, \nabla\varphi). \tag{4.14}$$

Thus, we have (4.9) with the first term in the form expressed in the underbrace.

3) We are left to show the equivalence

$$\min_{\substack{\boldsymbol{v} \in \boldsymbol{H}(\mathrm{div},\Omega) \\ \nabla\cdot\boldsymbol{v}=f}} \|\boldsymbol{u}_h - \boldsymbol{v}\| = \max_{\varphi \in H_0^1(\Omega); \|\nabla\varphi\|=1} \{(f, \varphi) + (\boldsymbol{u}_h, \nabla\varphi)\}. \tag{4.15}$$

3a) First, to motivate, we show that the right-hand side of (4.15) is bounded by the left-hand side one. Let $\boldsymbol{v} \in \boldsymbol{H}(\mathrm{div},\Omega)$ such that $\nabla\cdot\boldsymbol{v} = f$ be arbitrary. Then the Green theorem (3.3) gives

$$(f, \varphi) + (\boldsymbol{u}_h, \nabla\varphi) = (\nabla\cdot\boldsymbol{v}, \varphi) + (\boldsymbol{u}_h, \nabla\varphi) = (\boldsymbol{u}_h - \boldsymbol{v}, \nabla\varphi).$$

Consequently, by the Cauchy–Schwarz inequality,

$$\max_{\varphi \in H_0^1(\Omega); \|\nabla\varphi\|=1} \{(f, \varphi) + (\boldsymbol{u}_h, \nabla\varphi)\} \leq \min_{\substack{\boldsymbol{v} \in \boldsymbol{H}(\mathrm{div},\Omega) \\ \nabla\cdot\boldsymbol{v}=f}} \|\boldsymbol{u}_h - \boldsymbol{v}\|. \tag{4.16}$$

3b) We now show the equality (4.15). The argument of the constrained minimization in (4.15) is

$$\boldsymbol{\sigma} := \arg \min_{\substack{\boldsymbol{v} \in \boldsymbol{H}(\mathrm{div},\Omega) \\ \nabla\cdot\boldsymbol{v}=f}} \|\boldsymbol{u}_h - \boldsymbol{v}\|$$

and is characterized by the Euler–Lagrange conditions as a function $\boldsymbol{\sigma} \in \boldsymbol{H}(\mathrm{div},\Omega)$ with $\nabla\cdot\boldsymbol{\sigma} = f$ such that

$$(\boldsymbol{\sigma}, \boldsymbol{v}) = (\boldsymbol{u}_h, \boldsymbol{v}) \qquad \forall \boldsymbol{v} \in \boldsymbol{H}(\mathrm{div},\Omega) \text{ with } \nabla\cdot\boldsymbol{v} = 0.$$

Imposing the constraint $\nabla\cdot\boldsymbol{\sigma} = f$ equivalently with test functions $q \in L^2(\Omega)$ and introducing the Lagrange multiplier $r \in L^2(\Omega)$, using that $\nabla\cdot\boldsymbol{H}(\mathrm{div},\Omega) = L^2(\Omega)$, the above problem is equivalent to finding $\boldsymbol{\sigma} \in \boldsymbol{H}(\mathrm{div},\Omega)$ and $r \in L^2(\Omega)$ such that

$$(\boldsymbol{\sigma}, \boldsymbol{v}) - (r, \nabla\cdot\boldsymbol{v}) = (\boldsymbol{u}_h, \boldsymbol{v}) \qquad \forall \boldsymbol{v} \in \boldsymbol{H}(\mathrm{div},\Omega), \tag{4.17a}$$

$$(\nabla\cdot\boldsymbol{\sigma}, q) = (f, q) \qquad \forall q \in L^2(\Omega). \tag{4.17b}$$

Now, (4.17a) implies by Definition 3.1 of the weak partial derivative that $r \in H_0^1(\Omega)$ with $\nabla r = -(\boldsymbol{\sigma} - \boldsymbol{u}_h)$. Consequently, by the Green theorem (3.3),

$$\min_{\substack{\boldsymbol{v} \in \boldsymbol{H}(\mathrm{div},\Omega) \\ \nabla\cdot\boldsymbol{v}=f}} \|\boldsymbol{u}_h - \boldsymbol{v}\| = \|\boldsymbol{u}_h - \boldsymbol{\sigma}\| = \|\nabla r\| = \max_{\substack{\varphi \in H_0^1(\Omega) \\ \|\nabla\varphi\|=1}} (\nabla r, \nabla\varphi)$$

$$= \max_{\substack{\varphi \in H_0^1(\Omega) \\ \|\nabla\varphi\|=1}} (-\boldsymbol{\sigma} + \boldsymbol{u}_h, \nabla\varphi) = \max_{\substack{\varphi \in H_0^1(\Omega) \\ \|\nabla\varphi\|=1}} \{(f, \varphi) + (\boldsymbol{u}_h, \nabla\varphi)\}, \tag{4.18}$$

which is (4.15). $\qquad\qquad\square$

4.7 Cell-centered finite volume discretizations on a simplicial mesh

Consider now a simplicial mesh $\mathcal{T}_h$ of the domain Ω as defined in Section 3.3. We additionally assume that 1) there exists a point $\boldsymbol{x}_K$ associated with each simplex $K \in \mathcal{T}_h$; 2) there exists a point $\boldsymbol{x}_{K,\sigma}$ associated with each face σ lying in the boundary $\partial\Omega$, $\sigma \in \mathcal{F}_K \cap \mathcal{F}_h^{\mathrm{ext}}$, $K \in \mathcal{T}_h$, lying in the interior of the face σ; 3) all the points $\boldsymbol{x}_K$ and $\boldsymbol{x}_{K,\sigma}$ are distinct; 4) the straight line connecting $\boldsymbol{x}_K$ and $\boldsymbol{x}_L$ for two neighboring elements K and L is orthogonal to their common face $\sigma_{K,L} = \partial K \cap \partial L \in \mathcal{F}_h^{\mathrm{int}}$; 5) the straight line connecting $\boldsymbol{x}_K$ and $\boldsymbol{x}_{K,\sigma}$ is orthogonal to σ for $\sigma \in \mathcal{F}_K \cap \mathcal{F}_h^{\mathrm{ext}}$. The points $\boldsymbol{x}_K$ typically, but not necessarily, lie in the interior of the element $K \in \mathcal{T}_h$. In particular, Delaunay meshes satisfy this requirement of "admissibility" in the sense of [126, Definition 9.1]. An illustration is provided in Figure 10.

The cell-centered finite volume scheme for the Poisson equation (4.1) on the simplicial mesh $\mathcal{T}_h$ reads: find the real values p_K, $K \in \mathcal{T}_h$, the approximations to the mean values of p in the mesh elements K, such that

$$\boxed{\sum_{\sigma \in \mathcal{F}_K} U_{K,\sigma} = (f, 1)_K \qquad \forall K \in \mathcal{T}_h,} \tag{4.19a}$$
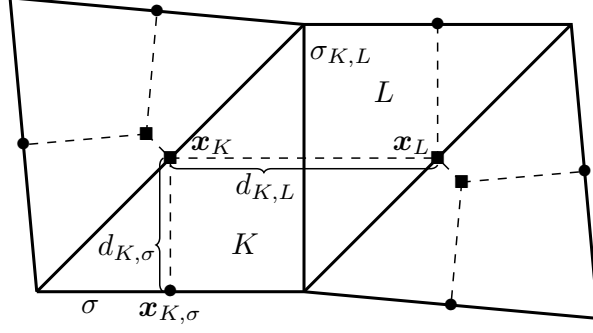
Figure 10: Admissible triangular mesh $\mathcal{T}_h$ and notation for the cell-centered finite volume scheme
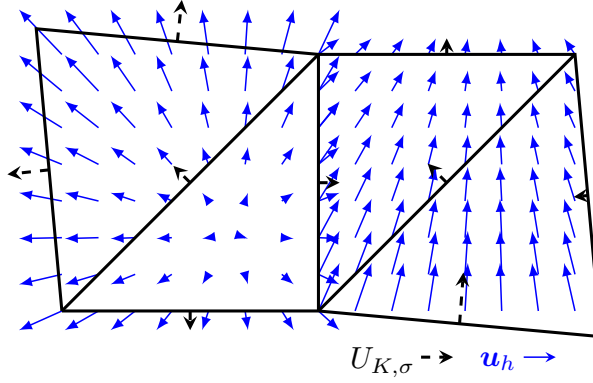
Figure 11: Finite volume face fluxes $U_{K,\sigma}$ (dashed black arrows) and flux reconstruction $\boldsymbol{u}_h$ of Definition 4.14 (blue arrows)

where $U_{K,\sigma} \in \mathbb{R}$ for each face $\sigma \in \mathcal{F}_K$ approximates the normal (out)flux $\langle \boldsymbol{u} \cdot \boldsymbol{n}_K, 1 \rangle_\sigma$ from K over the face σ (recall the notation (3.12)) by

$$U_{K,\sigma} := -\frac{|\sigma_{K,L}|}{d_{K,L}}(p_L - p_K) \qquad \sigma = \sigma_{K,L} \in \mathcal{F}_h^{\mathrm{int}}, \tag{4.19b}$$

$$U_{K,\sigma} := -\frac{|\sigma|}{d_{K,\sigma}}(0 - p_K) \qquad \sigma \in \mathcal{F}_K \cap \mathcal{F}_h^{\mathrm{ext}}. \tag{4.19c}$$

Recall that $|\sigma|$ is the $(d-1)$-dimensional measure of the face σ. Moreover, $d_{K,L} = |\boldsymbol{x}_K - \boldsymbol{x}_L|$ is the Euclidean distance of the points $\boldsymbol{x}_K$ and $\boldsymbol{x}_L$ for an interior face $\sigma = \sigma_{K,L} \in \mathcal{F}_h^{\mathrm{int}}$, unless the positions of $\boldsymbol{x}_K$ and $\boldsymbol{x}_L$ are interchanged with respect to the positions of K and L, where we take $d_{K,L} = -|\boldsymbol{x}_K - \boldsymbol{x}_L|$. Since we suppose $\boldsymbol{x}_K$ and $\boldsymbol{x}_L$ distinct, $d_{K,L} \neq 0$. Similarly, $d_{K,\sigma} = |\boldsymbol{x}_K - \boldsymbol{x}_{K,\sigma}|$ is the Euclidean distance of the points $\boldsymbol{x}_K$ and $\boldsymbol{x}_{K,\sigma}$ for a boundary face $\sigma \in \mathcal{F}_K \cap \mathcal{F}_h^{\mathrm{ext}}$, unless $\boldsymbol{x}_K$ lies outside of K past σ, where we take $d_{K,\sigma} = -|\boldsymbol{x}_K - \boldsymbol{x}_{K,\sigma}|$. Figure 10 provides an illustration.

Remark 4.13 (Flux conservation and equilibrium in finite volumes on simplicial meshes). *From* (4.19b), *there holds*

$$U_{K,\sigma} = -U_{L,\sigma} \qquad \forall \sigma = \sigma_{K,L} \in \mathcal{F}_h^{\mathrm{int}}. \tag{4.20}$$

This is the flux conservation in finite volumes: what flows from the element K out through the face $\sigma_{K,L}$ flows in the neighbor element L through the same face $\sigma_{K,L}$, cf. Lemma 3.11 and definition (3.15). The requirement (4.19a) then stipulates the equilibrium of these face fluxes with the load f, mimicking $\nabla \cdot \boldsymbol{u} = f$ from Proposition 4.1.

4.8 Flux reconstruction by lifting the face normal fluxes

Following [127], we now lift the normal fluxes $U_{K,\sigma}$ from (4.19b)–(4.19c) to form a vector-valued approximate flux field $\boldsymbol{u}_h$ that is $\boldsymbol{H}(\mathrm{div}, \Omega)$-conforming and in equilibrium with the load f. More precisely, we construct $\boldsymbol{u}_h$ in the piecewise polynomial Raviart–Thomas space $\mathcal{RT}_0(\mathcal{T}_h) \cap \boldsymbol{H}(\mathrm{div}, \Omega)$ from (3.15).

Recalling the definition (3.10) of the space $\boldsymbol{\mathcal{RT}}_0(K)$ for a simplex K and the notation from Figure 6, we start with:

Definition 4.14 (Flux reconstruction)**.** *Let the cell-centered finite volume discretization on a simplicial mesh be given by (4.19). Then, on each simplex $K \in \mathcal{T}_h$, define*

$$\boldsymbol{u}_h|_K \in \boldsymbol{\mathcal{RT}}_0(K), \quad \langle \boldsymbol{u}_h \cdot \boldsymbol{n}_K, 1 \rangle_\sigma = U_{K,\sigma} \qquad \forall \sigma \in \mathcal{F}_K, \tag{4.21}$$

or, equivalently, using the Raviart–Thomas basis functions $\boldsymbol{v}_\sigma$ from (3.13),

$$(\boldsymbol{u}_h|_K)(\boldsymbol{x}) := \sum_{\sigma \in \mathcal{F}_K} U_{K,\sigma} \boldsymbol{v}_\sigma(\boldsymbol{x}) = \sum_{\sigma \in \mathcal{F}_K} U_{K,\sigma} \frac{1}{d|K|}(\boldsymbol{x} - \boldsymbol{a}_{K,\sigma}), \qquad \boldsymbol{x} \in K. \tag{4.22}$$

An illustration of the flux reconstruction $\boldsymbol{u}_h$ from the finite volume face fluxes $U_{K,\sigma}$ is provided in Figure 11. Crucially, we have:

Lemma 4.15 (Flux reconstruction)**.** *Let $\boldsymbol{u}_h$ be given by Definition 4.14. Then*

$$\boldsymbol{u}_h \in \boldsymbol{\mathcal{RT}}_0(\mathcal{T}_h) \cap \boldsymbol{H}(\mathrm{div}, \Omega), \qquad (\nabla \cdot \boldsymbol{u}_h)|_K = \frac{(f, 1)_K}{|K|} \qquad \forall K \in \mathcal{T}_h. \tag{4.23}$$

Proof. From the finite volume conservation (4.20) and definition (4.21), we have the normal trace continuity needed in (3.15); this yields the first property in (4.23). As for the second property in (4.23), we write, using the Green theorem, definition (4.21), and the finite volume scheme (4.19a)

$$(\nabla \cdot \boldsymbol{u}_h, 1)_K = \sum_{\sigma \in \mathcal{F}_K} \langle \boldsymbol{u}_h \cdot \boldsymbol{n}_K, 1 \rangle_\sigma \overset{(4.21)}{=} \sum_{\sigma \in \mathcal{F}_K} U_{K,\sigma} \overset{(4.19a)}{=} (f, 1)_K.$$

The conclusion follows since $\nabla \cdot \boldsymbol{u}_h$ is constant by virtue of (3.11). $\qquad \square$

4.9 Potential reconstruction by elementwise postprocessing and averaging

Supposing for the moment that f is piecewise constant, so that $\boldsymbol{u}_h \in \boldsymbol{H}(\mathrm{div}, \Omega)$ with $\nabla \cdot \boldsymbol{u}_h = f$ from (4.23), we will be in position to use Corollary 4.6 to estimate the error between the finite volume flux reconstruction $\boldsymbol{u}_h$ of Definition 4.14 and the exact flux $\boldsymbol{u} = -\nabla p$. Inequality (4.8) tells us that we will need a suitable scalar-valued object in the primal space $H_0^1(\Omega)$ whose negative gradient is as close as possible to $\boldsymbol{u}_h$. Since the piecewise constant function given by the finite volume values p_K on each $K \in \mathcal{T}_h$ cannot be used directly, as it does not lie in the space $H_0^1(\Omega)$, we proceed in two steps.

Following [247, Section 4.1] and [249, Section 3.2], we first create a scalar-valued piecewise polynomial whose negative gradient equals $\boldsymbol{u}_h$ and whose mean values are given by p_K on each mesh element $K \in \mathcal{T}_h$. Recalling the piecewise polynomial space $\mathcal{P}_k(\mathcal{T}_h)$ from (3.8), we define:

Definition 4.16 (Potential postprocessing)**.** *Let p_K, $K \in \mathcal{T}_h$, be the finite volume approximate solution of (4.19). Let $\boldsymbol{u}_h$ be given by Definition 4.14. Define the potential postprocessing $\tilde{p}_h$ as a piecewise quadratic polynomial on the simplicial mesh $\mathcal{T}_h$, $\tilde{p}_h \in \mathcal{P}_2(\mathcal{T}_h)$, given by*

$$-\nabla \tilde{p}_h|_K = \boldsymbol{u}_h|_K, \qquad \frac{(\tilde{p}_h, 1)_K}{|K|} = p_K \qquad \forall K \in \mathcal{T}_h. \tag{4.24}$$

It is easy to verify that any object from $\boldsymbol{\mathcal{RT}}_0(K)$ for a simplex K is a gradient of some polynomial which is at most second-order, since fields in $\boldsymbol{\mathcal{RT}}_0(K)$ are curl-free and at most affine. Thus (4.24) is well-posed.

Unfortunately, in general $\tilde{p}_h$ from Definition 4.16 does not lie in $H_0^1(\Omega)$; if this was the case, then, actually, by Proposition 4.4 or Theorem 4.9, $\tilde{p}_h$ would be the weak solution p. An illustration of a prototypical situation is provided in Figure 12, left. For this reason, we continue with a second step, in the spirit of averaging operators in Durán and Padra [112], Achdou *et al.* [3], Karakashian and Pascal [157], Ainsworth [12], or Burman and Ern [68]. Recall the Lagrange spaces and nodes of Section 3.5.2 and Figure 4.
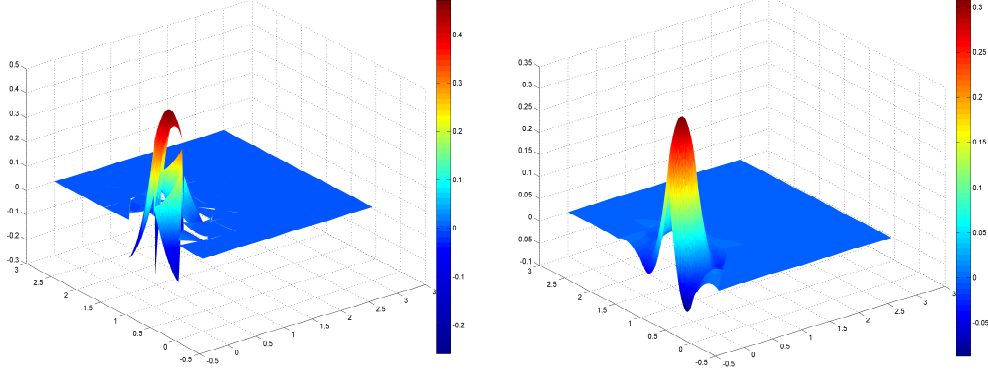
Figure 12: Potential postprocessing $\tilde{p}_h \in \mathcal{P}_2(\mathcal{T}_h)$ (left) and potential reconstruction by averaging $\zeta_h \in \mathcal{P}_2(\mathcal{T}_h) \cap H^1(\Omega)$ (right)

Definition 4.17 (Potential reconstruction by averaging). *Let $\tilde{p}_h \in \mathcal{P}_2(\mathcal{T}_h)$ be given by Definition 4.16. We call a potential reconstruction by averaging a piecewise quadratic and continuous polynomial $\zeta_h \in \mathcal{P}_2(\mathcal{T}_h) \cap H_0^1(\Omega)$ that is prescribed by its values in the Lagrange nodes of the space $\mathcal{P}_2(\mathcal{T}_h) \cap H_0^1(\Omega)$ by*

$$\zeta_h(\boldsymbol{x}) := \frac{1}{|\mathcal{T}_{\boldsymbol{x}}|} \sum_{K \in \mathcal{T}_{\boldsymbol{x}}} \tilde{p}_h|_K(\boldsymbol{x}) \qquad \boldsymbol{x} \text{ is a Lagrange node of } \mathcal{P}_2(\mathcal{T}_h) \cap H_0^1(\Omega) \text{ included in } \Omega, \qquad (4.25a)$$

$$\zeta_h(\boldsymbol{x}) := 0 \qquad \boldsymbol{x} \text{ is a Lagrange node of } \mathcal{P}_2(\mathcal{T}_h) \cap H_0^1(\Omega) \text{ included in } \partial\Omega, \qquad (4.25b)$$

where $\mathcal{T}_{\boldsymbol{x}}$ denotes the set of elements of the mesh $\mathcal{T}_h$ that contain the point $\boldsymbol{x}$ and $|\mathcal{T}_{\boldsymbol{x}}|$ is the cardinality (number of elements) of this set.

We simply average all the (typically) different values that ζ_h takes in the Lagrange nodes. An illustration is given in Figure 12, right (where the 0 values at the boundary $\partial\Omega$ are not imposed).

4.10 A guaranteed a posteriori error estimate

With the above developments, we are now in a position to present a guaranteed a posteriori error estimate for the cell-centered finite volume discretization (4.19) of the Poisson equation (4.1) on a simplicial mesh $\mathcal{T}_h$. We follow [249, Theorem 4.1].

Theorem 4.18 (A guaranteed a posteriori error estimate).

Let $p \in H_0^1(\Omega)$ be the exact potential given by (4.2) and let $\boldsymbol{u}$ be the exact flux given by (4.3). Let the cell-centered finite volume discretization be given by (4.19) and let the approximate flux $\boldsymbol{u}_h \in \mathcal{RT}_0(\mathcal{T}_h) \cap \boldsymbol{H}(\mathrm{div}, \Omega)$ be constructed following Definition 4.14. Let the potential postprocessing $\tilde{p}_h \in \mathcal{P}_2(\mathcal{T}_h)$ be given by Definition 4.16 and the potential reconstruction by averaging $\zeta_h \in \mathcal{P}_2(\mathcal{T}_h) \cap H_0^1(\Omega)$ by Definition 4.17. Then there holds

$$\|\boldsymbol{u} - \boldsymbol{u}_h\| \le \eta := \left\{ \sum_{K \in \mathcal{T}_h} \eta_K^2 \right\}^{1/2}, \qquad (4.26a)$$

where

$$\eta_K^2 := \|\boldsymbol{u}_h + \nabla\zeta_h\|_K^2 + \frac{h_K^2}{\pi^2} \left\| f - \frac{(f,1)_K}{|K|} \right\|_K^2. \qquad (4.26b)$$

Proof. If f is piecewise constant, then $\boldsymbol{u}_h \in \boldsymbol{H}(\mathrm{div}, \Omega)$ with $\nabla \cdot \boldsymbol{u}_h = f$ from (4.23). In this case, we merely employ inequality (4.8) from Corollary 4.6.

In the general case $f \in L^2(\Omega)$, we rather use equality (4.9) from Theorem 4.9. The second term on the right-hand side of (4.9) gives rise to $\|\boldsymbol{u}_h + \nabla\zeta_h\|^2$. As for the first one, let $\varphi \in H_0^1(\Omega)$ with $\|\nabla\varphi\| = 1$ be fixed. By the Green theorem (3.3), the flux reconstruction property (4.23), the Cauchy–Schwarz

inequality, and the Poincaré inequality

$$\left\| \varphi - \frac{(\varphi,1)_K}{|K|} \right\|_K \leq \frac{h_K}{\pi} \|\nabla\varphi\|_K \qquad \varphi \in H^1(K), \tag{4.27}$$

cf. [212, 37], we see

$$
(f,\varphi) + (\boldsymbol{u}_h, \nabla\varphi) \stackrel{(3.3)}{=} (f - \nabla\cdot\boldsymbol{u}_h, \varphi) = \sum_{K\in\mathcal{T}_h} (f - \nabla\cdot\boldsymbol{u}_h, \varphi)_K \stackrel{(4.23)}{=} \sum_{K\in\mathcal{T}_h} \left(f - \frac{(f,1)_K}{|K|}, \varphi \right)_K
$$

$$
= \sum_{K\in\mathcal{T}_h} \left(f - \frac{(f,1)_K}{|K|}, \varphi - \frac{(\varphi,1)_K}{|K|} \right)_K \stackrel{(4.27)}{\leq} \sum_{K\in\mathcal{T}_h} \left(\left\| f - \frac{(f,1)_K}{|K|} \right\|_K \frac{h_K}{\pi} \|\nabla\varphi\|_K \right)
$$

$$
\leq \left\{ \sum_{K\in\mathcal{T}_h} \left(\frac{h_K^2}{\pi^2} \left\| f - \frac{(f,1)_K}{|K|} \right\|_K^2 \right) \right\}^{1/2} \underbrace{\|\nabla\varphi\|}_{=1} = \left\{ \sum_{K\in\mathcal{T}_h} \left(\frac{h_K^2}{\pi^2} \left\| f - \frac{(f,1)_K}{|K|} \right\|_K^2 \right) \right\}^{1/2}.
$$

Thus (4.26) follows. $\qquad\square$

Remark 4.19 (Local efficiency). *Local efficiency of the estimate of* (4.26) *as per Section* 2.5.2 *holds true, see [249, Theorem 4.2].*

Remark 4.20 (Rectangular meshes). *All the above results are in [249] also presented on rectangular meshes.*

4.11   Numerical experiments

In this Section, we illustrate the application of the a posteriori error estimates of the form of Theorem 4.18 to the finite volume method on two test cases in two space dimensions.

4.11.1   Regular solution

In this test, we consider a model problem from [192] of form (4.1) with $\Omega = (0,1) \times (0,1)$ and the load term f given such that the (regular exact solution reads

$$p(x,y) = \frac{10^2}{4} x(1-x)y(1-y) \exp\left(-100 \left((x-0.75)^2 + (y-0.75)^2 \right) \right). \tag{4.28}$$

This solution is regular but contains a localized exponential peak. The numerical tests are performed on a sequence of uniformly refined rectangular meshes $\mathcal{T}_0, \mathcal{T}_1, \ldots, \mathcal{T}_J$, $J = 3$, cf. Remarks 3.12 and 4.20. Figure 13 illustrates the approximate solution given by the values p_K, $K \in \mathcal{T}_h$, the elementwise errors $\|\boldsymbol{u} - \boldsymbol{u}_h\|_K$, and the corresponding a posteriori error estimators η_K arising from Theorem 4.18 on mesh $\mathcal{T}_3$. We observe that the error is localized in a circular zone around the peak and that the actual and predicted error distributions match very closely.

In the left part of Figure 14, we plot the relative exact errors $\|\boldsymbol{u} - \boldsymbol{u}_h\|/\|\boldsymbol{u}_h\|$ on each of the meshes $\mathcal{T}_0$–$\mathcal{T}_3$ together with their relative a posteriori error estimates $\eta/\|\boldsymbol{u}_h\|$ from Theorem 4.18, with respect to the number of mesh elements/number of unknowns. We indeed observe that the estimators give a guaranteed upper bound on the errors. In the right part of Figure 14 we then plot the effectivity indices defined as per (2.22) by the ratio of the estimator to the error,

$$I_{\text{eff}} := \frac{\eta}{\|\boldsymbol{u} - \boldsymbol{u}_h\|}. \tag{4.29}$$

We observe that they are remarkably close to one and tend to the optimal value of one with mesh refinement, showing asymptotic exactness as per Section 2.5.4.

4.11.2   Singular solution

We consider a benchmark test case defined on the L-shaped domain

$$\Omega = (-1,1) \times (-1,1) \setminus (-1,0] \times (-1,0]$$

Figure 13: [Section 4.11.1, mesh $\mathcal{T}_3$] The approximate solution given by the values p_K, $K \in \mathcal{T}_h$, (*left*), the exact errors $\|\boldsymbol{u} - \boldsymbol{u}_h\|_K$ (*middle*), and the error estimators η_K from Theorem 4.18 (*right*).
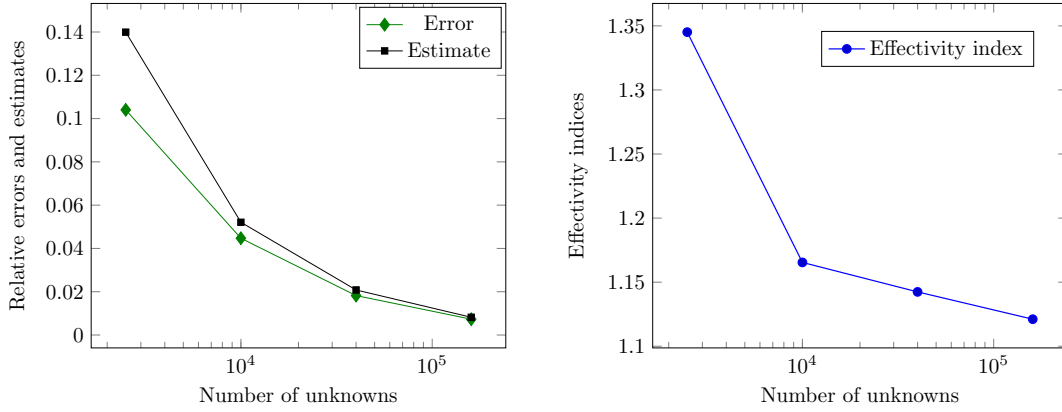
Figure 14: [Section 4.11.1, uniformly refined meshes $\mathcal{T}_0$–$\mathcal{T}_3$] Relative errors $\|\boldsymbol{u} - \boldsymbol{u}_h\|/\|\boldsymbol{u}_h\|$ and relative estimates $\eta/\|\boldsymbol{u}_h\|$ from Theorem 4.18 (*left*), effectivity indices I_{eff} from (4.29) (*right*)

with the exact solution

$$p(r,\theta) = r^{\frac{2}{3}} \sin\left(\frac{2}{3}\theta + \frac{3}{2}\pi\right),$$

where (r,θ) are the polar coordinates. The Dirichlet boundary condition is partly inhomogeneous here and given by the value of the exact solution on $\partial\Omega$. We illustrate in Figure 15 the approximate solution given by the values p_K, $K \in \mathcal{T}_h$, the elementwise errors $\|\boldsymbol{u} - \boldsymbol{u}_h\|_K$, and the corresponding a posteriori error estimators η_K from Theorem 4.18. We observe that the a posteriori error estimators η_K detect perfectly the error related in the vicinity of the singularity at $(0,0)$.

The convergence histories on the sequence of the uniformly-refined meshes $\mathcal{T}_0 - \mathcal{T}_2$ are depicted in the left part of Figure 16. The right part of Figure 16 then displays the corresponding effectivity indices from (4.29). We see that, for this test case, the effectivity indices are still close to the optimal value of one, even with singular solution, though no asymptotic exactness is observed here.

4.12   Bibliographic resources

One of the first a posteriori error analyses for finite volume methods was performed by Angermann [20]. Later, Achdou *et al.* [3], Afif *et al.* [6], Lazarov and Tomov [178], Carstensen *et al.* [73], and Amaziane *et al.* [18] also considered the so-called vertex-centered finite volume discretizations, closely related to lowest-order finite elements. Cell-centered finite volumes of the form (4.19a) were then addressed in Agouzal and Oudin [8], Achdou *et al.* [3], Nicaise [196], Vohralík [249], and Erath and Praetorius [116].
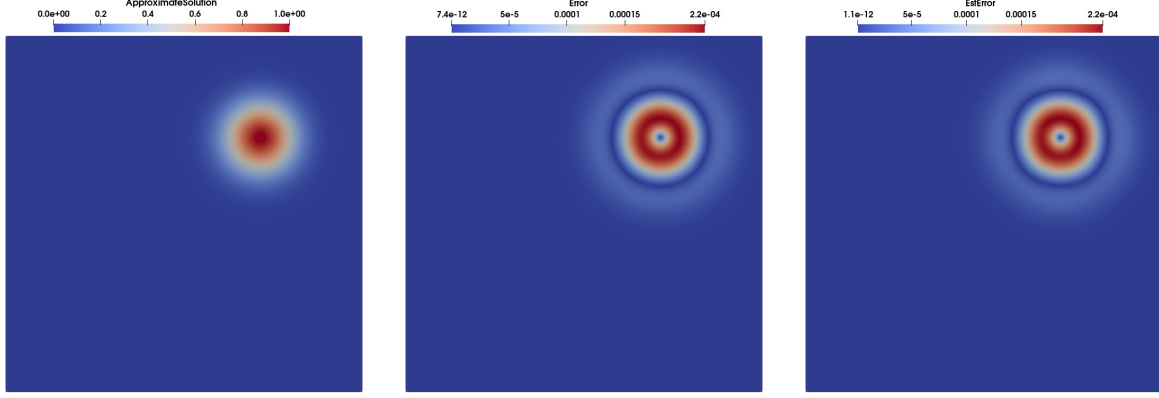
28
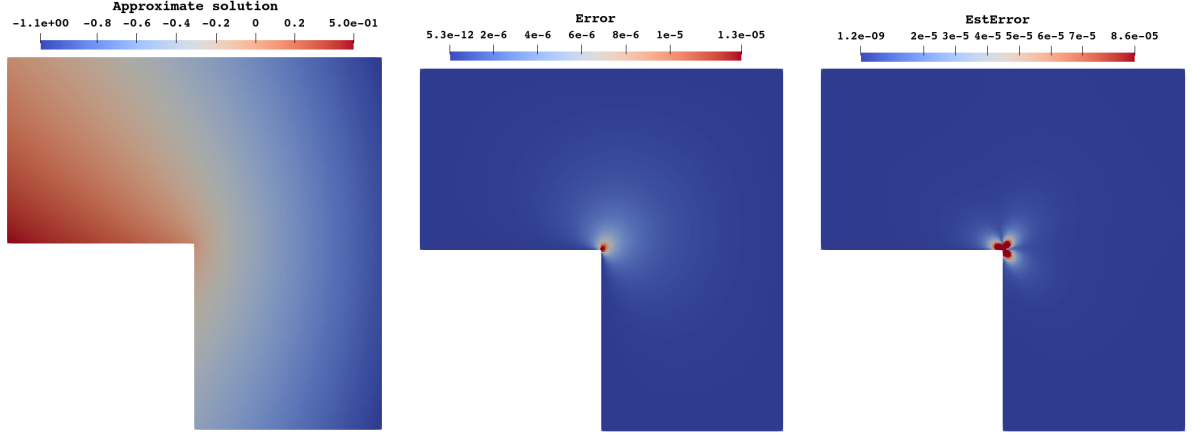
Figure 15: [Section 4.11.2, mesh $\mathcal{T}_2$] The approximate solution given by the values p_K, $K \in \mathcal{T}_h$, (*left*), the exact errors $\|\boldsymbol{u} - \boldsymbol{u}_h\|_K$ (*middle*), and the error estimators η_K from Theorem 4.18 (*right*).
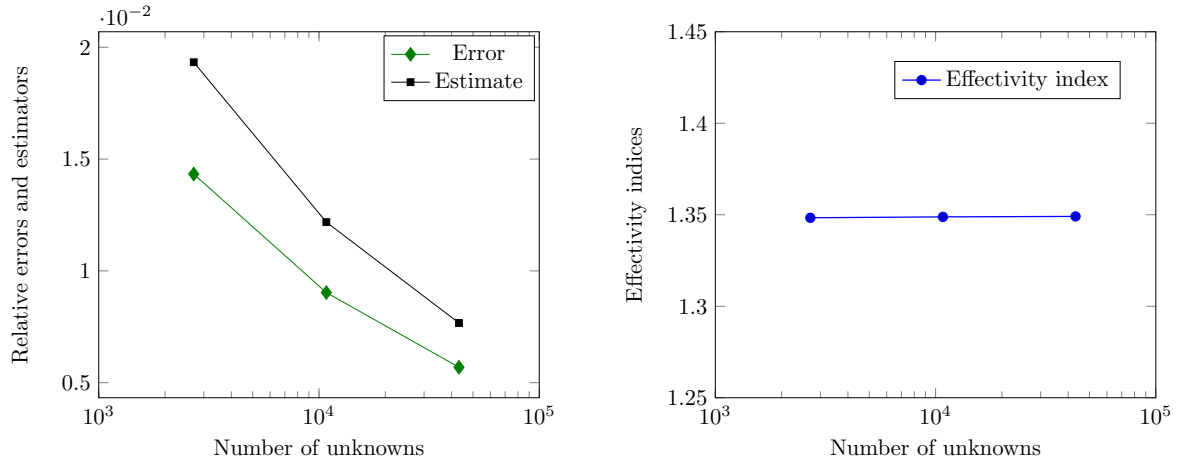
Figure 16: [Section 4.11.2, uniformly refined meshes $\mathcal{T}_0$–$\mathcal{T}_2$] Relative errors $\|\boldsymbol{u} - \boldsymbol{u}_h\|/\|\boldsymbol{u}_h\|$ and relative estimates $\eta/\|\boldsymbol{u}_h\|$ from Theorem 4.18 (*left*), effectivity indices I_{eff} from (4.29) (*right*)

5 Steady linear pure diffusion problems. Polytopal meshes and inexpensive implementation and evaluation

We describe here first a model steady linear problem, the Darcy flow problem. This extends the Poisson equation (4.1) from Section 4 and forms the basis for complex multiphase compositional porous media flows as that of Section 7. We also focus on extending the finite volume discretizations on simplicial meshes from Section 4.7 to general polytopal meshes. Finally, we investigate how the guaranteed a posteriori error estimates of the form of Theorem 4.18 can be implemented and evaluated in the most inexpensive way, and this on polytopal meshes. In particular, the flux and potential reconstructions of Sections 4.8–4.9 will only be performed here virtually, avoiding their factual construction. Instead, the a posteriori error estimates will only need *one flux value per polytopal face* and *one potential value per vertex of a (virtual) simplicial submesh*. The former are directly available from the numerical scheme (these are the generalizations of the face normal fluxes $U_{K,\sigma}$ in (4.19)) and the latter will be trivially obtained by some postprocessing and averaging of the elementwise potential values generalizing p_K in (4.19).

5.1 Singlephase steady linear Darcy flow

Consider the problem of finding $p : \Omega \to \mathbb{R}$ such that

$$-\nabla{\cdot}(\underline{\boldsymbol{K}}\nabla p) = f \qquad \text{in } \Omega, \tag{5.1a}$$
$$p = 0 \qquad \text{on } \partial\Omega. \tag{5.1b}$$

Here f is a source term and $\underline{\boldsymbol{K}}$ is a symmetric, positive definite, and bounded diffusion tensor with values in $\mathbb{R}^{d\times d}$; we suppose for simplicity here that both f and $\underline{\boldsymbol{K}}$ are piecewise constant with respect to the polytopal mesh $\mathcal{T}_H$.

5.2 Weak solution and its properties

The Sobolev space $H_0^1(\Omega)$ from Definition 3.4 still gives a proper mathematical setting to define a unique solution p of problem (5.1), which is $p \in H_0^1(\Omega)$ such that

$$(\underline{\boldsymbol{K}}\nabla p, \nabla v) = (f, v) \qquad \forall v \in H_0^1(\Omega). \tag{5.2}$$

From the pressure head p, we can define the Darcy velocity

$$\boldsymbol{u} := -\underline{\boldsymbol{K}}\nabla p; \tag{5.3}$$

recall that we refer to p as *potential* and to $\boldsymbol{u}$ as *flux*. As in Proposition 4.1, it follows from (5.2) that

$$\boldsymbol{u} \in \boldsymbol{H}(\mathrm{div}, \Omega), \quad \nabla{\cdot}\boldsymbol{u} = f. \tag{5.4}$$

Thus, just like in Remark 4.3, the scalar-valued pressure head p is trace continuous, cf. Figure 1, whereas the vector-valued Darcy velocity $\boldsymbol{u}$ is normal-trace continuous, cf. Figure 2. Moreover, $\boldsymbol{u}$ is locally mass conservative, in equilibrium with the load, satisfying $\nabla{\cdot}\boldsymbol{u} = f$.

5.3 Spatial discretization error

Applying the abstract reflections of Section 2.4.2, we will evaluate the spatial discretization error as

$$\|\boldsymbol{u} - \boldsymbol{u}_h\|_{\underline{\boldsymbol{K}}^{-\frac{1}{2}}} \tag{5.5}$$

for an approximate Darcy velocity $\boldsymbol{u}_h$. Here, the energy norm on $\omega \subset \Omega$, for $\boldsymbol{v} \in \left[L^2(\omega)\right]^d$, writes as

$$\|\boldsymbol{v}\|_{\underline{\boldsymbol{K}}^{-\frac{1}{2}},\omega} := \|\underline{\boldsymbol{K}}^{-1/2}\boldsymbol{v}\|_{L^2(\omega)} = \left\{\int_\omega \left|\underline{\boldsymbol{K}}^{-1/2}(\boldsymbol{x})\boldsymbol{v}(\boldsymbol{x})\right|^2 \mathrm{d}\boldsymbol{x}\right\}^{\frac{1}{2}}; \tag{5.6}$$

when $\omega = \Omega$, we as usual drop the subscript ω.

5.4 Generic discretizations on a polytopal mesh

Let $\mathcal{T}_H$ be a polytopal mesh of Ω in the sense of Section 3.6. Define the load vector $\mathsf{F} := \{\mathsf{F}_K\}_{K\in\mathcal{T}_H} \in \mathbb{R}^{|\mathcal{T}_H|}$ by

$$\mathsf{F}_K := (f, 1)_K \qquad \forall K \in \mathcal{T}_H. \tag{5.7}$$

We consider any numerical discretization of (5.1) that can be written under the following abstract lowest-order locally conservative form:

Assumption 5.1 (Locally conservative discretization on a polytopal mesh)**.** *Find the algebraic vector* $\mathsf{P} := \{\mathsf{P}_K\}_{K\in\mathcal{T}_H} \in \mathbb{R}^{|\mathcal{T}_H|}$ *such that the flux balance*

$$\sum_{\sigma\in\mathcal{F}_K} \mathsf{U}_\sigma \boldsymbol{n}_K{\cdot}\boldsymbol{n}_\sigma = \mathsf{F}_K \quad \forall K \in \mathcal{T}_H \tag{5.8}$$

is satisfied. Here $\mathsf{U} := \{\mathsf{U}_\sigma\}_{\sigma\in\mathcal{F}_H} \in \mathbb{R}^{|\mathcal{F}_H|}$*, and* $\mathsf{U}_\sigma \in \mathbb{R}$ *for each face* $\sigma \in \mathcal{F}_H$ *approximates the normal flux* $\langle\boldsymbol{u}{\cdot}\boldsymbol{n}_\sigma, 1\rangle_\sigma$ *over the face* σ *and depends linearly on* P. *The unknowns* $\mathsf{P}_K \in \mathbb{R}$ *for each element* $K \in \mathcal{T}_H$ *approximate the potential* p *in the element* K.

Assumption 5.1 is very generic (allows for any lowest-order locally conservative method). Note in particular that for the a posteriori error analysis performed below, we *do not need to know* how the flux unknowns U depend on the pressure head unknowns P. Recall that on admissible simplicial meshes, the easiest example for this relation is the two-point formula (4.19b)–(4.19c).

Many lowest-order locally conservative methods take a form comprised in Assumption 5.1 but more specific. We consider them separately under the following assumption, in the spirit of mimetic finite differences [64, Theorem 5.1], mixed finite elements [254, Theorems 7.2 and 7.3], and the unifying frameworks [110, 111]:

Assumption 5.2 (Locally conservative saddle-point discretization on a polytopal mesh)**.** *Find the algebraic vectors* $\mathsf{U} := \{\mathsf{U}_\sigma\}_{\sigma \in \mathcal{F}_H} \in \mathbb{R}^{|\mathcal{F}_H|}$ *and* $\mathsf{P} := \{\mathsf{P}_K\}_{K \in \mathcal{T}_H} \in \mathbb{R}^{|\mathcal{T}_H|}$ *such that*

$$\begin{pmatrix} \mathbb{A} & \mathbb{B}^{\mathrm{t}} \\ \mathbb{B} & 0 \end{pmatrix} \begin{pmatrix} \mathsf{U} \\ \mathsf{P} \end{pmatrix} = \begin{pmatrix} 0 \\ -\mathsf{F} \end{pmatrix}, \tag{5.9}$$

where

1) *the unknown* $\mathsf{U}_\sigma \in \mathbb{R}$ *for each face* $\sigma \in \mathcal{F}_H$ *approximates the normal flux* $\langle \boldsymbol{u}\cdot\boldsymbol{n}_\sigma, 1\rangle_\sigma$ *over the face* σ *and the unknown* $\mathsf{P}_K \in \mathbb{R}$ *for each element* $K \in \mathcal{T}_H$ *approximates the potential* p *in the element* K*;*

2) *the matrix* $\mathbb{B} \in \mathbb{R}^{|\mathcal{T}_H| \times |\mathcal{F}_H|}$ *has a full rank;*

3) *for each polytopal cell* $K \in \mathcal{T}_H$ *and face* $\sigma \in \mathcal{F}_H$, $\mathbb{B}_{K,\sigma} = -\boldsymbol{n}_K\cdot\boldsymbol{n}_\sigma$ *if* σ *is a face of* K, $\sigma \in \mathcal{F}_K$, *and* $\mathbb{B}_{K,\sigma} = 0$ *otherwise;*

4) *the matrix* $\mathbb{A} \in \mathbb{R}^{|\mathcal{F}_H| \times |\mathcal{F}_H|}$ *is invertible;*

5) *for faces* $\sigma, \sigma' \in \mathcal{F}_H$, $\mathbb{A}_{\sigma,\sigma'} = 0$ *if* σ *and* σ' *are not faces of the same element* $K \in \mathcal{T}_H$, $\{\sigma, \sigma'\} \not\subset \mathcal{F}_K$ *for some* $K \in \mathcal{T}_H$, *and*

$$\mathbb{A}_{\sigma,\sigma'} = \sum_{K \in \mathcal{T}_H, \{\sigma,\sigma'\} \subset \mathcal{F}_K} \boldsymbol{n}_K\cdot\boldsymbol{n}_\sigma \, \boldsymbol{n}_K\cdot\boldsymbol{n}_{\sigma'} (\widehat{\mathbb{A}}_K)_{\sigma,\sigma'} \tag{5.10}$$

otherwise, where $\widehat{\mathbb{A}}_K \in \mathbb{R}^{|\mathcal{F}_K| \times |\mathcal{F}_K|}$ *are symmetric and positive definite (the "element matrices" of the given method for each* $K \in \mathcal{T}_H$*);*

6) *on each polytopal cell* $K \in \mathcal{T}_H$, *there exists a lifting* $\tilde{\boldsymbol{u}}_h|_K \in \boldsymbol{H}(\mathrm{div}, K)$ *of the face normal fluxes* $\mathsf{U}_K^{\mathrm{ext}} = \{\mathsf{U}_\sigma\}_{\sigma \in \mathcal{F}_K}$ *to the interior of* K *such that*

$$\langle \tilde{\boldsymbol{u}}_h|_K\cdot\boldsymbol{n}_\sigma, 1\rangle_\sigma = \mathsf{U}_\sigma \quad \forall \sigma \in \mathcal{F}_K, \tag{5.11a}$$

$$\nabla\cdot\tilde{\boldsymbol{u}}_h|_K = f|_K, \tag{5.11b}$$

$$\|\tilde{\boldsymbol{u}}_h\|^2_{\underline{\boldsymbol{K}}^{-\frac{1}{2}}, K} = (\mathsf{U}_K^{\mathrm{ext}})^{\mathrm{t}}\widehat{\mathbb{A}}_K\mathsf{U}_K^{\mathrm{ext}}. \tag{5.11c}$$

5.5 Face normal fluxes

In Theorem 5.18 below, we will need a vector of real values with one value U_σ per face $\sigma \in \mathcal{F}_{H,h}$ (approximate face normal fluxes):

Definition 5.3 (Face normal fluxes)**.** *Let a discretization scheme of the form of Assumption 5.1 or 5.2 be given, leading to the face normal fluxes* $\mathsf{U} := \{\mathsf{U}_\sigma\}_{\sigma \in \mathcal{F}_H} \in \mathbb{R}^{|\mathcal{F}_H|}$. *For each polytopal face* $\sigma \in \mathcal{F}_H$ *and each simplicial subface* $\sigma' \in \mathcal{F}_{H,h}$ *as per Section 3.6, define the face normal flux*

$$\mathsf{U}_{\sigma'} := \frac{\mathsf{U}_\sigma}{|\sigma|}|\sigma'| \quad \forall \sigma' \in \mathcal{F}_{H,h}, \, \sigma' \subset \sigma, \, \sigma \in \mathcal{F}_H \tag{5.12}$$

and collect them in the element vectors $\mathsf{U}_K^{\mathrm{ext}} \in \mathbb{R}^{|\mathcal{F}_{K,h}^{\mathrm{ext}}|}$

$$\mathsf{U}_K^{\mathrm{ext}} := \{\mathsf{U}_\sigma\}_{\sigma \in \mathcal{F}_{K,h}^{\mathrm{ext}}} \tag{5.13}$$

for each polytopal cell $K \in \mathcal{T}_H$.

Remark that if the faces of the simplicial mesh $\mathcal{T}_h$ do not subdivide the faces of the polytopal mesh $\mathcal{T}_H$, then $\mathcal{F}_{H,h} = \mathcal{F}_H$ and consequently $\mathsf{U}_{\sigma'} = \mathsf{U}_\sigma$ for all $\sigma \in \mathcal{F}_K$ and $\mathsf{U}_K^{\mathrm{ext}} = \{\mathsf{U}_\sigma\}_{\sigma \in \mathcal{F}_K}$. Figures 7 and 9 give an illustration in two space dimensions.

5.6 Fictitious flux reconstruction by lifting the face normal fluxes

In Definition 4.14, on simplicial meshes, we have reconstructed a discrete $\boldsymbol{H}(\text{div}, \Omega)$-conforming flux $\boldsymbol{u}_h$ by lifting the face normal fluxes of a finite volume scheme. We extend here this procedure to general polytopal meshes. In contrast to Theorem 4.18, however, Theorem 5.18 below will not need $\boldsymbol{u}_h$ to evaluate the posteriori error estimate, so one does *not need to perform* physically/on the computer this reconstruction. Congruently, we may not have a closed formula for the reconstruction, which on top may not be discrete. This is the reason for our naming *fictitious*.

A (fictitious) flux reconstruction under Assumption 5.1

We first treat the generic case of Assumption 5.1; the case of Assumption 5.2 is treated below. Following the concept of lifting operators used in, e.g., Eymard *et al.* [127, Section 1.2], Kuznetsov and Repin [172], Brezzi *et al.* [64, Theorem 5.1], Beirão da Veiga [42, Section 2.1], Beirão da Veiga and Manzini [43, Section 2.4], Kuznetsov [173], Vohralík [249, Section 3.2], or Sboui *et al.* [230], we will extend the fluxes U_σ, $\sigma \in \mathcal{F}_K$, to the interior of each polytopal element $K \in \mathcal{T}_H$. The obtained flux $\boldsymbol{u}_h|_K$ will be the approximation of the exact flux $\boldsymbol{u}$ from (5.3) for which we will estimate the energy error $\|\boldsymbol{u} - \boldsymbol{u}_h\|_{\underline{\boldsymbol{K}}^{-\frac{1}{2}}}$. Let us first present a motivation.

Motivation 5.4 (Motivation for the fictitious flux reconstruction). *Following [249, equation (3.14)], consider in each element $K \in \mathcal{T}_H$ the infinite-dimensional Darcy problem: for the fluxes $\{\mathsf{U}_\sigma\}_{\sigma \in \mathcal{F}_K}$ and the potential P_K from Assumption 5.1, find $\tilde{p}_K : K \to \mathbb{R}$ such that*

$$-\nabla\cdot(\underline{\boldsymbol{K}}\nabla\tilde{p}_K) = f|_K, \tag{5.14a}$$

$$\frac{(\tilde{p}_K, 1)_K}{|K|} = \mathsf{P}_K, \tag{5.14b}$$

$$-\underline{\boldsymbol{K}}\nabla\tilde{p}_K \cdot \boldsymbol{n}_\sigma = \frac{\mathsf{U}_\sigma}{|\sigma|} \quad \forall\sigma \in \mathcal{F}_K. \tag{5.14c}$$

We have actually already seen this problem, in Definition 4.16 of Section 4.9. There, on a simplicial mesh, the solution was easily practically computable and discrete: the second-order polynomial denoted as $\tilde{p}_h|_K$ (indeed, (4.24) gives a solution to (5.14)) on a simplicial mesh. Unfortunately, for a general polytopal mesh element $K \in \mathcal{T}_H$, (5.14) amounts to a partial differential equation on K, with the solution $\tilde{p}_K$ living in the infinite-dimensional Sobolev space $H^1(K)$ of Definition 3.3 and the associated flux $\tilde{\boldsymbol{u}}_K := -\underline{\boldsymbol{K}}\nabla\tilde{p}_K$ living in the infinite-dimensional Sobolev space $\boldsymbol{H}(\text{div}, K)$ of Definition 3.7.

Based on Motivation 5.4, we would like to produce a discrete version of (5.14), which will in particular enable us to derive (easily) computable a posteriori error estimates. Following [249, Theorem 5.1] and [254, Theorems 7.2 and 7.3], we consider the mixed finite element method on the simplicial submesh $\mathcal{T}_K$ of K to *approximate the solution* $\tilde{p}_K$ of (5.14). Note that this is to be done separately for each $K \in \mathcal{T}_H$, so it is an intrinsically parallel procedure. Recall Definition 5.3 and consider the following sets and spaces of *piecewise polynomials* over the *simplicial submesh* $\mathcal{T}_K$ of K:

$$
\begin{aligned}
\boldsymbol{V}_{h,\mathrm{N}}^K &:= \{\boldsymbol{v}_h \in \boldsymbol{\mathcal{RT}}_0(\mathcal{T}_K) \cap \boldsymbol{H}(\text{div}, K); \langle\boldsymbol{v}_h\cdot\boldsymbol{n}_\sigma, 1\rangle_\sigma = \mathsf{U}_\sigma \quad \forall\sigma \in \mathcal{F}_{K,h}^{\text{ext}}\}, \\
\boldsymbol{V}_{h,0}^K &:= \{\boldsymbol{v}_h \in \boldsymbol{\mathcal{RT}}_0(\mathcal{T}_K) \cap \boldsymbol{H}(\text{div}, K); \langle\boldsymbol{v}_h\cdot\boldsymbol{n}_\sigma, 1\rangle_\sigma = 0 \quad \forall\sigma \in \mathcal{F}_{K,h}^{\text{ext}}\}, \\
Q_{h,\mathrm{N}}^K &:= \{q_h \in \mathcal{P}_0(\mathcal{T}_K); \frac{(q_h, 1)_K}{|K|} = \mathsf{P}_K\}, \\
Q_{h,0}^K &:= \{q_h \in \mathcal{P}_0(\mathcal{T}_K); (q_h, 1)_K = 0\}.
\end{aligned} \tag{5.15}
$$

Here, $\boldsymbol{V}_{h,\mathrm{N}}^K$ and $\boldsymbol{V}_{h,0}^K$ are subsets/subspaces of the lowest-order Raviart–Thomas spaces as defined in Section 3.5.3, whereas $Q_{h,\mathrm{N}}^K$ and $Q_{h,0}^K$ are subsets/subspaces of the piecewise polynomial spaces as defined in Section 3.5.1. We then let:

Definition 5.5 (Flux reconstruction on polytopal meshes under Assumption 5.1). *Let f be constant on each $K \in \mathcal{T}_H$ and let a discretization scheme of the form of Assumption 5.1 be given, leading to the normal face fluxes $\mathsf{U}_K^{\text{ext}} = \{\mathsf{U}_\sigma\}_{\sigma \in \mathcal{F}_K}$, $K \in \mathcal{T}_H$, and consequently to $\{\mathsf{U}_\sigma\}_{\sigma \in \mathcal{F}_{K,h}^{\text{ext}}}$ as per Definition 5.3. For all $K \in \mathcal{T}_H$, define $\boldsymbol{u}_h|_K$ by the mixed finite element discretization of (5.14), described by the constrained quadratic minimization problem*

$$\boldsymbol{u}_h|_K := \arg \min_{\boldsymbol{v}_h \in \boldsymbol{V}_{h,\mathrm{N}}^K, \nabla\cdot\boldsymbol{v}_h=\text{constant}=f|_K} \|\boldsymbol{v}_h\|_{\underline{\boldsymbol{K}}^{-\frac{1}{2}}, K}^2. \tag{5.16}$$
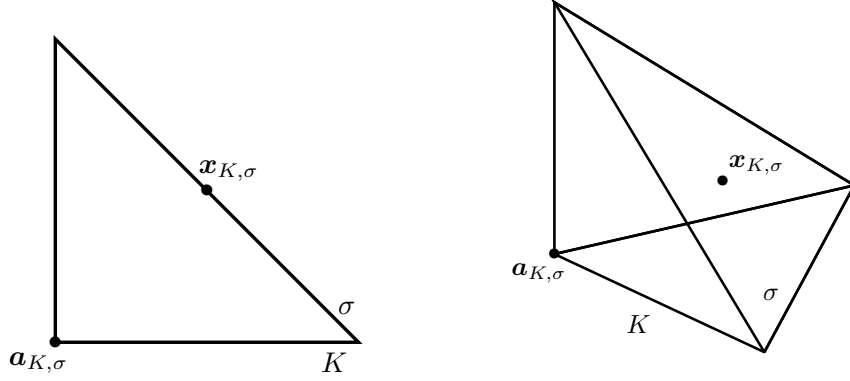
Figure 17: Face barycentres $\boldsymbol{x}_{K,\sigma}$ and vertices $\boldsymbol{a}_{K,\sigma}$ opposite to a face $\sigma \in \mathcal{F}_K$. Simplex $K \in \mathcal{T}_h$ in space dimensions $d = 2$ (left) and $d = 3$ (right)

Equivalently, writing the Euler–Lagrange conditions of (5.16) and imposing the divergence constraint with a Lagrange multiplier, $\boldsymbol{u}_h|_K \in \boldsymbol{V}_{h,\mathrm{N}}^K$ *together with* $p_h|_K \in Q_{h,\mathrm{N}}^K$ *are given by*

$$(\underline{\boldsymbol{K}}^{-1}\boldsymbol{u}_h, \boldsymbol{v}_h)_K - (p_h, \nabla{\cdot}\boldsymbol{v}_h)_K = 0 \qquad\qquad \forall \boldsymbol{v}_h \in \boldsymbol{V}_{h,0}^K, \qquad (5.17\text{a})$$

$$-(\nabla{\cdot}\boldsymbol{u}_h, q_h)_K = -(f, q_h)_K = 0 \qquad \forall q_h \in Q_{h,0}^K. \qquad (5.17\text{b})$$

It is worth noting that the divergence constraint in (5.16) can be written with either a generic constant, or more specifically with $f|_K$; indeed, the assumption that f is constant on each $K \in \mathcal{T}_H$, the definition (5.15) of $\boldsymbol{V}_{h,\mathrm{N}}^K$, and the flux balance (5.8) give the equivalence. Congruently, (5.17b) can be equivalently written as

$$(\nabla{\cdot}\boldsymbol{u}_h, q_h)_K = (f, q_h)_K \qquad \forall q_h \in \mathcal{P}_0(\mathcal{T}_K),$$

which is further equivalent to $\nabla{\cdot}\boldsymbol{u}_h|_K = f|_K$. Note also that since the normal fluxes U_σ are univalued on the faces from $\mathcal{F}_{H,h}$, the resulting flux reconstruction $\boldsymbol{u}_h$ has the normal trace continuous over all interior mesh faces from $\mathcal{F}_{H,h}$ and thus belongs to the Raviart–Thomas space $\boldsymbol{\mathcal{RT}}_0(\mathcal{T}_h) \cap \boldsymbol{H}(\mathrm{div}, \Omega)$ associated with the simplicial mesh $\mathcal{T}_h$ of the entire domain Ω. In consequence, we have:

Lemma 5.6 (Flux reconstruction on polytopal meshes under Assumption 5.1). *Let f be constant on each $K \in \mathcal{T}_H$. Then the flux reconstruction of Definition 5.5 satisfies*

$$\boldsymbol{u}_h \in \boldsymbol{\mathcal{RT}}_0(\mathcal{T}_h) \cap \boldsymbol{H}(\mathrm{div}, \Omega) \quad \text{with} \quad \nabla{\cdot}\boldsymbol{u}_h = f.$$

Construction 5.7 (Matrix form of problem (5.17)). *Let $\boldsymbol{v}_\sigma$, $\sigma \in \mathcal{F}_{K,h}^{\mathrm{int}}$, be the basis functions of the space $\boldsymbol{V}_{h,0}^K$, and $\boldsymbol{v}_\sigma$, $\sigma \in \mathcal{F}_{K,h}^{\mathrm{ext}}$, be the remaining basis functions of the space $\boldsymbol{\mathcal{RT}}_0(\mathcal{T}_K) \cap \boldsymbol{H}(\mathrm{div}, K)$, see (3.13) and Figures 6 and 17. Similarly, let q_i, $1 \le i \le |Q_{h,0}^K|$, be the basis functions of the space $Q_{h,0}^K$ (piecewise constants with mean value 0). Consider these basis functions of $\boldsymbol{V}_{h,0}^K$ and $Q_{h,0}^K$ as test functions $\boldsymbol{v}_h$ and q_h in (5.17) and develop*

$$\boldsymbol{u}_h|_K = \underbrace{\sum_{\sigma \in \mathcal{F}_{K,h}^{\mathrm{ext}}} (\mathsf{U}_K^{\mathrm{ext}})_\sigma \boldsymbol{v}_\sigma}_{\boldsymbol{u}_{h,K}^{\mathrm{ext}}} + \underbrace{\sum_{\sigma \in \mathcal{F}_{K,h}^{\mathrm{int}}} (\mathsf{U}_K^{\mathrm{int}})_\sigma \boldsymbol{v}_\sigma}_{\boldsymbol{u}_{h,K}^{\mathrm{int}}}$$

and

$$p_h|_K = \mathsf{P}_K + \underbrace{\sum_{i=1}^{|Q_{h,0}^K|} (\mathsf{P}_K^0)_i q_i}_{p_{h,K}^0},$$

where $\{\mathsf{U}_\sigma\}_{\sigma \in \mathcal{F}_K}$ and P_K are the given data of the problem (5.14) and the data $\mathsf{U}_K^{\mathrm{ext}} := \{\mathsf{U}_\sigma\}_{\sigma \in \mathcal{F}_{K,h}^{\mathrm{ext}}}$ appearing in problem (5.16) are given by Definition 5.3. Then (5.17) in matrix form corresponds to: find $\mathsf{U}_K^{\mathrm{int}} \in \mathbb{R}^{|\boldsymbol{V}_{h,0}^K|}$ and $\mathsf{P}_K^0 \in \mathbb{R}^{|Q_{h,0}^K|}$ such that

$$\begin{pmatrix} \mathbb{A}_K^{\mathrm{int,int}} & (\mathbb{B}_K^{0,\mathrm{int}})^{\mathrm{t}} \\ \mathbb{B}_K^{0,\mathrm{int}} & 0 \end{pmatrix} \begin{pmatrix} \mathsf{U}_K^{\mathrm{int}} \\ \mathsf{P}_K^0 \end{pmatrix} = \begin{pmatrix} -\mathbb{A}_K^{\mathrm{int,ext}} \mathsf{U}_K^{\mathrm{ext}} \\ -\mathbb{B}_K^{0,\mathrm{ext}} \mathsf{U}_K^{\mathrm{ext}} \end{pmatrix}, \qquad (5.18)$$

33

where

$$\mathbb{A}_K^{\text{int,int}} \in \mathbb{R}^{|\mathcal{F}_{K,h}^{\text{int}}| \times |\mathcal{F}_{K,h}^{\text{int}}|}, \qquad (\mathbb{A}_K^{\text{int,int}})_{\sigma,\sigma'} := (\underline{\boldsymbol{K}}^{-1}\boldsymbol{v}_{\sigma'}, \boldsymbol{v}_\sigma)_K \qquad \sigma, \sigma' \in \mathcal{F}_{K,h}^{\text{int}}, \tag{5.19a}$$

$$\mathbb{A}_K^{\text{int,ext}} \in \mathbb{R}^{|\mathcal{F}_{K,h}^{\text{int}}| \times |\mathcal{F}_{K,h}^{\text{ext}}|} \qquad (\mathbb{A}_K^{\text{int,ext}})_{\sigma,\sigma'} := (\underline{\boldsymbol{K}}^{-1}\boldsymbol{v}_{\sigma'}, \boldsymbol{v}_\sigma)_K \qquad \sigma \in \mathcal{F}_{K,h}^{\text{int}}, \sigma' \in \mathcal{F}_{K,h}^{\text{ext}}, \tag{5.19b}$$

$$\mathbb{B}_K^{0,\text{int}} \in \mathbb{R}^{|Q_{h,0}^K| \times |\mathcal{F}_{K,h}^{\text{int}}|} \qquad (\mathbb{B}_K^{0,\text{int}})_{i,\sigma'} := -(\nabla{\cdot}\boldsymbol{v}_{\sigma'}, q_i)_K \qquad 1 \le i \le |Q_{h,0}^K|, \sigma' \in \mathcal{F}_{K,h}^{\text{int}}, \tag{5.19c}$$

$$\mathbb{B}_K^{0,\text{ext}} \in \mathbb{R}^{|Q_{h,0}^K| \times |\mathcal{F}_{K,h}^{\text{ext}}|} \qquad (\mathbb{B}_K^{0,\text{ext}})_{i,\sigma'} := -(\nabla{\cdot}\boldsymbol{v}_{\sigma'}, q_i)_K \qquad 1 \le i \le |Q_{h,0}^K|, \sigma' \in \mathcal{F}_{K,h}^{\text{ext}}. \tag{5.19d}$$

It will also be useful to define

$$(\mathbb{A}_K^{\text{ext,ext}})_{\sigma,\sigma'} := (\underline{\boldsymbol{K}}^{-1}\boldsymbol{v}_{\sigma'}, \boldsymbol{v}_\sigma)_K \qquad \sigma, \sigma' \in \mathcal{F}_{K,h}^{\text{ext}}.$$

Definition 5.5 is practical and we can solve problem (5.17) or (5.18) to have $\boldsymbol{u}_h$ available. We, however, want to avoid this (unless we want to compute the energy error $\|\boldsymbol{u} - \boldsymbol{u}_h\|_{\underline{\boldsymbol{K}}^{-\frac{1}{2}}}$ in model cases with known exact solution $\boldsymbol{u}$). In Theorem 5.18 below, we will principally only need the energy $\|\boldsymbol{u}_h\|_{\underline{\boldsymbol{K}}^{-\frac{1}{2}}, K}$. In this respect, we observe that Definition 5.5 lifts the information from the boundary of each polytopal element $K \in \mathcal{T}_H$ given by the fluxes $\mathsf{U}_K^{\text{ext}}$ to the interior of the element K. It is thus clear that $\|\boldsymbol{u}_h\|_{\underline{\boldsymbol{K}}^{-\frac{1}{2}}, K}$ must only depend on $\mathsf{U}_K^{\text{ext}}$ and can be obtained directly therefrom. It turns out that this link can be expressed by a single element matrix constructed solely from the geometry of the simplicial submesh $\mathcal{T}_K$ of K, using Construction 5.7. As in [254, proof of Theorem 7.3], we indeed have:

Lemma 5.8 (Energy norm and MFE element matrix)**.** *Let a discretization scheme of the form of Assumption 5.1 be given, leading to the normal face fluxes* $\mathsf{U}_K^{\text{ext}} = \{\mathsf{U}_\sigma\}_{\sigma \in \mathcal{F}_K}$, $K \in \mathcal{T}_H$, *and consequently to* $\{\mathsf{U}_\sigma\}_{\sigma \in \mathcal{F}_{K,h}^{\text{ext}}}$ *as per Definition 5.3. For each* $K \in \mathcal{T}_H$, *let* $\boldsymbol{u}_h|_K \in \boldsymbol{V}_{h,\text{N}}^K$ *be given by Definition 5.5. Relying on Construction 5.7, define the* element matrix *of the mixed finite element method*

$$\widehat{\mathbb{A}}_{\text{MFE},K} := \mathbb{A}_K^{\text{ext,ext}} - \begin{pmatrix} \mathbb{A}_K^{\text{int,ext}} \\ \mathbb{B}_K^{0,\text{ext}} \end{pmatrix}^{\text{t}} \begin{pmatrix} \mathbb{A}_K^{\text{int,int}} & (\mathbb{B}_K^{0,\text{int}})^{\text{t}} \\ \mathbb{B}_K^{0,\text{int}} & 0 \end{pmatrix}^{-1} \begin{pmatrix} \mathbb{A}_K^{\text{int,ext}} \\ \mathbb{B}_K^{0,\text{ext}} \end{pmatrix}, \tag{5.20}$$

where the building matrices are given by (5.19). *Then*

$$\|\boldsymbol{u}_h\|_{\underline{\boldsymbol{K}}^{-\frac{1}{2}}, K}^2 = (\mathsf{U}_K^{\text{ext}})^{\text{t}} \widehat{\mathbb{A}}_{\text{MFE},K} \mathsf{U}_K^{\text{ext}}. \tag{5.21}$$

Proof. Let $K \in \mathcal{T}_H$ be a fixed polytopal mesh element. Following Construction 5.7, decompose $\boldsymbol{u}_h|_K = \boldsymbol{u}_{h,K}^{\text{ext}} + \boldsymbol{u}_{h,K}^{\text{int}}$ and $p_h|_K = \mathsf{P}_K + p_{h,K}^0$. Note that choosing $\boldsymbol{u}_{h,K}^{\text{int}} \in \boldsymbol{V}_{h,0}^K$ as the test function in (5.17a) and $p_{h,K}^0 \in Q_{h,0}^K$ as the test function in (5.17b), one has

$$(\underline{\boldsymbol{K}}^{-1}\boldsymbol{u}_h, \boldsymbol{u}_{h,K}^{\text{int}})_K \overset{(5.17\text{a})}{=} (p_h, \nabla{\cdot}\boldsymbol{u}_{h,K}^{\text{int}})_K = (p_{h,K}^0, \nabla{\cdot}\boldsymbol{u}_{h,K}^{\text{int}})_K \overset{(5.17\text{b})}{=} -(p_{h,K}^0, \nabla{\cdot}\boldsymbol{u}_{h,K}^{\text{ext}})_K,$$

where we have also used $(\mathsf{P}_K, \nabla{\cdot}\boldsymbol{u}_{h,K}^{\text{int}})_K = 0$ in the middle equality. Consequently,

$$\begin{aligned}
\|\boldsymbol{u}_h\|_{\underline{\boldsymbol{K}}^{-\frac{1}{2}}, K}^2 &= (\underline{\boldsymbol{K}}^{-1}\boldsymbol{u}_h, \boldsymbol{u}_h)_K \\
&= (\underline{\boldsymbol{K}}^{-1}\boldsymbol{u}_h, \boldsymbol{u}_{h,K}^{\text{int}})_K + (\underline{\boldsymbol{K}}^{-1}\boldsymbol{u}_{h,K}^{\text{int}}, \boldsymbol{u}_{h,K}^{\text{ext}})_K + (\underline{\boldsymbol{K}}^{-1}\boldsymbol{u}_{h,K}^{\text{ext}}, \boldsymbol{u}_{h,K}^{\text{ext}})_K \\
&= -(p_{h,K}^0, \nabla{\cdot}\boldsymbol{u}_{h,K}^{\text{ext}})_K + (\underline{\boldsymbol{K}}^{-1}\boldsymbol{u}_{h,K}^{\text{int}}, \boldsymbol{u}_{h,K}^{\text{ext}})_K + (\underline{\boldsymbol{K}}^{-1}\boldsymbol{u}_{h,K}^{\text{ext}}, \boldsymbol{u}_{h,K}^{\text{ext}})_K \\
&= (\mathsf{U}_K^{\text{ext}})^{\text{t}} \begin{pmatrix} \mathbb{A}_K^{\text{int,ext}} \\ \mathbb{B}_K^{0,\text{ext}} \end{pmatrix}^{\text{t}} \begin{pmatrix} \mathsf{U}_K^{\text{int}} \\ \mathsf{P}_K^0 \end{pmatrix} + (\mathsf{U}_K^{\text{ext}})^{\text{t}} \mathbb{A}_K^{\text{ext,ext}} \mathsf{U}_K^{\text{ext}}.
\end{aligned}$$

Combining this with the MFE matrix form (5.18) and the MFE element matrix definition (5.20) yields the claim (5.21). $\qquad\square$

A fictitious flux reconstruction under Assumption 5.2

There exist other ways how to lift the face normal fluxes $\mathsf{U}_K^{\text{ext}}$ into the interior of each polytope $K \in \mathcal{T}_H$ than via Definition 5.5. For instance, one could use a finite volume discretization of (5.14) (like (4.19a)) in place of the mixed finite element one (5.17). Under Assumption 5.2, which defines $\tilde{\boldsymbol{u}}_h$, a much more direct and simple procedure suggests itself: to use directly the lifting operator from (5.11). This is in particular the theoretical vehicle in mimetic finite differences according to, e.g., [64, Theorem 5.1]. In

contrast to (5.16), where we can solve for $\boldsymbol{u}_h$ if needed, the reconstruction $\tilde{\boldsymbol{u}}_h$ we now introduce may be truly fictitious in the sense that it may not be practically possible to construct it. Thus, we will not be able to compute the actual error $\|\boldsymbol{u} - \tilde{\boldsymbol{u}}_h\|_{\underline{\boldsymbol{K}}^{-\frac{1}{2}}}$ in model cases with known exact flux $\boldsymbol{u}$, but otherwise this not be an issue (we will still be able to compute the a posteriori error estimate of Theorem 5.18 below). In particular, thanks to (5.11c), we have the following trivial counterpart of Lemma 5.8:

Lemma 5.9 (Energy norm and given discretization scheme element matrix)**.** *Let a discretization scheme of the form of Assumption 5.2 be given, leading to the element matrices $\widehat{\mathbb{A}}_K$ and normal face fluxes $\mathsf{U}_K^{\mathrm{ext}} = \{\mathsf{U}_\sigma\}_{\sigma \in \mathcal{F}_K}$, $K \in \mathcal{T}_H$. Then the lifting $\tilde{\boldsymbol{u}}_h$ of (5.11) satisfies*

$$\|\tilde{\boldsymbol{u}}_h\|_{\underline{\boldsymbol{K}}^{-\frac{1}{2}},K}^2 = (\mathsf{U}_K^{\mathrm{ext}})^{\mathrm{t}} \widehat{\mathbb{A}}_K \mathsf{U}_K^{\mathrm{ext}}. \tag{5.22}$$

Following [254, Theorem 7.3], the mixed finite element element matrix $\widehat{\mathbb{A}}_{\mathrm{MFE},K}$ defined by (5.20) belongs to the same family as the element matrices $\widehat{\mathbb{A}}_K$ of mimetic finite differences, mixed finite volumes, and hybrid finite volumes of Assumption 5.2 when the simplicial faces of $\mathcal{T}_h$ do not subdivide the faces of the polytopal mesh $\mathcal{T}_H$. This leads to the following strong interconnection of Lemmas 5.8 and 5.9 (this is an illuminating observation not needed in practice/thory):

Lemma 5.10 (Approximate energy norm)**.** *Let the assumptions of Lemmas 5.8 and 5.9 and of [64, Theorem 5.1] be satisfied. Let the simplicial faces of $\mathcal{T}_h$ do not subdivide the faces of the polytopal mesh $\mathcal{T}_H$. Then there exist two constants $c, C > 0$ independent of the matrices $\widehat{\mathbb{A}}_K$, $\widehat{\mathbb{A}}_{\mathrm{MFE},K}$, and the mesh size but possibly depending on the shape-regularity of the simplicial submesh $\mathcal{T}_K$ such that*

$$c \underbrace{\|\tilde{\boldsymbol{u}}_h\|_{\underline{\boldsymbol{K}}^{-\frac{1}{2}},K}^2}_{(\mathsf{U}_K^{\mathrm{ext}})^{\mathrm{t}} \widehat{\mathbb{A}}_K \mathsf{U}_K^{\mathrm{ext}}} \leq \underbrace{\|\boldsymbol{u}_h\|_{\underline{\boldsymbol{K}}^{-\frac{1}{2}},K}^2}_{(\mathsf{U}_K^{\mathrm{ext}})^{\mathrm{t}} \widehat{\mathbb{A}}_{\mathrm{MFE},K} \mathsf{U}_K^{\mathrm{ext}}} \leq C \underbrace{\|\tilde{\boldsymbol{u}}_h\|_{\underline{\boldsymbol{K}}^{-\frac{1}{2}},K}^2}_{(\mathsf{U}_K^{\mathrm{ext}})^{\mathrm{t}} \widehat{\mathbb{A}}_K \mathsf{U}_K^{\mathrm{ext}}} \tag{5.23}$$

that we denote by

$$\|\boldsymbol{u}_h\|_{\underline{\boldsymbol{K}}^{-\frac{1}{2}},K}^2 \approx (\mathsf{U}_K^{\mathrm{ext}})^{\mathrm{t}} \widehat{\mathbb{A}}_K \mathsf{U}_K^{\mathrm{ext}}. \tag{5.24}$$

Proof. Any scheme of the form of Assumption 5.2 with the lifting operator according to [64, Theorem 5.1] satisfies the equality (5.7) from [64], i.e., (5.11c). As mixed finite elements on polytopal meshes belong to the mimetic finite difference family by [254, Theorem 7.3], this implies (5.23). □

5.7 Potential point values

In Theorem 5.18 below, we will need a vector of real values with one value $\mathsf{Z}_{\boldsymbol{a}}$ per vertex $\boldsymbol{a} \in \mathcal{V}_h$ of the simplicial submesh $\mathcal{T}_h$ of $\mathcal{T}_H$ and one real value Z_σ per face $\sigma \in \mathcal{F}_{H,h}$ of the simplicial submesh $\mathcal{T}_h$ of $\mathcal{T}_H$ (approximate point values of the pressure heads). We identify them here for the two Assumptions 5.1 and 5.2.

Potential point values under Assumption 5.1

Here, under the general Assumption 5.1, we define the vectors $\{\mathsf{Z}_{\boldsymbol{a}}\}_{\boldsymbol{a} \in \mathcal{V}_h}$ and $\{\mathsf{Z}_\sigma\}_{\sigma \in \mathcal{F}_{H,h}}$ by average values of P_K in the neighboring polytopal elements:

Definition 5.11 (Potential point values on polytopal meshes under Assumption 5.1)**.** *Let a discretization scheme of the form of Assumption 5.1 be given, leading to the elementwise pressure heads $\mathsf{P} := \{\mathsf{P}_K\}_{K \in \mathcal{T}_H}$. For each vertex $\boldsymbol{a}$ of the simplicial mesh $\mathcal{T}_h$ lying on ∂K for some $K \in \mathcal{T}_H$ but not on $\partial\Omega$, let $\mathcal{T}_{\boldsymbol{a}}$ denote the set of polytopal elements $K \in \mathcal{T}_H$ sharing $\boldsymbol{a}$. We set*

$$\mathsf{Z}_{\boldsymbol{a}} := \frac{1}{|\mathcal{T}_{\boldsymbol{a}}|} \sum_{K \in \mathcal{T}_{\boldsymbol{a}}} \mathsf{P}_K, \tag{5.25a}$$

where, recall, $|\mathcal{T}_{\boldsymbol{a}}|$ is the cardinality of the set $\mathcal{T}_{\boldsymbol{a}}$, i.e., the number of polytopal elements $K \in \mathcal{T}_H$ sharing $\boldsymbol{a}$. We also set

$$\mathsf{Z}_{\boldsymbol{a}} := \mathsf{P}_K \text{ for any vertex } \boldsymbol{a} \text{ of } \mathcal{T}_h \text{ lying inside some } K \in \mathcal{T}_H, \tag{5.25b}$$
$$\mathsf{Z}_{\boldsymbol{a}} := 0 \text{ for any vertex } \boldsymbol{a} \text{ lying simultaneously on } \partial K \text{ for some } K \in \mathcal{T}_H \text{ and on } \partial\Omega, \tag{5.25c}$$

$$\mathcal{V}_K = \{\mathbf{a}_i\}_{i=1}^{7}$$
$$\mathcal{V}_{K,h} = \{\mathbf{a}_i\}_{i=1}^{8}$$
$$\mathcal{F}_{K,h}^{\text{ext}} = \{\sigma_i\}_{i=1}^{7}$$
$$\mathsf{Z}_K = \{\mathsf{Z}_{\mathbf{a}_i}\}_{i=1}^{8}$$
$$\mathsf{Z}_K^{\text{ext}} = \{\mathsf{Z}_{\sigma_i}\}_{i=1}^{7}$$

Figure 18: Example of element vectors Z_K and $\mathsf{Z}_K^{\text{ext}}$ of nodal and facial potential point values

where the latter condition is set in accordance with the boundary condition (5.1b). *We will also need the face values* $\{\mathsf{Z}_\sigma\}_{\sigma \in \mathcal{F}_{H,h}}$

$$\mathsf{Z}_\sigma := \frac{1}{d} \sum_{\mathbf{a} \in \mathcal{V}_\sigma} \mathsf{Z}_{\mathbf{a}} \tag{5.26}$$

for any simplicial face $\sigma \in \mathcal{F}_{H,h}$, *where, recall,* $\mathcal{V}_\sigma$ *collects the vertices of the given face* $\sigma \in \mathcal{F}_{H,h}$ *and* d *is the space dimension. Then, for each polytopal cell* $K \in \mathcal{T}_H$, *we let the element vector* $\mathsf{Z}_K \in \mathbb{R}^{|\mathcal{V}_{K,h}|}$ *collect the values* $\mathsf{Z}_{\mathbf{a}}$ *for the vertices* $\mathbf{a}$ *of the simplicial mesh* $\mathcal{T}_K$ *lying on the boundary* ∂K *and the value(s) inside* K

$$\mathsf{Z}_K := \{\mathsf{Z}_{\mathbf{a}}\}_{\mathbf{a} \in \mathcal{V}_{K,h}} \tag{5.27}$$

and the element vector $\mathsf{Z}_K^{\text{ext}} \in \mathbb{R}^{|\mathcal{F}_{K,h}^{\text{ext}}|}$ *collect the values associated with faces* $\sigma \in \mathcal{F}_{K,h}^{\text{ext}}$

$$\mathsf{Z}_K^{\text{ext}} := \{\mathsf{Z}_\sigma\}_{\sigma \in \mathcal{F}_{K,h}^{\text{ext}}}. \tag{5.28}$$

Figure 18 gives an illustration in two space dimensions.

Potential point values under Assumption 5.2

A more straightforward and precise potential reconstruction can be designed under Assumption 5.2. The main idea is that any scheme of the form (5.9) can be hybridized, following, e.g., Roberts and Thomas [224], Brezzi and Fortin [63], or Droniou *et al.* [110], giving rise to one Lagrange multiplier Λ_σ per face $\sigma \in \mathcal{F}_H$:

Definition 5.12 (Lagrange multipliers under Assumption 5.2)**.** *Let a discretization scheme of the form of Assumption 5.2 be given and fix a face* $\sigma \in \mathcal{F}_H$. *Consider the line associated with* σ *in the first block equation of* (5.9). *For an interior face* σ, *let it be shared by two polytopes* $K, L \in \mathcal{T}_H$ *such that* $\boldsymbol{n}_\sigma$ *points from* K *to* L. *From the structure of the matrices* $\mathbb{A}$ *and* $\mathbb{B}$ *supposed in Assumption 5.2, it follows that*

$$\sum_{\sigma' \in \mathcal{F}_K} \boldsymbol{n}_K \cdot \boldsymbol{n}_{\sigma'} (\widehat{\mathbb{A}}_K)_{\sigma,\sigma'} \mathsf{U}_{\sigma'} - \mathsf{P}_K = \sum_{\sigma' \in \mathcal{F}_L} \boldsymbol{n}_L \cdot \boldsymbol{n}_{\sigma'} (\widehat{\mathbb{A}}_L)_{\sigma,\sigma'} \mathsf{U}_{\sigma'} - \mathsf{P}_L,$$

so that these expressions are univalued from both elements K *and* L. *Such an expression is also clearly univalued on boundary faces. It allows to define the Lagrange multipliers*

$$\Lambda_\sigma := \mathsf{P}_K - \sum_{\sigma' \in \mathcal{F}_K} \boldsymbol{n}_K \cdot \boldsymbol{n}_{\sigma'} (\widehat{\mathbb{A}}_K)_{\sigma,\sigma'} \mathsf{U}_{\sigma'} \tag{5.29}$$

for each element $K \in \mathcal{T}_H$ *and each face* $\sigma \in \mathcal{F}_K$.

Definition 5.13 (Potential point values on polytopal meshes under Assumption 5.2)**.** *Let a discretization scheme of the form of Assumption 5.2 be given, leading to the normal face fluxes* $\mathsf{U} = \{\mathsf{U}_\sigma\}_{\sigma \in \mathcal{F}_H}$ *and the elementwise pressure heads* $\mathsf{P} := \{\mathsf{P}_K\}_{K \in \mathcal{T}_H}$. *Let the Lagrange multipliers* Λ_σ *be given by Definition 5.12. For each vertex* $\mathbf{a}$ *of the simplicial mesh* $\mathcal{T}_h$ *lying on* ∂K *for some* $K \in \mathcal{T}_H$ *but not on* $\partial \Omega$, *let* $\mathcal{F}_{\mathbf{a}}$ *denote the set of polytopal faces* $\sigma \in \mathcal{F}_H$ *sharing* $\mathbf{a}$ *and* $|\mathcal{F}_{\mathbf{a}}|$ *its cardinality. We set*

$$\mathsf{Z}_{\mathbf{a}} := \frac{1}{|\mathcal{F}_{\mathbf{a}}|} \sum_{\sigma \in \mathcal{F}_{\mathbf{a}}} \Lambda_\sigma. \tag{5.30a}$$

We also set

$$\mathsf{Z_a} := \mathsf{P}_K \text{ for any vertex } \boldsymbol{a} \text{ of } \mathcal{T}_h \text{ lying inside some } K \in \mathcal{T}_H, \tag{5.30b}$$

$$\mathsf{Z_a} := 0 \text{ for any vertex } \boldsymbol{a} \text{ lying simultaneously on } \partial K \text{ for some } K \in \mathcal{T}_H \text{ and on } \partial\Omega, \tag{5.30c}$$
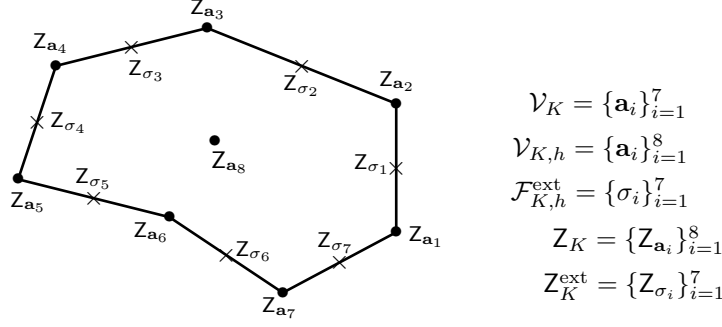
where the latter condition is set in accordance with the boundary condition (5.1b). *We then define the face values* $\{\mathsf{Z}_\sigma\}_{\sigma \in \mathcal{F}_{H,h}}$ *and the element vectors* $\mathsf{Z}_K \in \mathbb{R}^{|\mathcal{V}_{K,h}|}$ *and* $\mathsf{Z}_K^{\mathrm{ext}} \in \mathbb{R}^{|\mathcal{F}_{K,h}^{\mathrm{ext}}|}$ *by respectively* (5.26), (5.27), *and* (5.28) *as in Definition* 5.11.

5.8 Fictitious potential reconstruction by respecting the point values

In Definition 4.17, we have reconstructed a discrete $H_0^1(\Omega)$-conforming potential ζ_h from a finite volume approximation on a simplicial mesh. We extend here this procedure to general polytopal meshes. As in Section 5.6, in contrast to Theorem 4.18, however, Theorem 5.18 below will not need ζ_h to evaluate the posteriori error estimate, so one again does *not need to perform* physically/on the computer this reconstruction. This is the reason for our naming *fictitious* (though we will give a closed formula for ζ_h and ζ_h will be discrete and possible to construct if desired). Recall the notation from Section 3.6.

Given the point values $\mathsf{Z_a}$, either prescribed by (5.25) from Definition 5.11 or by (5.30) of Definition 5.13, the potential reconstruction $\zeta_h \in \mathcal{P}_1(\mathcal{T}_h) \cap H_0^1(\Omega)$ (piecewise affine with respect to the simplicial submesh $\mathcal{T}_h$ and $H_0^1(\Omega)$-conforming) is trivial:

Definition 5.14 (Potential reconstruction on polytopal meshes)**.** *Let the potential point values* $\mathsf{Z_a}$ *be given by either Definition* 5.11 *or Definition* 5.13. *Then the potential reconstruction* $\zeta_h \in \mathcal{P}_1(\mathcal{T}_h) \cap H_0^1(\Omega)$ *is prescribed by*

$$\zeta_h(\boldsymbol{a}) := \mathsf{Z_a} \qquad \forall \boldsymbol{a} \in \mathcal{V}_h.$$

Note that by (5.26), since ζ_h is affine on each face,

$$(\mathsf{Z}_K^{\mathrm{ext}})_\sigma = \mathsf{Z}_\sigma = \zeta_h(\boldsymbol{x}_\sigma), \tag{5.31}$$

i.e., the face values Z_σ are the punctual values of the reconstruction ζ_h in the face barycenters $\boldsymbol{x}_\sigma$, $\sigma \in \mathcal{F}_{K,h}^{\mathrm{ext}}$, cf. Figure 17.

Consider the usual hat basis functions $\psi_{\boldsymbol{a}}$ of the space $\mathcal{P}_1(\mathcal{T}_h) \cap H^1(\Omega)$ on the simplicial mesh $\mathcal{T}_h$: $\psi_{\boldsymbol{a}}$ is piecewise affine on $\mathcal{T}_h$ and $H^1(\Omega)$-conforming, $\psi_{\boldsymbol{a}}(\boldsymbol{a}') = 1$ if $\boldsymbol{a} = \boldsymbol{a}'$ and 0 otherwise, where $\boldsymbol{a}, \boldsymbol{a}' \in \mathcal{V}_h$ are the vertices of $\mathcal{T}_h$. For any polytopal element $K \in \mathcal{T}_H$, we will need below the element *stiffness matrix* $\widehat{\mathbb{S}}_{\mathrm{FE},K} \in \mathbb{R}^{|\mathcal{V}_{K,h}| \times |\mathcal{V}_{K,h}|}$ defined by

$$(\widehat{\mathbb{S}}_{\mathrm{FE},K})_{\boldsymbol{a},\boldsymbol{a}'} := (\underline{\boldsymbol{K}} \nabla \psi_{\boldsymbol{a}'}, \nabla \psi_{\boldsymbol{a}})_K \qquad \boldsymbol{a}, \boldsymbol{a}' \in \mathcal{V}_{K,h} \tag{5.32}$$

and the element *mass matrix* $\widehat{\mathbb{M}}_{\mathrm{FE},K} \in \mathbb{R}^{|\mathcal{V}_{K,h}| \times |\mathcal{V}_{K,h}|}$ defined by

$$(\widehat{\mathbb{M}}_{\mathrm{FE},K})_{\boldsymbol{a},\boldsymbol{a}'} := (\psi_{\boldsymbol{a}'}, \psi_{\boldsymbol{a}})_K \qquad \boldsymbol{a}, \boldsymbol{a}' \in \mathcal{V}_{K,h}. \tag{5.33}$$

An immediate consequence is:

Lemma 5.15 (Energy norm and integration using the stiffness and mass matrices)**.** *Let the potential reconstruction be given by Definition* 5.14 *and the stiffness and mass matrices respectively by* (5.32) *and* (5.33). *Then there holds*

$$\|\underline{\boldsymbol{K}} \nabla \zeta_h\|_{\underline{\boldsymbol{K}}^{-\frac{1}{2}}, K}^2 = \mathsf{Z}_K^{\mathrm{t}} \widehat{\mathbb{S}}_{\mathrm{FE},K} \mathsf{Z}_K \tag{5.34}$$

and

$$(1, \zeta_h)_K = \mathbf{1}^{\mathrm{t}} \widehat{\mathbb{M}}_{\mathrm{FE},K} \mathsf{Z}_K. \tag{5.35}$$

Remark 5.16 (Element matrices $\widehat{\mathbb{S}}_{\mathrm{FE},K}$ and $\widehat{\mathbb{M}}_{\mathrm{FE},K}$)**.** *To obtain the finite element stiffness and mass matrices* $\widehat{\mathbb{S}}_{\mathrm{FE},K}$ *of* (5.32) *and* $\widehat{\mathbb{M}}_{\mathrm{FE},K}$ *of* (5.33), *one could use a finite element assembly code linked to a simplicial mesh on the polytopal cell* $K \in \mathcal{T}_H$. *Note, however, that there exist simple analytical formulas for the basis functions* $\psi_{\boldsymbol{a}}$ *of piecewise affine Lagrange finite elements and for their gradients* $\nabla \psi_{\boldsymbol{a}}$ *that merely necessitate the position of the vertices of the polygonal element* K *and of the central point. In this sense, we say that neither* $\mathcal{T}_h$ *nor* $\mathcal{T}_K$ *need not be constructed in practice in order to obtain* $\widehat{\mathbb{S}}_{\mathrm{FE},K}$ *and* $\widehat{\mathbb{M}}_{\mathrm{FE},K}$.

We finish this section by a remark on an alternative potential reconstruction in the spirit of Definitions 4.16–4.17. This is still more precise, though slightly more computationally demanding; we will use it for comparison in numerical experiments below.

Remark 5.17 (Piecewise quadratic potential reconstruction). *Let $(\boldsymbol{u}_h|_K, p_h|_K)$ be given by Definition 5.5 on all $K \in \mathcal{T}_H$. We let $\tilde{p}_h$ be a piecewise quadratic polynomial on the simplicial mesh $\mathcal{T}_h$ given by*

$$-\underline{\boldsymbol{K}} \nabla \tilde{p}_h|_\kappa = \boldsymbol{u}_h|_\kappa, \qquad \frac{(\tilde{p}_h, 1)_\kappa}{|\kappa|} = p_h|_\kappa \qquad \forall \kappa \in \mathcal{T}_h.$$

Then a piecewise quadratic potential reconstruction $\zeta_h \in \mathcal{P}_2(\mathcal{T}_h) \cap H_0^1(\Omega)$ can be obtained by averaging the values of $\tilde{p}_h$ in all Lagrange nodes of freedom as per Definition 4.17.

5.9 A guaranteed a posteriori error estimate with inexpensive implementation and evaluation

We now present our a posteriori error estimate for the model problem (5.1). Our main result is given for the error between the exact Darcy velocity $\boldsymbol{u} \in \boldsymbol{H}(\mathrm{div}, \Omega)$ of (5.2)–(5.3) and the reconstruction $\boldsymbol{u}_h \in \mathcal{R}\mathcal{T}_0(\mathcal{T}_h) \cap \boldsymbol{H}(\mathrm{div}, \Omega)$ of Definition 5.5 evaluated as per (5.5), under the general Assumption 5.1:

Theorem 5.18 (A guaranteed a posteriori error estimate with inexpensive implementation and evaluation).

Let f be constant on each $K \in \mathcal{T}_H$ and let the exact flux $\boldsymbol{u}$ be given by (5.2)–(5.3). For any polytopal discretization satisfying Assumption 5.1, let, for each polytopal element $K \in \mathcal{T}_H$, the element vectors of face normal fluxes $\mathsf{U}_K^{\mathrm{ext}}$ be given by Definition 5.3. Let the element vectors of potential point values Z_K and $\mathsf{Z}_K^{\mathrm{ext}}$ be given by Definition 5.11. Let finally the matrices $\widehat{\mathbb{A}}_{\mathrm{MFE},K}$, $\widehat{\mathbb{S}}_{\mathrm{FE},K}$, and $\widehat{\mathbb{M}}_{\mathrm{FE},K}$ be respectively defined by (5.20), (5.32), and (5.33). Then there holds

$$\|\boldsymbol{u} - \boldsymbol{u}_h\|_{\underline{\boldsymbol{K}}^{-\frac{1}{2}}} \leq \left\{ \sum_{K \in \mathcal{T}_H} \eta_K^2 \right\}^{1/2}, \tag{5.36}$$

where

$$\eta_K^2 := (\mathsf{U}_K^{\mathrm{ext}})^{\mathrm{t}} \widehat{\mathbb{A}}_{\mathrm{MFE},K} \mathsf{U}_K^{\mathrm{ext}} + \mathsf{Z}_K^{\mathrm{t}} \widehat{\mathbb{S}}_{\mathrm{FE},K} \mathsf{Z}_K + 2(\mathsf{U}_K^{\mathrm{ext}})^{\mathrm{t}} \mathsf{Z}_K^{\mathrm{ext}} - 2\mathsf{F}_K |K|^{-1} \mathbb{1}^{\mathrm{t}} \widehat{\mathbb{M}}_{\mathrm{FE},K} \mathsf{Z}_K. \tag{5.37}$$

Here the flux reconstruction $\boldsymbol{u}_h \in \mathcal{R}\mathcal{T}_0(\mathcal{T}_h) \cap \boldsymbol{H}(\mathrm{div}, \Omega)$ with $\nabla \cdot \boldsymbol{u}_h = f$ is obtained following Definition 5.5.

Proof. Recall from Lemma 5.6 that $\boldsymbol{u}_h \in \boldsymbol{H}(\mathrm{div}, \Omega)$ with $\nabla \cdot \boldsymbol{u}_h = f$. Thus, considering the diffusion tensor $\underline{\boldsymbol{K}}$ in place of the identity matrix and proceeding as for the Prager–Synge equality of Corollary 4.6, we have

$$\|\boldsymbol{u} - \boldsymbol{u}_h\|_{\underline{\boldsymbol{K}}^{-\frac{1}{2}}} = \min_{v \in H_0^1(\Omega)} \|\boldsymbol{u}_h + \underline{\boldsymbol{K}} \nabla v\|_{\underline{\boldsymbol{K}}^{-\frac{1}{2}}}. \tag{5.38}$$

Consequently, for an arbitrary $\zeta_h \in H_0^1(\Omega)$,

$$\|\boldsymbol{u} - \boldsymbol{u}_h\|_{\underline{\boldsymbol{K}}^{-\frac{1}{2}}} \leq \|\boldsymbol{u}_h + \underline{\boldsymbol{K}} \nabla \zeta_h\|_{\underline{\boldsymbol{K}}^{-\frac{1}{2}}}.$$

We now choose for ζ_h the fictitious potential reconstruction $\zeta_h \in \mathcal{P}_1(\mathcal{T}_h) \cap H_0^1(\Omega)$ of Definition 5.14, continuous and piecewise affine with respect to the simplicial submesh $\mathcal{T}_h$ and given by the nodal values of the vector Z from Definition 5.11. We focus on the evaluation of $\|\boldsymbol{u}_h + \underline{\boldsymbol{K}} \nabla \zeta_h\|_{\underline{\boldsymbol{K}}^{-\frac{1}{2}}}$. Developing, we obtain, for each polytopal mesh element $K \in \mathcal{T}_H$,

$$\|\boldsymbol{u}_h + \underline{\boldsymbol{K}} \nabla \zeta_h\|_{\underline{\boldsymbol{K}}^{-\frac{1}{2}}, K}^2 = \|\boldsymbol{u}_h\|_{\underline{\boldsymbol{K}}^{-\frac{1}{2}}, K}^2 + 2(\boldsymbol{u}_h, \nabla \zeta_h)_K + \|\underline{\boldsymbol{K}} \nabla \zeta_h\|_{\underline{\boldsymbol{K}}^{-\frac{1}{2}}, K}^2.$$

We now use Lemma 5.8 for the first term and (5.34) from Lemma 5.15 for the last one. For the middle term, recall first that the normal components of vector fields in $\mathcal{R}\mathcal{T}_0(\mathcal{T}_h) \cap \boldsymbol{H}(\mathrm{div}, \Omega)$ are constant on

each face. Thus the Green theorem together with (5.26) and its consequence (5.31), $\nabla\cdot\boldsymbol{u}_h = f$ and (5.7), and (5.35) from Lemma 5.15 give

$$(\boldsymbol{u}_h, \nabla\zeta_h)_K = \langle \boldsymbol{u}_h\cdot\boldsymbol{n}, \zeta_h\rangle_{\partial K} - (\nabla\cdot\boldsymbol{u}_h, \zeta_h)_K = (\mathsf{U}_K^{\mathrm{ext}})^{\mathrm{t}}\mathsf{Z}_K^{\mathrm{ext}} - \mathsf{F}_K|K|^{-1}\mathbf{1}^{\mathrm{t}}\widehat{\mathbb{M}}_{\mathrm{FE},K}\mathsf{Z}_K. \tag{5.39}$$

Thus the proof is finished. $\qquad\square$

Note that the a posteriori error estimators η_K of Theorem 5.18 take a simple form of *matrix-vector multiplication* on each polytopal mesh element $K \in \mathcal{T}_H$, only need the element vectors $\mathsf{U}_K^{\mathrm{ext}}$, Z_K, and $\mathsf{Z}_K^{\mathrm{ext}}$ together with the element matrices $\widehat{\mathbb{A}}_{\mathrm{MFE},K}$, $\widehat{\mathbb{S}}_{\mathrm{FE},K}$, and $\widehat{\mathbb{M}}_{\mathrm{FE},K}$, but *not the reconstructions* $\boldsymbol{u}_h$ and ζ_h (which become fictitious), and yet (5.36) delivers a *guaranteed upper bound* on the Darcy velocity error.

For schemes of the form of Assumption 5.2, Definition 5.13 can alternatively be used in place of Definition 5.11 in Theorem 5.18. In this case, however, it is probably still more attractive to proceed as:

Corollary 5.19 (A simple guaranteed estimate with the given element matrices $\widehat{\mathbb{A}}_K$)**.** *Let f be constant on each $K \in \mathcal{T}_H$ and let the exact flux $\boldsymbol{u}$ be given by (5.2)–(5.3). For any polytopal discretization satisfying Assumption 5.2, let, for each polytopal element $K \in \mathcal{T}_H$, the element vectors of face normal fluxes $\mathsf{U}_K^{\mathrm{ext}}$ be given by Definition 5.3. Let the element vectors of potential point values Z_K and $\mathsf{Z}_K^{\mathrm{ext}}$ be given by Definition 5.13. Let the matrices $\widehat{\mathbb{A}}_K$ be the element matrices from Assumption 5.2. Let finally $\widehat{\mathbb{S}}_{\mathrm{FE},K}$ and $\widehat{\mathbb{M}}_{\mathrm{FE},K}$ be respectively defined by (5.32) and (5.33). Then there holds*

$$\|\boldsymbol{u} - \tilde{\boldsymbol{u}}_h\|_{\underline{K}^{-\frac{1}{2}}} \leq \left\{ \sum_{K\in\mathcal{T}_H} \eta_K^2 \right\}^{1/2}, \tag{5.40}$$

where

$$\eta_K^2 := (\mathsf{U}_K^{\mathrm{ext}})^{\mathrm{t}}\widehat{\mathbb{A}}_K\mathsf{U}_K^{\mathrm{ext}} + \mathsf{Z}_K^{\mathrm{t}}\widehat{\mathbb{S}}_{\mathrm{FE},K}\mathsf{Z}_K + 2(\mathsf{U}_K^{\mathrm{ext}})^{\mathrm{t}}\mathsf{Z}_K^{\mathrm{ext}} - 2\mathsf{F}_K|K|^{-1}\mathbf{1}^{\mathrm{t}}\widehat{\mathbb{M}}_{\mathrm{FE},K}\mathsf{Z}_K, \tag{5.41}$$

i.e., the same form as (5.37) but with $\widehat{\mathbb{A}}_K$ in place of $\widehat{\mathbb{A}}_{\mathrm{MFE},K}$. Here the fictitious flux reconstruction $\tilde{\boldsymbol{u}}_h \in \boldsymbol{\mathcal{RT}}_0(\mathcal{T}_h) \cap \boldsymbol{H}(\mathrm{div}, \Omega)$ with $\nabla\cdot\boldsymbol{u}_h = f$ is obtained following (5.11).

Proof. The proof is similar to that of Theorem 5.18, using the facts that $\tilde{\boldsymbol{u}}_h \in \boldsymbol{H}(\mathrm{div}, \Omega)$, $\nabla\cdot\tilde{\boldsymbol{u}}_h = f$, and $\tilde{\boldsymbol{u}}_h\cdot\boldsymbol{n}_\sigma = \mathsf{U}_\sigma/|\sigma|$ for all $\sigma \in \mathcal{F}_K$ following (5.11a), and replacing Lemma 5.8 by Lemma 5.9. $\qquad\square$

Finally, relying on Lemma 5.10, we also have:

Corollary 5.20 (A simple estimate with the given element matrices $\widehat{\mathbb{A}}_K$)**.** *Let the assumptions of Corollary 5.19 hold, let $\mathcal{F}_{K,h}^{\mathrm{ext}} = \mathcal{F}_K$, i.e., the simplicial faces of $\mathcal{T}_h$ do not subdivide the faces of the polytopal mesh $\mathcal{T}_H$, and let the flux $\boldsymbol{u}_h \in \boldsymbol{\mathcal{RT}}_0(\mathcal{T}_h) \cap \boldsymbol{H}(\mathrm{div}, \Omega)$ be constructed by Definition 5.5. Then*

$$\|\boldsymbol{u} - \boldsymbol{u}_h\|_{\underline{K}^{-\frac{1}{2}}} \lesssim \left\{ \sum_{K\in\mathcal{T}_H} \eta_K^2 \right\}^{1/2}, \tag{5.42}$$

where η_K is given by (5.41) and the "approximately less than or equal to" denoted by $\lesssim$ stems from the approximation (5.24).

A few comments are in order:

Remark 5.21 (Comparison of Theorem 5.18 with Corollaries 5.19 and 5.20)**.** *Theorem 5.18 is applicable to any scheme of the form of Assumption 5.1 and relies on the the mixed finite element matrices $\widehat{\mathbb{A}}_{\mathrm{MFE},K}$ that one needs to construct by (5.20). In contrast, Corollaries 5.19 and 5.20 are only applicable to schemes of the form of Assumption 5.2, but use directly the matrices $\widehat{\mathbb{A}}_K$. The estimates (5.36) and (5.42) hold for the reconstruction $\boldsymbol{u}_h$ that one can compute in practice if needed by Definition 5.5, whereas (5.40) only holds for the generally unavailable lifting $\tilde{\boldsymbol{u}}_h$, as in [42, 43]. Finally, both (5.36) and (5.40) are guaranteed, with the $\leq$ inequality, whereas (5.42) is not. In the numerical experiments in Section 5.10 below, though, it is hard to distinguish the two estimators (5.37) and (5.41).*

Remark 5.22 (Simplicial submeshes and solution of local problems)**.** *We can avoid the physical construction of the simplicial submeshes $\mathcal{T}_K$ under the conditions that 1) the given scheme takes the form of Assumption 5.2; 2) the element matrices $\widehat{\mathbb{A}}_K$ are explicitly given; 3) the stiffness matrix $\widehat{\mathbb{S}}_{\mathrm{FE},K}$ of (5.32) and the mass matrix $\widehat{\mathbb{M}}_{\mathrm{FE},K}$ of (5.33) of the conforming finite element method on $\mathcal{T}_K$ for each polytopal*
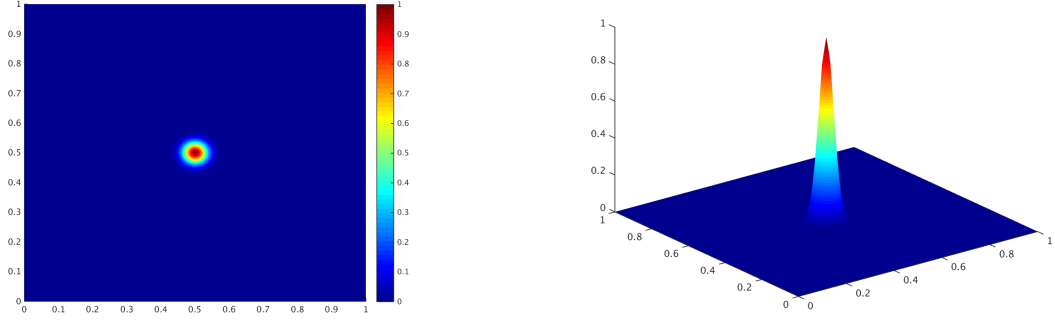
Figure 19: The exact solution (5.43)

cell $K \in \mathcal{T}_H$ are constructed from the geometry of $\mathcal{T}_K$ only, in the sense of Remark 5.16; 4) we use the estimator η_K (5.41) of Corollary 5.19 or 5.20. Note, however, that one may need a submesh of each polytopal cell K already to produce the scheme element matrix $\widehat{\mathbb{A}}_K$. Similarly, a typical implementation of $\widehat{\mathbb{A}}_{\mathrm{MFE},K}$ (5.20) will also need the submesh $\mathcal{T}_K$, though it can also be constructed from the geometry of $\mathcal{T}_K$ only. Remark finally that behind (5.20), there is the local Neumann problem of Definition 5.5.

Remark 5.23 (Higher-order locally conservative methods)**.** *Higher-order locally conservative methods can be treated similarly to the above exposition but are not subject of this work.*

5.10 Numerical experiments

The purpose of this section is to numerically illustrate the performance of the estimators of Theorem 5.18, as well as the simplified estimate (5.42) of Corollary 5.20 . We also compare this methodology to a reference estimate on the simplicial submesh in particular using the potential reconstruction of Remark 5.17. The test is taken from [188], where a collection of two-dimensional elliptic problems for testing adaptive grid refinement algorithms is proposed. We approximate $-\Delta p = f$ on the space domain $\Omega = (0,1)^2$ where the analytical solution is

$$p(x,y) = 2^{4\alpha} x^{\alpha} (1-x)^{\alpha} y^{\alpha} (1-y)^{\alpha}, \quad \alpha = 200, \tag{5.43}$$

see Figure 19. The source term f prescribed correspondingly and we neglect here that it is not piecewise constant, i.e., the data oscillation term as the second term in (4.26b). We use homogeneous Dirichlet boundary conditions.

The mesh $\mathcal{T}_H$ consists of general polygonal elements; together with the fictitious triangular submesh $\mathcal{T}_h$, it is shown in Figure 8. We consider the hybrid finite volume (HFV) discretization (Droniou *et al.* [128] or [110, Section 2.2]), taking the form (5.9) of Assumption 5.2 with the matrix $\mathbb{A}$ formed by local element matrices $\widehat{\mathbb{A}}_K$. We compare three versions of a posteriori error estimates: a) Theorem 5.18, with the estimators η_K evaluated via the matrices $\widehat{\mathbb{A}}_{\mathrm{MFE},K}$, $\widehat{\mathbb{S}}_{\mathrm{FE},K}$, and $\widehat{\mathbb{M}}_{\mathrm{FE},K}$ (polygonal MFE estimate); b) Corollary 5.20, using the element matrix $\widehat{\mathbb{A}}_K$ of the HFV scheme that is already available (polygonal HFV estimate); c) estimate (5.36) with the estimators η_K replaced by the expression $\|\underline{\boldsymbol{K}}^{-1/2}\boldsymbol{u}_h + \underline{\boldsymbol{K}}^{1/2}\nabla\zeta_h\|_K$, where $\boldsymbol{u}_h|_K \in \boldsymbol{V}_{h,\mathrm{N}}^K$ is given in Definition 5.5 and $\zeta_h \in \mathcal{P}_2(\mathcal{T}_h) \cap H_0^1(\Omega)$ is described in Remark 5.17 (triangular estimate, sharpest-possible, for comparison); In the first two cases, we use Definition 5.11 to obtain the vectors of the potential point values Z_K and $\mathsf{Z}_K^{\mathrm{ext}}$ by (5.25) and (5.26).

Note that if the mixed finite element method on polygonal meshes of [254, Theorem 7.2] was used instead of the HFV discretization, the third procedure would be, following [254, Remark 7.3], fully equivalent to solving the problem (5.1) directly on the simplicial mesh $\mathcal{T}_h$ by the lowest-order Raviart–Thomas mixed finite element method and applying the a posteriori error estimates of Theorem 4.18. This is why the triangular estimate serves here as a reference a posteriori error estimate. The first procedure only uses the piecewise affine potential reconstruction ζ_h of Definitions 5.11 and 5.14, which allows for the simple matrix form (5.37) of the estimators. The second procedure is definitely the easiest choice in practice, where only the already available element matrices $\widehat{\mathbb{A}}_K$ are used and there is no need to construct the mixed finite element matrices $\widehat{\mathbb{A}}_{\mathrm{MFE},K}$ via (5.20).

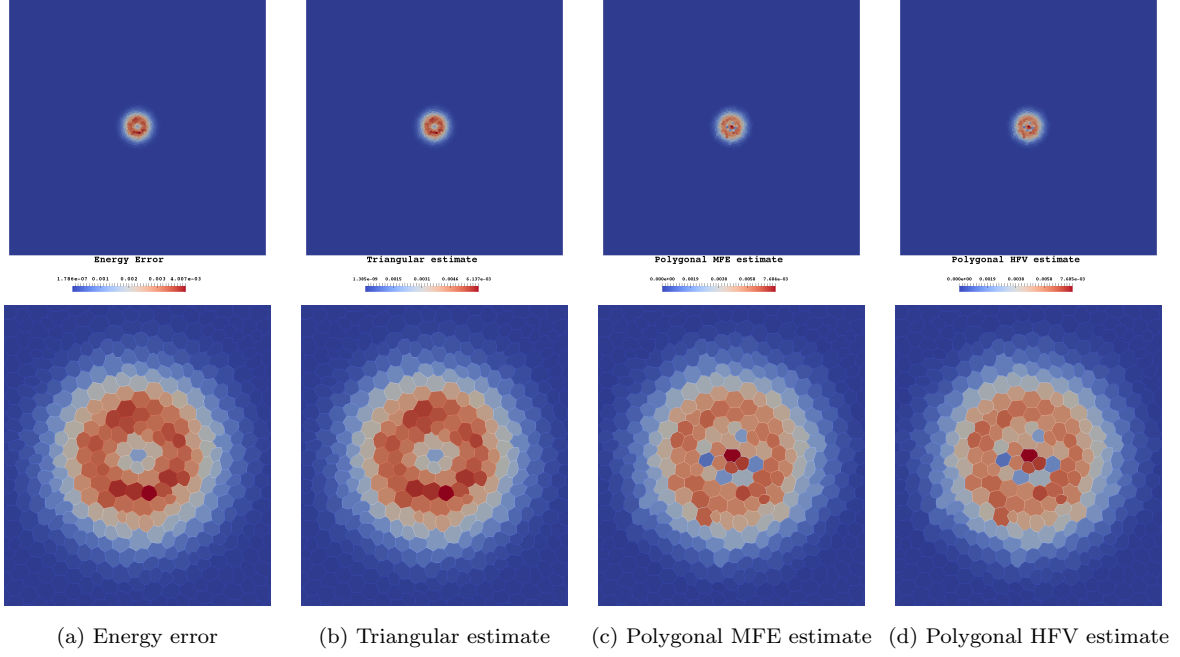(a) Energy error (b) Triangular estimate (c) Polygonal MFE estimate (d) Polygonal HFV estimate

Figure 20: Actual and estimated error distributions, entire domain (*top*) and center zoom (*bottom*)
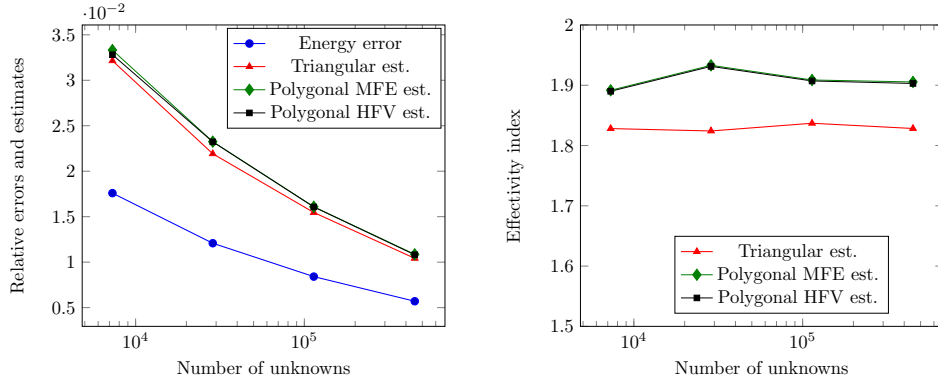
Figure 21: Relative error $\frac{\|\boldsymbol{u}-\boldsymbol{u}_h\|}{\|\boldsymbol{u}_h\|}$ and relative estimators and $\frac{\eta}{\|\boldsymbol{u}_h\|}$ (left) and effectivity indices (right), uniform mesh refinement

In Figure 20 we compare the actual and predicted error distributions. The energy error and the sharp triangular estimate distributions match perfectly. The polygonal MFE estimate and the polygonal HFV estimates give similar results and match also well with the energy error, though the error distribution around the peak is less well captured. We depict in Figure 21 the error and estimates as a function of the total number of unknowns and the corresponding effectivity indices for a uniform mesh refinement. This is in this test performed as follows: we refine uniformly the triangular submesh $\mathcal{T}_h$, giving actually rise to a Delaunay triangulation on each step, and then we merge triangles into polygons. All the three estimators behave in a similar way, with a slight advantage for the triangular estimate. The graphs confirm in particular that replacing the mixed finite element matrix $\widehat{\mathbb{A}}_{\mathrm{MFE},K}$ in (5.37) by $\widehat{\mathbb{A}}_K$ in (5.41) has a very small influence.

Figure 22 shows the results for adaptive mesh refinement, achieved by using the local distribution of the predicted error as an indicator to refine only the cells of the mesh where the error is important. More precisely, we refine the cells $K \in \mathcal{T}_H$ such that $\eta_K \geq 0.7 \max_{L \in \mathcal{T}_H} \eta_L$. We observe quasi-identical values of the polygonal MFE and polygonal HFV estimates; consequently, we obtain the same number of unknowns at each step of adaptivity and almost identical final energy error for these two estimators. The left part of Figure 22 displays the results where we present the estimators and two energy errors, one resulting with adaptivity based on the polygonal estimates and another with adaptivity based on the triangular estimates. Almost undistinguishable results are obtained. Finally, adaptive mesh refinement
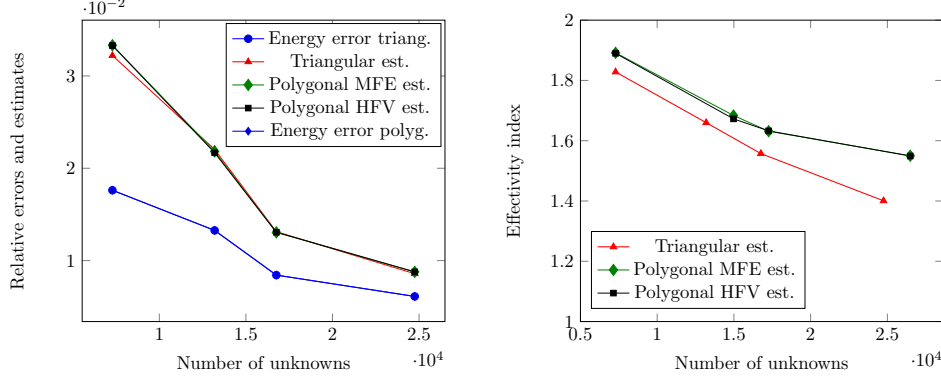
41

Figure 22: Relative error $\frac{\|\boldsymbol{u}-\boldsymbol{u}_h\|}{\|\boldsymbol{u}_h\|}$ and relative estimators and $\frac{\eta}{\|\boldsymbol{u}_h\|}$ (left) and effectivity indices (right), adaptive mesh refinement

leads to both smaller error and better effectivity indices in comparison with uniform mesh refinement.

5.11 Bibliographic resources

Literature on a posteriori analysis on polytopal meshes is much less plentiful than that discussed in Section 4.12 for simplicial meshes. Beirão da Veiga [42] and Beirão da Veiga and Manzini [43] derive a posteriori error estimates for low-order mimetic finite difference methods; extensions are presented in Antonietti *et al.* [22]. Omnes *et al.* [206] considered the discrete duality finite volume method. Recent contributions on the subject include Munar *et al.* [193] or Li *et al.* [180], see also the references therein. Here we have followed Vohralík and Yousef [255], developing the idea of the local Neumann problems (5.14) from Vohralík [249, Section 3.2]. Numerous additional results were obtained for more complex model problems, which we discuss in Section 7.12 below.

6 Steady nonlinear pure diffusion problems. Iterative linearization and algebraic errors

As a passage between the model steady linear Darcy flow of Section 5 and the complex multiphase compositional Darcy flow of Section 7, we consider here a steady nonlinear Darcy flow. We showcase that our methodology applies here in a similar simple fashion as in Section 5. We in particular give all the details how inexact linearization and linear solvers can be taken into account, following the concept in Ern and Vohralík [124], giving rise to a simple *adaptive inexact Newton method* on *polytopal meshes*. Its particularity is that it gives a guaranteed upper bound on the total error in the fluxes on *each linearization* and *each algebraic* solver *step*, distinguishes the *different error components*, and is obtained by *simple multiplications* of the same element matrices as in the linear case by vectors of face normal fluxes and potential values immediately available in each lowest-order locally conservative polytopal scheme.

6.1 Singlephase steady nonlinear Darcy flow

For simplicity of exposition, let us consider the following quasi-linear version of (5.1): find $p : \Omega \to \mathbb{R}$ such that

$$-\nabla\cdot(\underline{\boldsymbol{K}}(|\nabla p|)\nabla p) = f \qquad \text{in } \Omega, \tag{6.1a}$$
$$p = 0 \qquad \text{on } \partial\Omega. \tag{6.1b}$$

We suppose that the nonlinearity can be inverted in the sense that

$$\boldsymbol{v} = -\underline{\boldsymbol{K}}(|\boldsymbol{w}|)\boldsymbol{w} \quad \Longleftrightarrow \quad \boldsymbol{w} = -\tilde{\underline{\boldsymbol{K}}}(|\boldsymbol{v}|)\boldsymbol{v} \tag{6.2}$$

for all $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^d$. We suppose *strong monotonicity* and *Lipschitz-continuity*, i.e., that there exist two positive constants $c_{\underline{\boldsymbol{K}}}, C_{\underline{\boldsymbol{K}}}$ so that for all $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^d$,

$$c_{\tilde{\boldsymbol{K}}}|\boldsymbol{v} - \boldsymbol{w}|^2 \le (\boldsymbol{v} - \boldsymbol{w})\cdot(\tilde{\underline{\boldsymbol{K}}}(|\boldsymbol{v}|)\boldsymbol{v} - \tilde{\underline{\boldsymbol{K}}}(|\boldsymbol{w}|)\boldsymbol{w}), \tag{6.3a}$$
$$|\tilde{\underline{\boldsymbol{K}}}(|\boldsymbol{v}|)\boldsymbol{v} - \tilde{\underline{\boldsymbol{K}}}(|\boldsymbol{w}|)\boldsymbol{w}| \le C_{\tilde{\boldsymbol{K}}}|\boldsymbol{v} - \boldsymbol{w}|. \tag{6.3b}$$
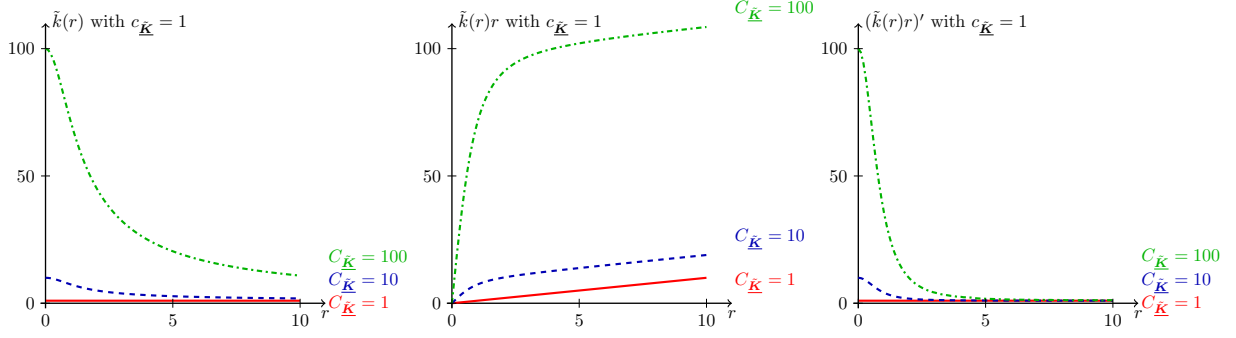
42

Figure 23: Prototypical monotonous and Lipschitz-continuous function $\tilde{k}(r) := c_{\tilde{\underline{K}}} + \frac{C_{\tilde{\underline{K}}} - c_{\tilde{\underline{K}}}}{\sqrt{1+r^2}}$ for $c_{\underline{\tilde{K}}} = 1$ and different values of $C_{\tilde{\underline{K}}}$ (left), $\tilde{k}(r)r$ (middle), and $(\tilde{k}(r)r)'$ (right)

Moreover, we let $\underline{K}$ and $\tilde{\underline{K}}$ take symmetric values and, namely to arrive at a simple matrix-vector multiplication form of the estimates, we also suppose that, for all $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^d$,

$$c_{\tilde{\underline{K}}}|\boldsymbol{v}|^2 \le \boldsymbol{v} \cdot \tilde{\underline{K}}(|\boldsymbol{w}|)\boldsymbol{v}, \qquad |\tilde{\underline{K}}(|\boldsymbol{w}|)\boldsymbol{v}| \le C_{\tilde{\underline{K}}}|\boldsymbol{v}|. \tag{6.4}$$

Note that for a linear problem, $\tilde{\underline{K}} = \underline{K}^{-1}$, $c_{\tilde{\underline{K}}}$ is simply the smallest eigenvalue of $\tilde{\underline{K}}$, i.e., the reciprocal of the largest eigenvalue of $\underline{K}$, and $C_{\tilde{\underline{K}}}$ is the largest eigenvalue of $\tilde{\underline{K}}$, i.e., the reciprocal of the smallest eigenvalue of $\underline{K}$.

A prototypical example satisfying (6.3)–(6.4) is, for two positive constant $c_{\tilde{\underline{K}}}$, $C_{\tilde{\underline{K}}}$,

$$\tilde{k}(r) := c_{\tilde{\underline{K}}} + \frac{C_{\tilde{\underline{K}}} - c_{\tilde{\underline{K}}}}{\sqrt{1+r^2}}, \qquad \tilde{\underline{K}} := \tilde{k}\underline{I}. \tag{6.5}$$

In this case, it can be easily verified that

$$c_{\tilde{\underline{K}}} \le \tilde{k}(r) \le C_{\tilde{\underline{K}}}, \qquad c_{\tilde{\underline{K}}} \le (\tilde{k}(r)r)' \le C_{\tilde{\underline{K}}}.$$

An illustration is provided in Figure 23. A much wider class of nonlinearities is considered below in Section 7.1.

6.2 Weak solution and its properties

In a generalization of Section 5.2, the weak formulation of (6.1) looks for $p \in H_0^1(\Omega)$ such that

$$(\underline{K}(|\nabla p|)\nabla p, \nabla v) = (f, v) \qquad \forall v \in H_0^1(\Omega). \tag{6.6}$$

From the pressure head p, the Darcy velocity is given by

$$\boldsymbol{u} := -\underline{K}(|\nabla p|)\nabla p, \tag{6.7}$$

and

$$\nabla p = -\tilde{\underline{K}}(|\boldsymbol{u}|)\boldsymbol{u} \tag{6.8}$$

from (6.2). Still as in Proposition 4.1, it follows from (6.6) that

$$\boldsymbol{u} \in \boldsymbol{H}(\mathrm{div}, \Omega), \quad \nabla \cdot \boldsymbol{u} = f. \tag{6.9}$$

6.3 Total error and its spatial discretization, linearization, and algebraic components

In the Poisson model problem in Section 4, we measure the error as $\|\boldsymbol{u} - \boldsymbol{u}_h\|$, see (4.5). In the Darcy problem in Section 5, the intrinsic error measure becomes $\|\boldsymbol{u} - \boldsymbol{u}_h\|_{\underline{K}^{-\frac{1}{2}}}$, see (5.5). From the mathematical setting in Section 6.1, a natural extension is here to consider

$$c_{\tilde{\underline{K}}}^{1/2}\|\boldsymbol{u} - \boldsymbol{u}_h^{k,i}\| \tag{6.10}$$

43

as the error, where the weight $c_{\tilde{\boldsymbol{K}}}$ in front of the $L^2(\Omega)$-norm is the monotonicity constant from (6.3a).

In (6.10), we have denoted the approximate solution by $\boldsymbol{u}_h^{k,i}$, anticipating that our numerical approximation will here consist in *spatial discretization* (with characteristic mesh size h), *iterative linearization* (with iteration index k), and *iterative algebraic resolution* (with iteration index i). Our goal will be to *distinguish the error components* as in (2.14) in Section 2.4.6 and to design an a posteriori error estimate of the form (2.23a), following the general concepts outlined in Section 2.5.

6.4 Generic discretizations on a polytopal mesh, iterative linearization, and iterative algebraic system solution

Proceeding as generally as possible, we accommodate any discretization of the form of Assumption 5.1. Let $\mathsf{F}_K := (f, 1)_K$ for all $K \in \mathcal{T}_H$, as in (5.7). For the model nonlinear problem (6.1), the flux balance (5.8) now takes the form: find the algebraic vector $\mathsf{P} := \{\mathsf{P}_K\}_{K \in \mathcal{T}_H} \in \mathbb{R}^{|\mathcal{T}_H|}$ such that

$$\boxed{\sum_{\sigma \in \mathcal{F}_K} (\mathsf{U}(\mathsf{P}))_\sigma \boldsymbol{n}_K \cdot \boldsymbol{n}_\sigma = \mathsf{F}_K \quad \forall K \in \mathcal{T}_H.} \tag{6.11}$$

Here $(\mathsf{U}(\mathsf{P}))_\sigma \in \mathbb{R}$ for each face $\sigma \in \mathcal{F}_H$ again approximates the normal flux $\langle \boldsymbol{u} \cdot \boldsymbol{n}_\sigma, 1 \rangle_\sigma$ over the face σ, but it depends in a *nonlinear way* on the pressures P, which is denoted by $\mathsf{U}(\mathsf{P})$. It follows that (6.11) is a system of $|\mathcal{T}_H|$ *nonlinear algebraic equations* for the $|\mathcal{T}_H|$ unknowns P.

The solution to (6.11) is typically sought by an *iterative linearization* such as the Newton or the fixed point method. Given an initial guess P^0, on each linearization step $k \geq 1$, this leads to a system of *linear algebraic equations* of the form: find the algebraic vector $\mathsf{P}^k := \{\mathsf{P}_K^k\}_{K \in \mathcal{T}_H} \in \mathbb{R}^{|\mathcal{T}_H|}$ such that

$$\boxed{\sum_{\sigma \in \mathcal{F}_K} (\mathsf{U}^{k-1}(\mathsf{P}^k))_\sigma \boldsymbol{n}_K \cdot \boldsymbol{n}_\sigma = \mathsf{F}_K \quad \forall K \in \mathcal{T}_H.} \tag{6.12}$$

Here, $\mathsf{U}^{k-1}(\mathsf{P}^k)$ are linearized face normal fluxes. These are obtained from the previous iterate P^{k-1} and $\mathsf{U}(\mathsf{P}^{k-1})$ but depend in a linear (affine) way on P^k. For the Newton linearization, in particular,

$$(\mathsf{U}^{k-1}(\mathsf{P}^k))_\sigma := \sum_{K' \in \mathcal{T}_H} \frac{\partial (\mathsf{U}(\mathsf{P}^{k-1}))_\sigma}{\partial \mathsf{P}_{K'}} \cdot \left(\mathsf{P}_{K'}^k - \mathsf{P}_{K'}^{k-1} \right) + (\mathsf{U}(\mathsf{P}^{k-1}))_\sigma. \tag{6.13}$$

In Section 7 below, (7.27) provides a general Newton linearization for a more involved problem.

The solution to (6.12) is typically sought by an *iterative algebraic solver*. Given an initial guess $\mathsf{P}^{k,0}$, typically the last algebraic solver iterate on the previous linearization step $k-1$ when $k > 1$, on each algebraic solver step $i \geq 1$, the constant

$$(\mathsf{R})_K^{k,i} := \mathsf{F}_K - \sum_{\sigma \in \mathcal{F}_K} (\mathsf{U}^{k-1}(\mathsf{P}^{k,i}))_\sigma \boldsymbol{n}_K \cdot \boldsymbol{n}_\sigma \tag{6.14}$$

is the *algebraic residual* on the element K, the misfit of linearized fluxes $\mathsf{U}^{k-1}(\mathsf{P}^{k,i})$ not to satisfy (6.12). Consequently, the flux balance on linearization step k and algebraic step i takes the form

$$\boxed{\sum_{\sigma \in \mathcal{F}_K} (\mathsf{U}^{k-1}(\mathsf{P}^{k,i}))_\sigma \boldsymbol{n}_K \cdot \boldsymbol{n}_\sigma = \mathsf{F}_K - (\mathsf{R})_K^{k,i} \quad \forall K \in \mathcal{T}_H,} \tag{6.15}$$

i.e., (6.12) up to the algebraic residual. In order to estimate the algebraic error, we will, following [140, 154, 124], also employ $j \geq 1$ additional algebraic solver steps in (6.15), giving rise to

$$\sum_{\sigma \in \mathcal{F}_K} (\mathsf{U}^{k-1}(\mathsf{P}^{k,i+j}))_\sigma \boldsymbol{n}_K \cdot \boldsymbol{n}_\sigma = \mathsf{F}_K - (\mathsf{R})_K^{k,i+j} \quad \forall K \in \mathcal{T}_H. \tag{6.16}$$

The motivation is to take j large enough to make the algebraic residual $(\mathsf{R})_K^{k,i+j}$ considerably smaller than $(\mathsf{R})_K^{k,i}$, see Remark 6.3 below.

Example 6.1 (Discretization of problem (6.1) by the finite volume scheme on a simplicial mesh with fixed-point linearization and an arbitrary iterative algebraic solver). *Consider the setting and notation*

of Section 4.7. The cell-centered finite volume scheme for the nonlinear Darcy problem (6.1) on the simplicial mesh $\mathcal{T}_h$ reads: find the real values p_K, $K \in \mathcal{T}_h$, the approximations to the mean values of p in the mesh elements K, such that

$$\sum_{\sigma \in \mathcal{F}_K} \xi_{K,\sigma} U_{K,\sigma} = (f, 1)_K \qquad \forall K \in \mathcal{T}_h, \tag{6.17a}$$

where the real number $\xi_{K,\sigma} U_{K,\sigma}$ for each face $\sigma \in \mathcal{F}_K$ approximates the normal (out)flux $\langle \boldsymbol{u} \cdot \boldsymbol{n}_K, 1 \rangle_\sigma = \langle (-\underline{\boldsymbol{K}}(|\nabla p|)\nabla p) \cdot \boldsymbol{n}_K, 1 \rangle_\sigma$ from K over the face σ. This is composed of the approximation of $\langle -\nabla p \cdot \boldsymbol{n}_K, 1 \rangle_\sigma$ given, as in (4.19b)–(4.19c), by

$$U_{K,\sigma} := -\frac{|\sigma_{K,L}|}{d_{K,L}}(p_L - p_K) \qquad \sigma = \sigma_{K,L} \in \mathcal{F}_h^{\text{int}}, \tag{6.17b}$$

$$U_{K,\sigma} := -\frac{|\sigma|}{d_{K,\sigma}}(0 - p_K) \qquad \sigma \in \mathcal{F}_K \cap \mathcal{F}_h^{\text{ext}} \tag{6.17c}$$

and of the approximation of $\underline{\boldsymbol{K}}(|\nabla p|)$ on σ given by

$$\xi_{K,\sigma} := \underline{\boldsymbol{K}}\left(\left| \frac{1}{2}(\boldsymbol{\xi}_h|_K(\boldsymbol{x}_{K,\sigma}) + \boldsymbol{\xi}_h|_L(\boldsymbol{x}_{K,\sigma})) \right| \right) \qquad \sigma = \sigma_{K,L} \in \mathcal{F}_h^{\text{int}}, \tag{6.17d}$$

$$\xi_{K,\sigma} := \underline{\boldsymbol{K}}(|\boldsymbol{\xi}_h|_K(\boldsymbol{x}_{K,\sigma})|) \qquad \sigma \in \mathcal{F}_K \cap \mathcal{F}_h^{\text{ext}}, \tag{6.17e}$$

where the approximation of $-\nabla p$ is, as in (4.22), given by

$$(\boldsymbol{\xi}_h|_K)(\boldsymbol{x}) := \sum_{\sigma \in \mathcal{F}_K} U_{K,\sigma} \frac{1}{d|K|}(\boldsymbol{x} - \boldsymbol{a}_{K,\sigma}), \qquad \boldsymbol{x} \in K, K \in \mathcal{T}_h, \tag{6.17f}$$

see the illustrations in Figure 17. Note that in contrast to (4.19b)–(4.19c), the face fluxes $\xi_{K,\sigma} U_{K,\sigma}$ are now nonlinear *functions of the pressure unknowns p_K and p_L (unless $\underline{\boldsymbol{K}}(\cdot)$ is a constant function). Also note that (6.17) writes in the form (6.11) with $(\mathsf{U}(\mathsf{P}))_\sigma = \xi_{K,\sigma} U_{K,\sigma} \boldsymbol{n}_K \cdot \boldsymbol{n}_\sigma$ (recall that $\boldsymbol{n}_K \cdot \boldsymbol{n}_\sigma = \pm 1$ only determines the sign).*

Iterative linearization of (6.17) by the fixed point method then reads: given the initial values p_K^0, $K \in \mathcal{T}_h$, on each linearization step $k \geq 1$, find the real values p_K^k, $K \in \mathcal{T}_h$, such that

$$\sum_{\sigma \in \mathcal{F}_K} \xi_{K,\sigma}^{k-1} U_{K,\sigma}^k = (f, 1)_K \qquad \forall K \in \mathcal{T}_h, \tag{6.18a}$$

where

$$U_{K,\sigma}^k := -\frac{|\sigma_{K,L}|}{d_{K,L}}(p_L^k - p_K^k) \qquad \sigma = \sigma_{K,L} \in \mathcal{F}_h^{\text{int}}, \tag{6.18b}$$

$$U_{K,\sigma}^k := -\frac{|\sigma|}{d_{K,\sigma}}(0 - p_K^k) \qquad \sigma \in \mathcal{F}_K \cap \mathcal{F}_h^{\text{ext}}, \tag{6.18c}$$

$$\xi_{K,\sigma}^{k-1} := \underline{\boldsymbol{K}}\left(\left| \frac{1}{2}(\boldsymbol{\xi}_h^{k-1}|_K(\boldsymbol{x}_{K,\sigma}) + \boldsymbol{\xi}_h^{k-1}|_L(\boldsymbol{x}_{K,\sigma})) \right| \right) \qquad \sigma = \sigma_{K,L} \in \mathcal{F}_h^{\text{int}}, \tag{6.18d}$$

$$\xi_{K,\sigma}^{k-1} := \underline{\boldsymbol{K}}(|\boldsymbol{\xi}_h^{k-1}|_K(\boldsymbol{x}_{K,\sigma})|) \qquad \sigma \in \mathcal{F}_K \cap \mathcal{F}_h^{\text{ext}}, \tag{6.18e}$$

$$(\boldsymbol{\xi}_h^{k-1}|_K)(\boldsymbol{x}) := \sum_{\sigma \in \mathcal{F}_K} U_{K,\sigma}^{k-1} \frac{1}{d|K|}(\boldsymbol{x} - \boldsymbol{a}_{K,\sigma}), \qquad \boldsymbol{x} \in K, K \in \mathcal{T}_h. \tag{6.18f}$$

We see that (6.18) takes the form requested in (6.12).

Finally, (6.15) for an arbitrary iterative algebraic solver follows by the definition (6.14).

6.5   Face normal fluxes and potential point values

The key for the guaranteed a posteriori error estimate with inexpensive implementation and evaluation for the linear Darcy problem (5.1) in Theorem 5.18 were the face normal fluxes of Section 5.5 and the potential point values of Section 5.7. We will now identify the face normal fluxes and potential point values for the present nonlinear Darcy problem (6.1). In order to devise a guaranteed a posteriori error

estimate distinguishing the error components, we will also identify *linearization error* and *algebraic error* face fluxes.

Let a polytopal cell $K \in \mathcal{T}_H$ with its face $\sigma \in \mathcal{F}_K$ be fixed. For a given linearization step $k \geq 1$, algebraic step $i \geq 1$, and $j \geq 1$ additional algebraic iterations, we set

$$(\mathsf{U}_K^{k,i})_\sigma := (\mathsf{U}(\mathsf{P}^{k,i}))_\sigma, \tag{6.19a}$$

$$(\mathsf{U}_{\mathrm{lin},K}^{k,i})_\sigma := (\mathsf{U}^{k-1}(\mathsf{P}^{k,i}))_\sigma - (\mathsf{U}(\mathsf{P}^{k,i}))_\sigma, \tag{6.19b}$$

$$(\mathsf{U}_{\mathrm{alg},K}^{k,i})_\sigma := (\mathsf{U}^{k-1}(\mathsf{P}^{k,i+j}))_\sigma - (\mathsf{U}^{k-1}(\mathsf{P}^{k,i}))_\sigma. \tag{6.19c}$$

Observe that $\mathsf{U}^{k-1}(\mathsf{P}^{k,i+j})$ are directly available as employed in (6.16) and that $\mathsf{U}^{k-1}(\mathsf{P}^{k,i})$ are directly available as employed in (6.15). In turn, $\mathsf{U}(\mathsf{P}^{k,i})$ are available upon plugging the current approximation $\mathsf{P}^{k,i}$ into the nonlinear flux function of (6.11). We call $\mathsf{U}_K^{k,i}$ *discretization face fluxes* since in limit $k, i \to \infty$, they tend to $\mathsf{U}(\mathsf{P})$ of (6.11). Similarly, we call $\mathsf{U}_{\mathrm{lin},K}^{k,i}$ *linearization error face fluxes* since they vanish with linearization and algebraic solver iterations $k, i \to \infty$. Finally, we call $\mathsf{U}_{\mathrm{alg},K}^{k,i}$ the *algebraic error face fluxes* since they vanish with algebraic solver iterations $i \to \infty$. Observe that

$$(\mathsf{U}_K^{k,i})_\sigma + (\mathsf{U}_{\mathrm{lin},K}^{k,i})_\sigma + (\mathsf{U}_{\mathrm{alg},K}^{k,i})_\sigma = (\mathsf{U}^{k-1}(\mathsf{P}^{k,i+j}))_\sigma,$$

so that from (6.16)

$$\sum_{\sigma \in \mathcal{F}_K} \left((\mathsf{U}_K^{k,i})_\sigma + (\mathsf{U}_{\mathrm{lin},K}^{k,i})_\sigma + (\mathsf{U}_{\mathrm{alg},K}^{k,i})_\sigma \right) \boldsymbol{n}_K \cdot \boldsymbol{n}_\sigma = \mathsf{F}_K - (\mathsf{R})_K^{k,i+j} \quad \forall K \in \mathcal{T}_H. \tag{6.20}$$

The fluxes per simplex face are then obtained as in Definition 5.3.

The potential point values are obtained as in Definition 5.11 or 5.13 directly from the cell pressure heads $\mathsf{P}^{k,i}$. This gives the vectors $\mathsf{Z}_K^{k,i}$ and $\mathsf{Z}_K^{k,i,\mathrm{ext}}$ of *potential point values* in the vertices of the simplicial submesh $\mathcal{T}_K$ of each polytopal element $K \in \mathcal{T}_H$.

6.6 Fictitious flux and potential reconstructions

We now extend the reconstructions of Sections 5.6 and 5.8 to the present nonlinear setting.

We have to face the fact that the diffusion tensor $\underline{\boldsymbol{K}}(|\nabla p|)$ is now unknown since p is unknown. We thus modify Definition 5.5 to

$$\boldsymbol{u}_h^{k,i}|_K := c_{\underline{\boldsymbol{K}}}^{-1} C_{\underline{\boldsymbol{K}}}^2 \arg \min_{\boldsymbol{v}_h \in \boldsymbol{V}_{h,\mathrm{N}}^K, \, \nabla \cdot \boldsymbol{v}_h = \mathrm{constant}} \|\boldsymbol{v}_h\|_K^2, \tag{6.21}$$

where the data in the spaces $\boldsymbol{V}_{h,\mathrm{N}}^K$ are now given by $\mathsf{U}_K^{k,i}$ from (6.19a) in place of U_σ (we systematically use the convention (5.12) (without change of the notation) to define the fluxes of the simplicial submesh faces in case they subdivide the original polytopal faces). Note that we cannot in general impose here $\nabla \cdot (\boldsymbol{u}_h^{k,i})|_K = f|_K$. This *discretization flux reconstruction* is again not needed to evaluate the a posteriori error estimators of Theorem 6.2 below (it would, however, be needed to evaluate the error $c_{\underline{\boldsymbol{K}}}^{1/2} \|\boldsymbol{u} - \boldsymbol{u}_h^{k,i}\|$.) For appropriate scaling, we have replaced $\underline{\boldsymbol{K}}^{-1}$ by $c_{\underline{\boldsymbol{K}}}^{-1} C_{\underline{\boldsymbol{K}}}^2 \underline{\boldsymbol{I}}$. It is important to stress that just like in (5.16), problem (6.21) is *linear* and can be written in a matrix form as in (5.18), with the *same matrices*, just replacing $\underline{\boldsymbol{K}}^{-1}$ by $c_{\underline{\boldsymbol{K}}}^{-1} C_{\underline{\boldsymbol{K}}}^2 \underline{\boldsymbol{I}}$. This in particular leads to the definition of the MFE element matrix $\widehat{\mathbb{A}}_{\mathrm{MFE},K}$ by (5.20) also in the present nonlinear setting, up to the replacement of $\underline{\boldsymbol{K}}^{-1}$ by $c_{\underline{\boldsymbol{K}}}^{-1} C_{\underline{\boldsymbol{K}}}^2 \underline{\boldsymbol{I}}$.

Similarly, we define the linearization error flux reconstruction $\boldsymbol{u}_{\mathrm{lin},h}^{k,i}$ by the lifting (6.21), employing $\mathsf{U}_{\mathrm{lin},K}^{k,i}$ from (6.19b) in the definition of the space $\boldsymbol{V}_{h,\mathrm{N}}^K$, and the algebraic error flux reconstruction $\boldsymbol{u}_{\mathrm{alg},h}^{k,i}$ by the lifting (6.21), employing $\mathsf{U}_{\mathrm{alg},K}^{k,i}$ from (6.19c) in the definition of the space $\boldsymbol{V}_{h,\mathrm{N}}^K$. The flux reconstructions $\boldsymbol{u}_{\mathrm{lin},h}^{k,i}$ and $\boldsymbol{u}_{\mathrm{alg},h}^{k,i}$ will only be used as a theoretical vehicle below. From (6.20), we observe by the Green theorem that

$$\left(\nabla \cdot (\boldsymbol{u}_h^{k,i} + \boldsymbol{u}_{\mathrm{lin},h}^{k,i} + \boldsymbol{u}_{\mathrm{alg},h}^{k,i}) \right)|_K = f|_K - |K|^{-1}(\mathsf{R})_K^{k,i+j} \quad \forall K \in \mathcal{T}_H. \tag{6.22}$$

Finally, the *potential reconstruction* $\zeta_h^{k,i}$ is simply obtained from the element vectors $\mathsf{Z}_K^{k,i}$ and $\mathsf{Z}_K^{k,i,\mathrm{ext}}$ assembled in Section 6.5 as in Definition 5.14, i.e., setting

$$\zeta_h^{k,i}(\boldsymbol{a}) := \mathsf{Z}_{\boldsymbol{a}}^{k,i} \qquad \forall \boldsymbol{a} \in \mathcal{V}_h.$$

Congruently, for any $K \in \mathcal{T}_H$, we define the stiffness matrix $\widehat{\mathbb{S}}_{\mathrm{FE},K}$ as in (5.32), with $c_{\tilde{\underline{K}}}^{-2} C_{\tilde{\underline{K}}} \underline{I}$ in place of $\underline{K}$, whereas the mass matrix $\widehat{\mathbb{M}}_{\mathrm{FE},K}$ is defined exactly as in (5.33).

6.7 A guaranteed a posteriori error estimate distinguishing the error components with inexpensive implementation and evaluation

Let C_{F} be the constant from the Friedrichs inequality

$$\|v\| \le C_{\mathrm{F}} h_\Omega \|\nabla v\| \qquad \forall v \in H_0^1(\Omega); \tag{6.23}$$

there holds $1/(\pi d) \le C_{\mathrm{F}} \le 1$, as $C_{\mathrm{F}} h_\Omega$ is the square root of the reciprocal of the smallest eigenvalue of the Laplace operator on Ω with homogenous Dirichlet boundary condition. The main result of this section is:

Theorem 6.2 (A guaranteed a posteriori error estimate distinguishing the error components with inexpensive implementation and evaluation)**.**

Let f be constant on each $K \in \mathcal{T}_H$ and let $\boldsymbol{u}$ be given by (6.6)–(6.7). Consider any polytopal discretization of the form (6.11), any iterative linearization (6.12) on step $k \ge 1$, and any iterative algebraic solver (6.15) on step $i \ge 1$. Consider $j \ge 1$ additional algebraic solver steps leading to (6.16). Let, for each polytopal element $K \in \mathcal{T}_H$, the element vectors of face normal fluxes $\mathsf{U}_K^{k,i}$, $\mathsf{U}_{\mathrm{lin},K}^{k,i}$, and $\mathsf{U}_{\mathrm{alg},K}^{k,i}$ be given by (6.19). Let the element vectors of potential point values $\mathsf{Z}_K^{k,i}$ and $\mathsf{Z}_K^{k,i,\mathrm{ext}}$ be constructed from the values $\mathsf{P}^{k,i}$ following Section 6.5. Let finally the element matrices $\widehat{\mathbb{A}}_{\mathrm{MFE},K}$ and $\widehat{\mathbb{S}}_{\mathrm{FE},K}$ be defined by (5.20) and (5.32), where one respectively takes the multiples of the identity matrix $c_{\tilde{\underline{K}}}^{-1} C_{\tilde{\underline{K}}}^2 \underline{I}$ in place of $\underline{K}^{-1}$ and $c_{\tilde{\underline{K}}}^{-2} C_{\tilde{\underline{K}}} \underline{I}$ in place of $\underline{K}$. Let also $\widehat{\mathbb{M}}_{\mathrm{FE},K}$ be given by (5.33). Then there holds

$$c_{\tilde{\underline{K}}}^{1/2} \|\boldsymbol{u} - \boldsymbol{u}_h^{k,i}\| \le \eta_{\mathrm{sp}}^{k,i} + \eta_{\mathrm{lin}}^{k,i} + \eta_{\mathrm{alg}}^{k,i} + \eta_{\mathrm{rem}}^{k,i} \tag{6.24}$$

with

$$\eta_\bullet^{k,i} = \left\{ \sum_{K \in \mathcal{T}_H} \left(\eta_{\bullet,K}^{k,i} \right)^2 \right\}^{1/2}, \qquad \bullet = \{\mathrm{sp}, \mathrm{lin}, \mathrm{alg}, \mathrm{rem}\},$$

where the (spatial) *discretization estimators* are given by

$$\left(\eta_{\mathrm{sp},K}^{k,i} \right)^2 := (\mathsf{U}_K^{k,i})^{\mathrm{t}} \widehat{\mathbb{A}}_{\mathrm{MFE},K} \mathsf{U}_K^{k,i} + (\mathsf{Z}_K^{k,i})^{\mathrm{t}} \widehat{\mathbb{S}}_{\mathrm{FE},K} \mathsf{Z}_K^{k,i} + 2 c_{\tilde{\underline{K}}}^{-1} C_{\tilde{\underline{K}}} \left[(\mathsf{U}_K^{k,i,\mathrm{ext}})^{\mathrm{t}} \mathsf{Z}_K^{k,i,\mathrm{ext}} - \mathsf{F}_K |K|^{-1} \mathsf{1}^{\mathrm{t}} \widehat{\mathbb{M}}_{\mathrm{FE},K} \mathsf{Z}_K^{k,i} \right],$$

the *linearization estimators* by

$$\left(\eta_{\mathrm{lin},K}^{k,i} \right)^2 := (\mathsf{U}_{\mathrm{lin},K}^{k,i})^{\mathrm{t}} \widehat{\mathbb{A}}_{\mathrm{MFE},K} \mathsf{U}_{\mathrm{lin},K}^{k,i},$$

the *algebraic estimators* by,

$$\left(\eta_{\mathrm{alg},K}^{k,i} \right)^2 := (\mathsf{U}_{\mathrm{alg},K}^{k,i})^{\mathrm{t}} \widehat{\mathbb{A}}_{\mathrm{MFE},K} \mathsf{U}_{\mathrm{alg},K}^{k,i},$$

and the *algebraic remainder estimators* by

$$\eta_{\mathrm{rem},K}^{k,i} := c_{\tilde{\underline{K}}}^{-1/2} C_{\tilde{\underline{K}}} C_{\mathrm{F}} h_\Omega |K|^{-1/2} |(\mathsf{R})_K^{k,i+j}|.$$

Here, the flux reconstruction $\boldsymbol{u}_h^{k,i} \in \mathcal{RT}_0(\mathcal{T}_h) \cap \boldsymbol{H}(\mathrm{div}, \Omega)$ is obtained following (6.21).

Remark 6.3 (Algebraic remainder estimators $\eta_{\mathrm{rem},K}^{k,i}$, choice of j additional linear solver iterations, and mass balance)**.** *From* (6.20) *or* (6.22), *at "convergence" of the iterative algebraic solver (or for a direct solver)(recall that we neglect rounding errors), as the algebraic residuals* $(\mathsf{R})_K^{k,i+j}$ *vanish, we recover exact mass balance on each linearization step* $k \ge 1$. *More importantly, for a sufficiently high number* j *of additional algebraic iterations, we "almost" have the mass balance on each linearization step* $k \ge 1$. *We typically choose* j *adaptively, following* [140, 154, 124], *so that this mass balance misfit related to the algebraic residuals* $(\mathsf{R})_K^{k,i+j}$ *is negligible: we choose* j *such that* $\eta_{\mathrm{rem}}^{k,i}$ *becomes negligible in comparison*

with the other estimators in (6.24). The presence of the terms $(\mathsf{R})_K^{k,i+j}$ is the disadvantage of the j additional iterations. In turn, the advantage of the j of additional algebraic iterations is the simplicity of the formula (6.19c). A more involved choice can be made following [210]. Then no j additional iterations are needed, the term $|K|^{-1}(\mathsf{R})_K^{k,i+j}$ in (6.20) and (6.22) and the term $\eta_{\text{rem}}^{k,i}$ in (6.24) vanish, there is an exact mass balance on each linearization step $k \geq 1$ and each linear algebraic step $i \geq 1$, and the a posteriori bound (6.24) is typically slightly more precise. Remark finally that these two choices respectively correspond to nonzero and zero $\rho_h^{k,i}$ in the general discussion in Section 2.7.

Proof. The proof follows in spirit of previous works by Kim [163], Jiránek *et al.* [154], and Ern and Vohralík [124, 125]. Let $\zeta \in H_0^1(\Omega)$ be the solution of

$$(\underline{\boldsymbol{K}}(|\nabla\zeta|)\nabla\zeta, \nabla v) = -(\boldsymbol{u}_h^{k,i}, \nabla v) \qquad \forall v \in H_0^1(\Omega) \tag{6.25}$$

and set $\boldsymbol{\zeta} := -\underline{\boldsymbol{K}}(|\nabla\zeta|)\nabla\zeta$, similarly to (6.7). Remark that if there holds $\boldsymbol{u}_h^{k,i} \in \boldsymbol{H}(\text{div},\Omega)$ with $\nabla\cdot\boldsymbol{u}_h^{k,i} = f$ (if (6.11) is satisfied exactly, no iterative linearization and exact algebraic solve), then, by the Green theorem, $-(\boldsymbol{u}_h^{k,i}, \nabla v) = (f, v)$, so that ζ coincides with the solution p of (6.6).

By the triangle inequality,

$$c_{\underline{\tilde{\boldsymbol{K}}}}^{1/2}\|\boldsymbol{u} - \boldsymbol{u}_h^{k,i}\| \leq c_{\underline{\tilde{\boldsymbol{K}}}}^{1/2}\|\boldsymbol{u} - \boldsymbol{\zeta}\| + c_{\underline{\tilde{\boldsymbol{K}}}}^{1/2}\|\boldsymbol{\zeta} - \boldsymbol{u}_h^{k,i}\|. \tag{6.26}$$

For the first term above, the strong monotonicity implies

$$c_{\tilde{\boldsymbol{K}}}\|\boldsymbol{u} - \boldsymbol{\zeta}\|^2 \leq (\underline{\boldsymbol{K}}(|\nabla p|)\nabla p - \underline{\boldsymbol{K}}(|\nabla\zeta|)\nabla\zeta, \nabla(p-\zeta)) \qquad\qquad \text{by (6.3a) and (6.2)}$$

$$= (\underline{\boldsymbol{K}}(|\nabla p|)\nabla p + \boldsymbol{u}_h^{k,i}, \nabla(p-\zeta)) \qquad\qquad \text{by (6.25)}$$

$$= (f, p-\zeta) + (\boldsymbol{u}_h^{k,i}, \nabla(p-\zeta)). \qquad\qquad \text{by (6.6)}$$

Note that by the Lipschitz-continuity (6.3b) together with (6.2),

$$\|\nabla(p-\zeta)\| \leq C_{\tilde{\boldsymbol{K}}}\|\boldsymbol{u} - \boldsymbol{\zeta}\|. \tag{6.27}$$

Thus,

$$c_{\underline{\tilde{\boldsymbol{K}}}}^{1/2}\|\boldsymbol{u} - \boldsymbol{\zeta}\| = \frac{(f, p-\zeta) + (\boldsymbol{u}_h^{k,i}, \nabla(p-\zeta))}{c_{\underline{\tilde{\boldsymbol{K}}}}^{1/2}\|\boldsymbol{u} - \boldsymbol{\zeta}\|}$$

$$\leq c_{\underline{\tilde{\boldsymbol{K}}}}^{-1/2} C_{\tilde{\boldsymbol{K}}} \max_{\varphi \in H_0^1(\Omega),\, \|\nabla\varphi\|=1} \{(f, \varphi) + (\boldsymbol{u}_h^{k,i}, \nabla\varphi)\}. \tag{6.28}$$

To estimate (6.28) by a computable quantity, we use (6.22). Fix now $\varphi \in H_0^1(\Omega)$ with $\|\nabla\varphi\| = 1$. Then adding and subtracting $(\boldsymbol{u}_{\text{lin},h}^{k,i} + \boldsymbol{u}_{\text{alg},h}^{k,i}, \nabla\varphi)$ and using the Green theorem,

$$(f, \varphi) + (\boldsymbol{u}_h^{k,i}, \nabla\varphi) = (f - \nabla\cdot(\boldsymbol{u}_h^{k,i} + \boldsymbol{u}_{\text{lin},h}^{k,i} + \boldsymbol{u}_{\text{alg},h}^{k,i}), \varphi) - (\boldsymbol{u}_{\text{lin},h}^{k,i} + \boldsymbol{u}_{\text{alg},h}^{k,i}, \nabla\varphi)$$

$$= \sum_{K \in \mathcal{T}_H} \{|K|^{-1}((\mathsf{R})_K^{k,i+j}, \varphi)_K - (\boldsymbol{u}_{\text{lin},h}^{k,i} + \boldsymbol{u}_{\text{alg},h}^{k,i}, \nabla\varphi)_K\}.$$

The Cauchy–Schwarz inequality gives

$$\sum_{K \in \mathcal{T}_H} -(\boldsymbol{u}_{\text{lin},h}^{k,i} + \boldsymbol{u}_{\text{alg},h}^{k,i}, \nabla\varphi)_K \leq \sum_{K \in \mathcal{T}_H} \|\boldsymbol{u}_{\text{lin},h}^{k,i} + \boldsymbol{u}_{\text{alg},h}^{k,i}\|_K \|\nabla\varphi\|_K \leq \|\boldsymbol{u}_{\text{lin},h}^{k,i} + \boldsymbol{u}_{\text{alg},h}^{k,i}\| \underbrace{\|\nabla\varphi\|}_{=1}$$

$$\leq \|\boldsymbol{u}_{\text{lin},h}^{k,i}\| + \|\boldsymbol{u}_{\text{alg},h}^{k,i}\|.$$

Similarly, again the Cauchy–Schwarz inequality leads to

$$\sum_{K \in \mathcal{T}_H} |K|^{-1}|((\mathsf{R})_K^{k,i+j}, \varphi)_K| \leq \sum_{K \in \mathcal{T}_H} |K|^{-1/2}|(\mathsf{R})_K^{k,i+j}|\|\varphi\|_K \leq \left\{\sum_{K \in \mathcal{T}_H} |K|^{-1}|(\mathsf{R})_K^{k,i+j}|^2\right\}^{1/2} \|\varphi\|_\Omega. \tag{6.29}$$

Summing up these developments with the Friedrichs inequality (6.23), infer from (6.28)

$$c_{\underline{\tilde{\boldsymbol{K}}}}^{1/2}\|\boldsymbol{u} - \boldsymbol{\zeta}\| \leq c_{\underline{\tilde{\boldsymbol{K}}}}^{-1/2} C_{\tilde{\boldsymbol{K}}} \left[C_{\text{F}} h_\Omega \left\{\sum_{K \in \mathcal{T}_H} |K|^{-1}|(\mathsf{R})_K^{k,i+j}|^2\right\}^{1/2} + \|\boldsymbol{u}_{\text{lin},h}^{k,i}\| + \|\boldsymbol{u}_{\text{alg},h}^{k,i}\|\right].$$

Finally, to evaluate the norm $\|\boldsymbol{u}_{\mathrm{lin},h}^{k,i}\|$, we use

$$\|\boldsymbol{u}_{\mathrm{lin},h}^{k,i}\|^2 = \sum_{K \in \mathcal{T}_H} \|\boldsymbol{u}_{\mathrm{lin},h}^{k,i}\|_K^2$$

and employ the constant matrix $c_{\tilde{\underline{K}}}^{-1} C_{\tilde{\underline{K}}}^2 \boldsymbol{I}$ in place of $\underline{K}^{-1}$ in Lemma 5.8 to infer

$$c_{\tilde{\underline{K}}}^{-1} C_{\tilde{\underline{K}}}^2 \|\boldsymbol{u}_{\mathrm{lin},h}^{k,i}\|_K^2 = \big(\mathsf{U}_{\mathrm{lin},K}^{k,i}\big)^{\mathrm{t}} \widehat{\mathbb{A}}_{\mathrm{MFE},K} \mathsf{U}_{\mathrm{lin},K}^{k,i} \tag{6.30}$$

for each polytopal element $K \in \mathcal{T}_H$. We proceed similarly for $\|\boldsymbol{u}_{\mathrm{alg},h}^{k,i}\|$, which gives

$$c_{\tilde{\underline{K}}}^{-1} C_{\tilde{\underline{K}}}^2 \|\boldsymbol{u}_{\mathrm{alg},h}^{k,i}\|_K^2 = \big(\mathsf{U}_{\mathrm{alg},K}^{k,i}\big)^{\mathrm{t}} \widehat{\mathbb{A}}_{\mathrm{MFE},K} \mathsf{U}_{\mathrm{alg},K}^{k,i}. \tag{6.31}$$

We finally estimate the second term in (6.26). For $v \in H_0^1(\Omega)$ arbitrary, the strong monotonicity in particular leads to

$$
\begin{aligned}
c_{\tilde{\underline{K}}} \|\boldsymbol{\zeta} - \boldsymbol{u}_h^{k,i}\|^2 &\leq (\underline{K}(|\nabla\zeta|)\nabla\zeta + \boldsymbol{u}_h^{k,i}, \nabla\zeta + \tilde{\underline{K}}(|\boldsymbol{u}_h^{k,i}|)\boldsymbol{u}_h^{k,i}) && \text{by (6.3a) and (6.2)}\\
&= (\underline{K}(|\nabla\zeta|)\nabla\zeta + \boldsymbol{u}_h^{k,i}, \nabla v + \tilde{\underline{K}}(|\boldsymbol{u}_h^{k,i}|)\boldsymbol{u}_h^{k,i}) && \text{by (6.25)}\\
&\leq \|\boldsymbol{\zeta} - \boldsymbol{u}_h^{k,i}\| \|\tilde{\underline{K}}(|\boldsymbol{u}_h^{k,i}|)\boldsymbol{u}_h^{k,i} + \nabla v\| && \text{by Cauchy–Schwarz}
\end{aligned}
$$

so that,

$$c_{\tilde{\underline{K}}}^{1/2} \|\boldsymbol{\zeta} - \boldsymbol{u}_h^{k,i}\| \leq c_{\tilde{\underline{K}}}^{-1/2} \min_{v \in H_0^1(\Omega)} \|\tilde{\underline{K}}(|\boldsymbol{u}_h^{k,i}|)\boldsymbol{u}_h^{k,i} + \nabla v\| \tag{6.32a}$$

$$\leq c_{\tilde{\underline{K}}}^{-1/2} \|\tilde{\underline{K}}(|\boldsymbol{u}_h^{k,i}|)\boldsymbol{u}_h^{k,i} + \nabla\zeta_h^{k,i}\| \tag{6.32b}$$

for an arbitrary $\zeta_h^{k,i} \in H_0^1(\Omega)$. We pick the potential reconstruction from Section 6.6. For the inexpensive implementation and evaluation, we proceed as in Section 5.9. Consequently, for each polytopal mesh element $K \in \mathcal{T}_H$, employing here also the eigenvalue hypothesis (6.4),

$$
\begin{aligned}
&\|\tilde{\underline{K}}(|\boldsymbol{u}_h^{k,i}|)\boldsymbol{u}_h^{k,i} + \nabla\zeta_h^{k,i}\|_K^2 \\
&= \|[\tilde{\underline{K}}(|\boldsymbol{u}_h^{k,i}|)]^{1/2}\big([\tilde{\underline{K}}(|\boldsymbol{u}_h^{k,i}|)]^{1/2}\boldsymbol{u}_h^{k,i} + [\tilde{\underline{K}}(|\boldsymbol{u}_h^{k,i}|)]^{-1/2}\nabla\zeta_h^{k,i}\big)\|_K^2 \\
&\leq C_{\tilde{\underline{K}}} \|[\tilde{\underline{K}}(|\boldsymbol{u}_h^{k,i}|)]^{1/2}\boldsymbol{u}_h^{k,i} + [\tilde{\underline{K}}(|\boldsymbol{u}_h^{k,i}|)]^{-1/2}\nabla\zeta_h^{k,i}\|_K^2 && \text{by (6.4)}\\
&= C_{\tilde{\underline{K}}}\Big((\tilde{\underline{K}}(|\boldsymbol{u}_h^{k,i}|)\boldsymbol{u}_h^{k,i}, \boldsymbol{u}_h^{k,i})_K + 2(\boldsymbol{u}_h^{k,i}, \nabla\zeta_h^{k,i})_K + ([\tilde{\underline{K}}(|\boldsymbol{u}_h^{k,i}|)]^{-1}\nabla\zeta_h^{k,i}, \nabla\zeta_h^{k,i})_K\Big) \\
&\leq C_{\tilde{\underline{K}}}^2 \|\boldsymbol{u}_h^{k,i}\|_K^2 + 2C_{\tilde{\underline{K}}}(\boldsymbol{u}_h^{k,i}, \nabla\zeta_h^{k,i})_K + c_{\tilde{\underline{K}}}^{-1} C_{\tilde{\underline{K}}} \|\nabla\zeta_h^{k,i}\|_K^2. && \text{by (6.4)}
\end{aligned}
$$

For the first term above, we proceed as in (6.30). For the second one, the development (5.39) remains unchanged. Finally, for the third term, we use (5.34) where $\underline{K}$ has been replaced by $c_{\tilde{\underline{K}}}^{-2} C_{\tilde{\underline{K}}} \boldsymbol{I}$ to see that

$$c_{\tilde{\underline{K}}}^{-2} C_{\tilde{\underline{K}}} \|\nabla\zeta_h^{k,i}\|_K^2 = (\mathsf{Z}_K^{k,i})^{\mathrm{t}} \widehat{\mathbb{S}}_{\mathrm{FE},K} \mathsf{Z}_K^{k,i}.$$

The proof is finished by combining the two estimates on $c_{\tilde{\underline{K}}}^{1/2} \|\boldsymbol{u} - \boldsymbol{\zeta}\|$ and $c_{\tilde{\underline{K}}}^{1/2} \|\boldsymbol{\zeta} - \boldsymbol{u}_h^{k,i}\|$. $\qquad\square$

Remark 6.4 (Inexpensive implementation and evaluation)**.** *Note that the estimate of Theorem 6.2 still takes the same simple matrix-vector multiplication form of Theorem 5.18, with in particular the same element matrices $\widehat{\mathbb{A}}_{\mathrm{MFE},K}$, $\widehat{\mathbb{S}}_{\mathrm{FE},K}$, $\widehat{\mathbb{M}}_{\mathrm{FE},K}$ as in the liner case. Similarly to Corollaries 5.19 and 5.20, a further simplification is to replace $\widehat{\mathbb{A}}_{\mathrm{MFE},K}$ by the element matrices $\widehat{\mathbb{A}}_K$ of the given scheme whenever they are available, as per Assumption 5.2.*

Remark 6.5 (Distinction of error components)**.** *The distinction of the different error components is based on the face fluxes (6.19), directly available form any finite volume discretization (6.11)–(6.16). Beyond simplicity, a remarkable property is that all the error component estimators have the same physical units (that of energy flux error) and take the same form. This stands in contrast to the usual practice where the algebraic error is typically treated by the (relative) L^2-norm of the algebraic residual (here, from (6.33), we rather employ the energy-scaled $H^{-1}(\Omega)$-norm of the residual) and the linearization error estimator is often the L^∞-norm of the difference of two consecutive iterates. This unified treatment of the different error components is a consequence of our holistic approach. In practice, the distinction of error components is to be used in an adaptive inexact Newton algorithm such as Algorithm 7.1 of the next section.*

Remark 6.6 (Error structure). *From* (6.26), (6.28), *and* (6.32a), *we see that*

$$c_{\underline{\tilde{K}}}^{1/2}\|\boldsymbol{u} - \boldsymbol{u}_h^{k,i}\| \leq c_{\underline{\tilde{K}}}^{-1/2}C_{\tilde{K}}\max_{\varphi\in H_0^1(\Omega),\,\|\nabla\varphi\|=1}\{(f,\varphi) + (\boldsymbol{u}_h^{k,i},\nabla\varphi)\} + c_{\underline{\tilde{K}}}^{-1/2}\min_{v\in H_0^1(\Omega)}\|\underline{\tilde{K}}(|\boldsymbol{u}_h^{k,i}|)\boldsymbol{u}_h^{k,i} + \nabla v\|$$

$$\leq 2c_{\underline{\tilde{K}}}^{-1/2}C_{\tilde{K}}\|\boldsymbol{u} - \boldsymbol{u}_h^{k,i}\|; \tag{6.33}$$

in the second inequality, we have used the Cauchy–Schwarz inequality and (6.8) *together with* (6.3b) *for the second term. This means that the energy error* $c_{\underline{\tilde{K}}}^{1/2}\|\boldsymbol{u}-\boldsymbol{u}_h^{k,i}\|$ *is equivalent to the sum of the* dual norm *of the residual and* nonconformity *evaluated as the* distance to the $H_0^1(\Omega)$ space. *This is an immediate extension of Theorem 4.9 from the linear case. Compared to Corollary 4.6 for the linear case with an exact algebraic solver giving* $\nabla\cdot\boldsymbol{u}_h = f$, *this in particular enables to treat the nonlinear case with* inexact *linearization and algebraic solvers, so that* $\nabla\cdot\boldsymbol{u}_h^{k,i} \neq f$. *Our motivation here was to evaluate the error as the* $\boldsymbol{L}^2(\Omega)$-*norm, which gives rise to the weights* $c_{\tilde{K}}$ *and* $C_{\tilde{K}}$ *and the factor* 2 *in* (6.33). *Alternatively, we could have stick to the* intrinsic error measure *of the form*

$$\left\{\max_{\varphi\in H_0^1(\Omega),\,\|\nabla\varphi\|=1}\{(f,\varphi) + (\boldsymbol{u}_h^{k,i},\nabla\varphi)\}^2 + \min_{v\in H_0^1(\Omega)}\|\underline{\tilde{K}}(|\boldsymbol{u}_h^{k,i}|)\boldsymbol{u}_h^{k,i} + \nabla v\|^2\right\}^{\frac{1}{2}},$$

cf. (4.9), *which would lead to constant-free estimates without the weights* $c_{\tilde{K}}$ *and* $C_{\tilde{K}}$. *In this way, we define the error measure* $\mathcal{N}^{n,k,i}$ *in* (7.17a) *in the complex multiphase flow below.*

6.8 Numerical experiments

In this section, we numerically illustrate the efficiency of our theoretical results of Section 6.7 on two different examples. Our main goals are to asses the sharpness of the guaranteed bound (6.24) and to examine the robustness of our estimates with respect to the ratio $C_{\tilde{K}}/c_{\tilde{K}}$ where we consider $C_{\tilde{K}}/c_{\tilde{K}} = 10^i$, $i \in \{1,...,6\}$. As in previous sections, the effectivity index is defined by the ratio of the estimator from Theorem 6.2 to the exact error,

$$I_{\text{eff}} := \frac{\eta^{k,i}}{c_{\underline{\tilde{K}}}^{1/2}\|\boldsymbol{u} - \boldsymbol{u}_h^{k,i}\|}. \tag{6.34}$$

Additionally, we study in the following examples a stopping criterion for the linearization algorithm based on balancing the error components of Theorem 6.2. The criterion is that the linearization iteration is pursued until step k such that

$$\eta_{\text{lin}}^{k,i} \leq \Gamma_{\text{lin}}\eta_{\text{sp}}^{k,i}, \tag{6.35}$$

with $\Gamma_{\text{lin}} \in (0,1)$ (we focus on the linearization only, the algebraic solver is used with a very small tolerance yielding $\eta_{\text{alg}}^{k,i}, \eta_{\text{rem}}^{k,i} \approx 0$). We compare this criterion with a classic one based on a fixed threshold on the relative linearization residual

$$e_{\text{lin}}^k \leq 10^{-8}, \tag{6.36}$$

where the relative linearization residual, for a resolution of a system of nonlinear algebraic equations $F(X) = 0$ by the Newton method, is given by

$$e_{\text{lin}}^k := \frac{\|F(X^k)\|}{\|F(X^0)\|}.$$

Note that the criterion (6.35) can be used in more general adaptive algorithm balancing all the error components such as in Algorithm 7.1.

6.8.1 Regular solution

In this test, we consider on the domain $\Omega = (0,1) \times (0,1)$ the exact solution

$$p(x,y) = 16x(1-x)y(1-y). \tag{6.37}$$

The source term f is then constructed through (6.1). We perform the numerical tests with the nonlinearity example given by (6.5) on a uniformly refined rectangular mesh of 10^4 elements. Figure 24 shows the
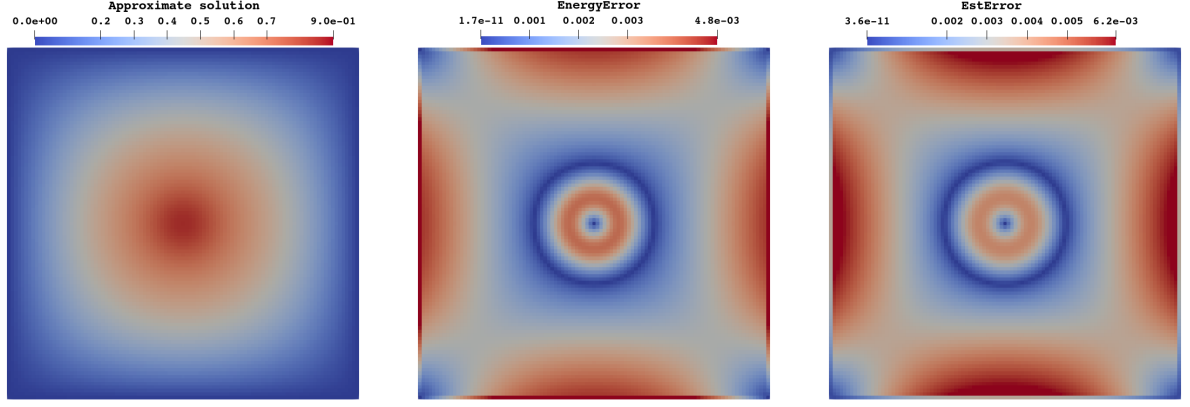
Figure 24: [Section 6.8.1] The approximate solution given by the values $p_K^{k,i}$, $K \in \mathcal{T}_h$, (*left*) the energy errors $c_{\tilde{\boldsymbol{K}}}^{1/2} \|\boldsymbol{u} - \boldsymbol{u}_h^{k,i}\|_K$ (*middle*), and the error estimators $\eta_{\mathrm{sp},K}^{k,i} + \eta_{\mathrm{lin},K}^{k,i} + \eta_{\mathrm{alg},K}^{k,i} + \eta_{\mathrm{rem},K}^{k,i}$ from Theorem 6.2 (*right*).
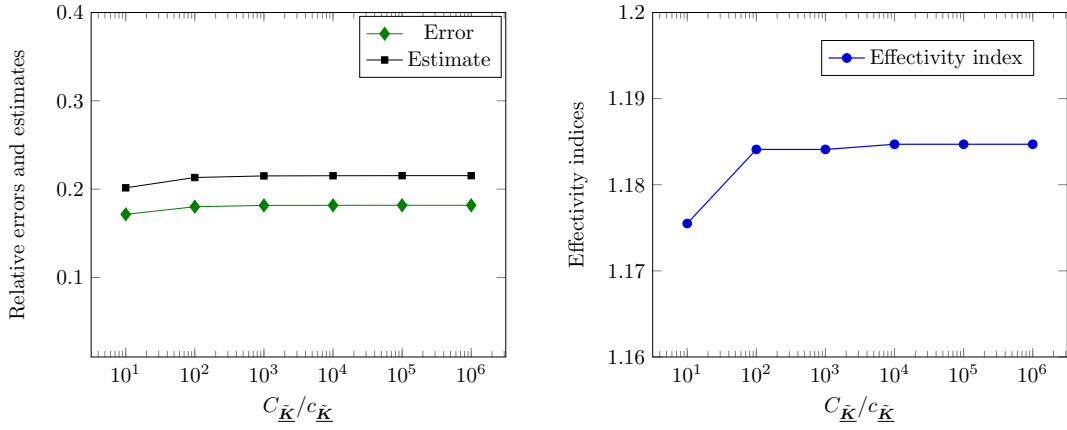
Figure 25: [Section 6.8.1], Relative errors $c_{\tilde{\boldsymbol{K}}}^{1/2} \|\boldsymbol{u} - \boldsymbol{u}_h^{k,i}\| / \|\boldsymbol{u}_h^{k,i}\|$ and relative estimates $\eta^{k,i} / \|\boldsymbol{u}_h^{k,i}\|$ from Theorem 6.2 (*left*), effectivity indices I_{eff} from (6.34) (*right*)

approximate solution given by the values $p_K^{k,i}$, $K \in \mathcal{T}_h$, the elementwise errors $c_{\tilde{\boldsymbol{K}}}^{1/2} \|\boldsymbol{u} - \boldsymbol{u}_h^{k,i}\|_K$, and the corresponding a posteriori error estimators $\eta_K^{k,i}$ from Theorem 6.2.

The robustness of our estimates with respect to the ratio $C_{\tilde{\boldsymbol{K}}}/c_{\tilde{\boldsymbol{K}}}$ is examined in Figure 25, where we fix the value of $c_{\tilde{\boldsymbol{K}}} = 1$ and plot, for different values of $C_{\tilde{\boldsymbol{K}}} \in \{10, 100, ..., 10^6\}$, the values of the relative energy error $c_{\tilde{\boldsymbol{K}}}^{1/2} \|\boldsymbol{u} - \boldsymbol{u}_h^{k,i}\| / \|\boldsymbol{u}_h^{k,i}\|$ and the corresponding relative total error estimate $\eta^{k,i} / \|\boldsymbol{u}_h^{k,i}\|$ from Theorem 6.2. The effectivity index plotted in the right part of Figure 25 reveal that, in this case and for the choice of nonlinearity example (6.5), the estimate is robust and not affected by the values of the ratio $C_{\tilde{\boldsymbol{K}}}/c_{\tilde{\boldsymbol{K}}}$.

Figure 26 is dedicated to study the stopping criterion (6.35) and its comparison with the usual stopping criterion (6.36). The left part of Figure 26 depicts the evolution of the total estimator $\eta^{k,i}$, the linearization estimator $\eta_{\mathrm{lin}}^{k,i}$, and the relative linearization residual e_{lin}^k as a function of the number of Newton iterations for a uniform fixed mesh of 10^4 elements and with the ratio $C_{\tilde{\boldsymbol{K}}}/c_{\tilde{\boldsymbol{K}}} = 10$. We observe that the linearization estimator and the relative linearization residual steadily decrease, while the total estimator stagnate starting from the third iteration. The stopping criterion (6.35) with $\Gamma_{\mathrm{1in}} = 0.1$, indicates that the resolution can be stopped at the sixth iteration, which already yields a sufficiently accurate approximate solution and helps avoid unnecessary additional iterations. The right part of Figure 26 shows a comparison of the required number of Newton iteration in order to satisfy the standard criterion $e_{\mathrm{lin}}^k \leq 10^{-8}$ and the adaptive one $\eta_{\mathrm{lin}}^{k,i} \leq 0.1\eta_{\mathrm{sp}}^{k,i}$ for different values of the ratio $C_{\tilde{\boldsymbol{K}}}/c_{\tilde{\boldsymbol{K}}}$. We see that we can save at least the half of Newton iterations on each resolution without altering the precision.

Figure 26: [Section 6.8.1], Standard linearization vs. adaptive linearization

Figure 27: [Section 6.8.2] The approximate solution given by the values $p_K^{k,i}$, $K \in \mathcal{T}_h$, (*left*), the energy errors $c_{\tilde{K}}^{1/2}\|\boldsymbol{u}-\boldsymbol{u}_h^{k,i}\|_K$ (*middle*), and the error estimators $\eta_{\text{sp},K}^{k,i}+\eta_{\text{lin},K}^{k,i}+\eta_{\text{alg},K}^{k,i}+\eta_{\text{rem},K}^{k,i}$ from Theorem 6.2 (*right*).
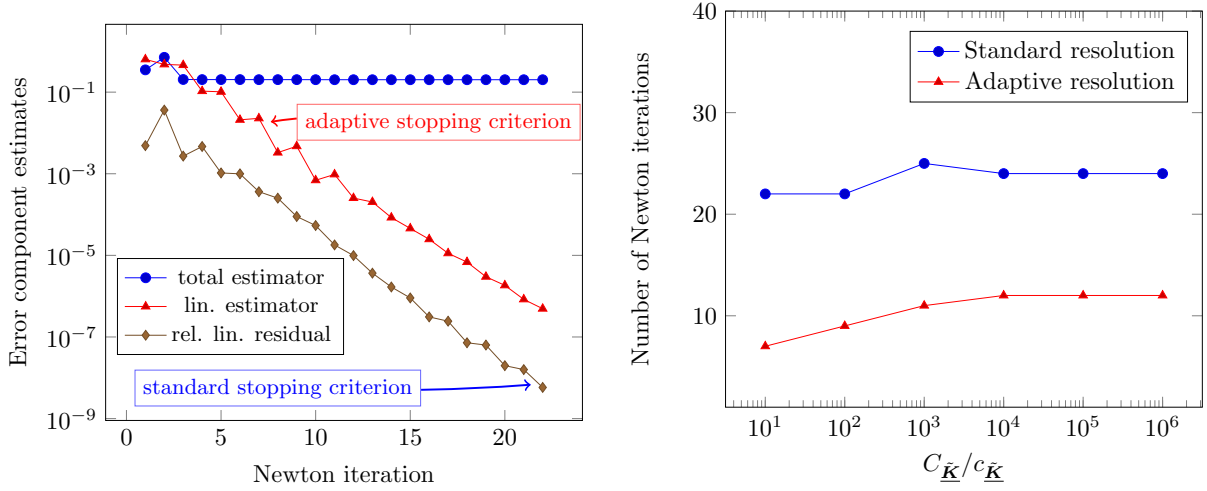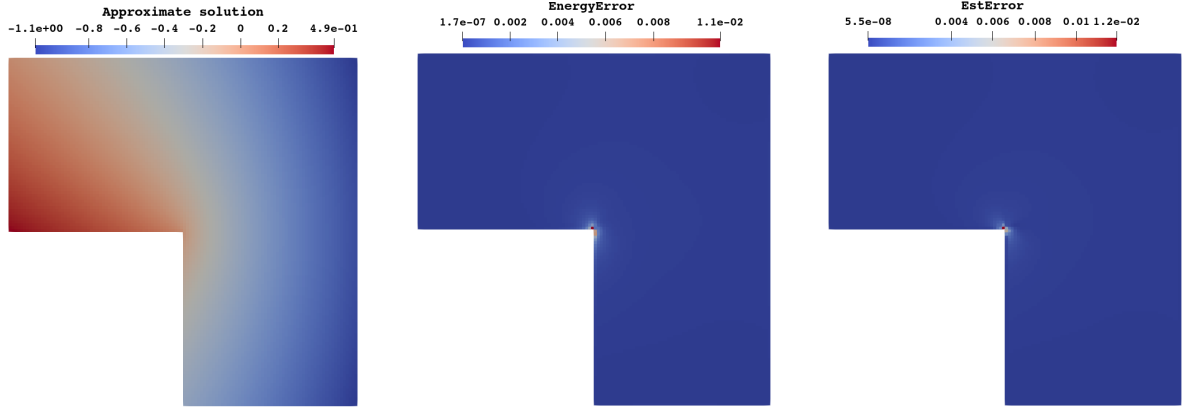
6.8.2 Singular solution

In this test, we consider the L-shaped domain

$$\Omega = (-1,1) \times (-1,1) \setminus (-1,0] \times (-1,0]$$

with the exact solution

$$p(r,\theta) = r^{\frac{2}{3}} \sin\left(\frac{2}{3}\theta + \frac{3}{2}\pi\right),$$

where (r,θ) are the polar coordinates. The Dirichlet boundary condition is partly inhomogeneous and given by the value of the exact solution on $\partial\Omega$. The following results are carried out with the choice of the nonlinearity example given by (6.5) and on a uniformly refined rectangular mesh of 10800 elements.

In Figure 27, we show the distribution of the approximate solution and we compare the distributions of the errors $c_{\tilde{K}}^{1/2}\|\boldsymbol{u} - \boldsymbol{u}_h^{k,i}\|_K$ and the corresponding a posteriori error estimators $\eta_K^{k,i}$ from Theorem 6.2. We can observe that the estimators identify perfectly the error in the zone around the singularity at (0,0).

In Figure 28, we study the effect of the ratio $C_{\tilde{K}}/c_{\tilde{K}}$ on our estimators. The left part of Figure 28 shows the relative exact errors $c_{\tilde{K}}^{1/2}\|\boldsymbol{u} - \boldsymbol{u}_h^{k,i}\|/\|\boldsymbol{u}_h^{k,i}\|$ together with their relative a posteriori error estimates $\eta^{k,i}/\|\boldsymbol{u}_h^{k,i}\|$ from Theorem 6.2. The right part of Figure 28 displays the corresponding effectivity indices,
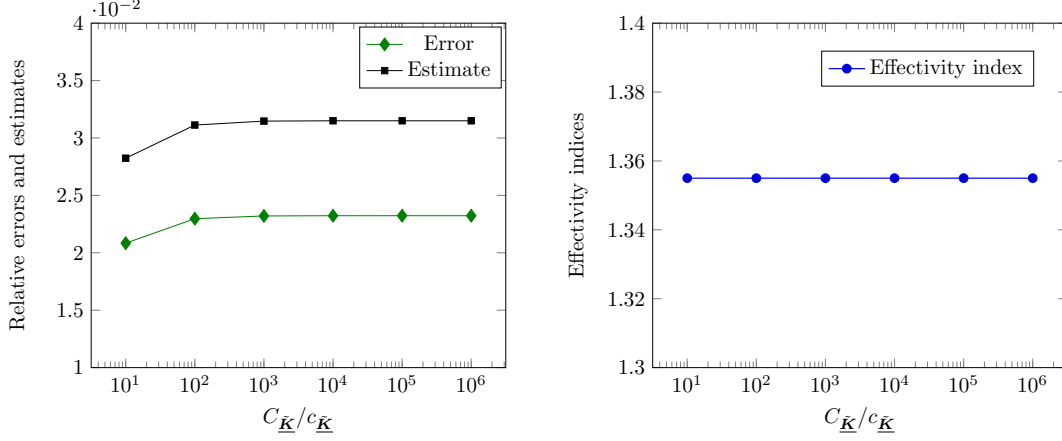
Figure 28: [Section 6.8.2], Relative errors $c_{\underline{\tilde{K}}}^{1/2}\|\boldsymbol{u} - \boldsymbol{u}_h^{k,i}\|/\|\boldsymbol{u}_h^{k,i}\|$ and relative estimates $\eta^{k,i}/\|\boldsymbol{u}_h^{k,i}\|$ from Theorem 6.2 (*left*), effectivity indices I_{eff} from (6.34) (*right*)
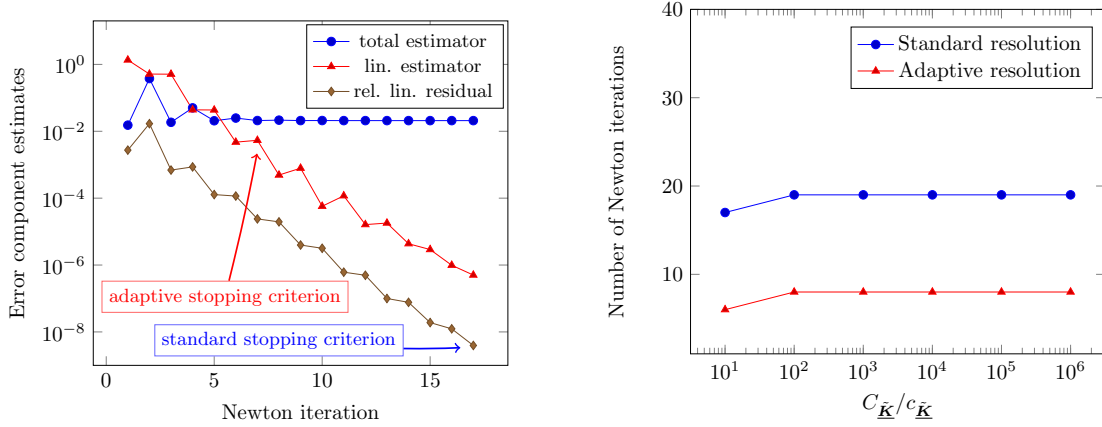

Figure 29: [Section 6.8.2], Standard linearization vs. adaptive linearization

given by the ratio of the estimate over the error (6.34). We see that the effectivity indices are not affected by the values of the ratio $C_{\underline{\tilde{K}}}/c_{\underline{\tilde{K}}}$, which numerically gives the robustness of the estimate.

In order to evaluate the potential of a stopping criterion for the linearization algorithm, we illustrate in the left part of Figure 29 the evolution of the total estimator $\eta^{k,i}$, the linearization estimator $\eta_{\mathrm{lin}}^{k,i}$, and the relative linearization residual e_{lin}^k as a function of the number of Newton iterations for a uniform fixed mesh of 10800 elements and with the ratio $C_{\underline{\tilde{K}}}/c_{\underline{\tilde{K}}} = 10$. These results are given with a standard resolution stopped when $e_{\mathrm{lin}}^k \leq 10^{-8}$. The linearization estimator and the relative linearization residual decrease as expected while the total estimator start to stagnate from the fifth iteration. Considering the stopping criterion (6.35) with $\Gamma_{\mathrm{lin}} = 0.1$ suggests that the resolution process can be halted after the seventh iteration, as it already provides a sufficiently accurate approximate solution while avoiding a significant number of unnecessary linearization iterations. In the right part of Figure 29, computational savings in terms of linearization iterations is observed for different values of the ratio $C_{\underline{\tilde{K}}}/c_{\underline{\tilde{K}}}$. These results make it apparent that using criteria based on balancing the error components gives a powerful tool for adaptivity algorithms. This is shown in the next section by considering Algorithm 7.1 applied to a realistic problem and achieving a significant gain in terms of the computation effort.

6.9 Bibliographic resources

One of the first a posteriori error estimates for finite volume methods for nonlinear problems is presented in Bergam *et al.* [48], though it concerns vertex-centered finite volumes, yielding $H_0^1(\Omega)$-conforming approximations as in the finite element case. Kim [163] was one of the first to consider locally conservative methods in a Prager–Synge-type framework of the form (6.32).

The distinction of different error components (and the associated adaptivity) has been initiated in

Han [143], Becker *et al.* [39], Kaltenbacher [156], Axelsson and Kaporin [27], Patera and Rønquist [211], Arioli [24], Bakushinsky and Smirnova [30], Chaillou and Suri [75, 74], Picasso [215], Ahusborde *et al.* [11], Bernardi *et al.* [49, 50], Jiránek *et al.* [154], El Alaoui *et al.* [114], Nordbotten *et al.* [200], Arioli *et al.* [26, 25], Rannacher and Vihharev [217], Ern and Vohralík [124], Keilegavlen and Nordbotten [158, 159], Rey *et al.* [222, 221, 220], and Becker *et al.* [38], see also the references therein. It has been reconsidered more recently, namely in the framework of conforming finite elements, in Heid and Wihler [147], Heid *et al.* [146], Gantner *et al.* [136], Haberl *et al.* [142], Kumar *et al.* [171], Chaudhry *et al.* [77], Dolgov and Vejchodský [106], Liu and Keyes [182], Févotte *et al.* [134], Bringmann *et al.* [66], Harnist *et al.* [145], Mitra and Vohralík [189], Stokke *et al.* [233], and Papež [209]. The classical books of Kelley [160], Saad [229], or Olshanskii and Tyrtyshnikov [205] do not discuss total a posteriori error estimates during iterative linearization and iterative linear algebraic solvers iteration but rather only treat the nonlinear algebra level (for iterative linearization) and linear algebra level (for iterative linear algebraic solvers), not addressing the discretization/partial differential equation level.

7 Unsteady nonlinear coupled degenerate advection–diffusion–reaction problems. Temporal discretization error

In this section, the methodology derived in Sections 4–6 is applied to polytopal discretizations of a model real-life *multiphase compositional Darcy flow* in porous media. In the following, we describe briefly the multiphase model, discuss its discretization by an implicit finite volume scheme, the linearization by the Newton or combined Newton and fixed point method, and the algebraic solution of the arising linear system. We explain that the required fictitious flux and potential reconstructions stay unchanged with respect to Section 6 and finally develop our a posteriori error estimate with inexpensive implementation and evaluation distinguishing the different error components.

By decomposing the estimators into the space, time, linearization, and algebraic error components, we can design an algorithm involving adaptivity of both solvers of meshes. We namely formulate criteria for stopping the iterative algebraic solver and the iterative linearization solver when the corresponding error components do not affect significantly the total error. Moreover, the spatial and temporal error components are balanced respectively by time step and space mesh adaptation, materializing the abstract considerations of Section 2.6. Denote by $\eta_{\mathrm{sp}}^{n,k,i}, \eta_{\mathrm{tm}}^{n,k,i}, \eta_{\mathrm{lin}}^{n,k,i}$, and $\eta_{\mathrm{alg}}^{n,k,i}$ respectively the total spatial, temporal, linearization, and algebraic estimators, on time step n, linearization step k, and algebraic solver step i. Let $\gamma_{\mathrm{lin}}, \gamma_{\mathrm{alg}} \in (0, 1)$ and $\Gamma_{\mathrm{tm}} > \gamma_{\mathrm{tm}} > 0$ be user-given fixed balancing parameters and fix the fractions (percentages) of cells to refine δ_{ref} and to derefine δ_{deref}, $0 < \delta_{\mathrm{deref}} < \delta_{\mathrm{ref}} < 1$. Following [124, 253, 70, 103, 104] and the references therein, we in particular derive in this section the following *fully adaptive algorithm*:

Algorithm 7.1 (Adaptive stopping criteria and adaptive time and space mesh refinement)**.**
  Set $n := 0$.
  while $t^n \leq t_{\mathrm{F}}$ **do** {time loop}
    Set $n := n + 1$.
    loop {spatial and temporal errors balancing loop}
      Set $k := 0$.
      loop {linearization loop}
        Set $k := k + 1$.
        Set up the linear system.
        Set $i := 0$.
        loop {algebraic solver loop}
          Perform a step of the iterative algebraic solver and set $i := i + 1$.
          Evaluate the different a posteriori error estimators.
          Terminate if: $\eta_{\mathrm{alg}}^{n,k,i} \leq \gamma_{\mathrm{alg}} \eta_{\mathrm{sp}}^{n,k,i}$.
        end loop {algebraic solver loop}
        Terminate if: $\eta_{\mathrm{lin}}^{n,k,i} \leq \gamma_{\mathrm{lin}} \eta_{\mathrm{sp}}^{n,k,i}$.
      end loop {linearization loop}
      Terminate if

$$\eta_{\mathrm{sp}}^{K,n,k,i} \geq \delta_{\mathrm{ref}} \max_{L \in \mathcal{T}_H^n} \left\{ \eta_{\mathrm{sp}}^{L,n,k,i} \right\} \qquad \forall K \in \mathcal{T}_H^n$$

$$\gamma_{\mathrm{tm}} \eta_{\mathrm{sp}}^{n,k,i} \leq \eta_{\mathrm{tm}}^{n,k,i} \leq \Gamma_{\mathrm{tm}} \eta_{\mathrm{sp}}^{n,k,i};$$

else

Refine the cells $K \in \mathcal{T}_H^n$ such that $\eta_{\mathrm{sp}}^K, n, k, i \geq \delta_{\mathrm{ref}} \max_{L \in \mathcal{T}_H^n} \{\eta_{\mathrm{sp}}^L, n, k, i\}$.

Derefine the cells $K \in \mathcal{T}_H^n$ such that $\eta_{\mathrm{sp}}^K, n, k, i \leq \delta_{\mathrm{deref}} \max_{L \in \mathcal{T}_H^n} \{\eta_{\mathrm{sp}}^L, n, k, i\}$.

Refine the time step if $\eta_{\mathrm{tm}}^{n,k,i} > \Gamma_{\mathrm{tm}} \eta_{\mathrm{sp}}^t n, k, i$, derefine the time step if $\gamma_{\mathrm{tm}} \eta_{\mathrm{sp}}^{n,k,i} > \eta_{\mathrm{tm}}^{n,k,i}$.

end loop {spatial and temporal errors balancing loop}

Update data.

end while {time loop}

7.1 Multiphase compositional unsteady Darcy flow

We consider the multiphase compositional Darcy flow model in porous media [4, 28, 93, 83, 261]. We use the generalization of Coats' formulation [94] to an arbitrary number of phases by [129].

Let $\mathcal{P} = \{p\}$ be the set of *phases*, $\mathcal{C} = \{c\}$ the set of *components*, and, for a given phase $p \in \mathcal{P}$, let $\mathcal{C}_p \subset \mathcal{C}$ be the set of its components. S_p then denotes the *saturation* of the phase p and $C_{p,c}$ the *molar fraction* of the component c in the phase p. For a given component $c \in \mathcal{C}$, denote by $\mathcal{P}_c$ the set of the phases which contain c.

We denote by P the *reference pressure* such that the *phase pressures* P_p, $p \in \mathcal{P}$, are expressed as

$$P_p := P + P_{\mathrm{c}_p}(S_p), \tag{7.1}$$

where $P_{\mathrm{c}_p} : \mathbb{R} \to \mathbb{R}$ is a given generalized capillary pressure function. We collect the unknowns of the model in the vector

$$\boldsymbol{X} := (P, (S_p)_{p \in \mathcal{P}}, (C_{p,c})_{p \in \mathcal{P}, c \in \mathcal{C}_p}). \tag{7.2}$$

For a phase $p \in \mathcal{P}$, let $k_{\mathrm{r},p} : \mathbb{R} \to \mathbb{R}$ be the relative permeability function of the unknown saturation S_p. We denote by ϕ the porosity of the medium, by $\underline{\boldsymbol{K}}$ the permeability tensor, by μ_p the dynamic viscosity, by ς_p the molar density, and by ρ_p the mass density. For simplicity, we suppose that all these porous media parameters are given as constants on each mesh element and on each time step; ϕ is actually supposed to be constant in time. We also define the mobility by $\nu_p := \varsigma_p \frac{k_{\mathrm{r},p}}{\mu_p}$. Finally, let q_c be the *source or sink* term, again piecewise constant on the space-time mesh.

Let $t_{\mathrm{F}} > 0$ be the final simulation time. The system of governing equations is posed in $\Omega \times (0, t_{\mathrm{F}})$ by

$$\boxed{\partial_t l_c + \nabla \cdot \boldsymbol{\theta}_c = q_c, \qquad \forall c \in \mathcal{C},} \tag{7.3}$$

where l_c is the *amount* (in moles) of component c per unit volume,

$$l_c := \phi \sum_{p \in \mathcal{P}_c} \varsigma_p S_p C_{p,c}. \tag{7.4}$$

In (7.3), for each component $c \in \mathcal{C}$, the *component flux* $\boldsymbol{\theta}_c$ has the following expression:

$$\boldsymbol{\theta}_c := \sum_{p \in \mathcal{P}_c} \boldsymbol{\theta}_{p,c}, \qquad \boldsymbol{\theta}_{p,c} := \boldsymbol{\theta}_{p,c}(\boldsymbol{X}) = \nu_p C_{p,c} \boldsymbol{v}_p(P_p), \tag{7.5}$$

where for all phases $p \in \mathcal{P}$, $\boldsymbol{v}_p(P_p)$ represents the average *phase velocity* given by Darcy's law,

$$\boldsymbol{v}_p(P_p) := -\underline{\boldsymbol{K}}(\nabla P_p + \rho_p g \nabla z). \tag{7.6}$$

Here, z is the vertical space coordinate and g is the gravitation acceleration constant. Note at this occasion that only the first part of the Darcy velocity,

$$\boldsymbol{u}_p(P_p) := -\underline{\boldsymbol{K}} \nabla P_p, \tag{7.7}$$

is a gradient of the phase pressure. We will use this observation in our error measure below. We assume that no-flow boundary conditions are prescribed for all the component fluxes,

$$\boldsymbol{\theta}_c \cdot \boldsymbol{n}_\Omega = 0 \qquad \text{on } \partial\Omega \times (0, t_{\mathrm{F}}) \qquad \forall c \in \mathcal{C}, \tag{7.8}$$

where $\partial\Omega$ denotes the boundary of Ω and $\boldsymbol{n}_\Omega$ its unit outward normal. At $t = 0$, we prescribe the *initial amount* of each component,

$$l_c(\cdot, 0) = l_c^0 \qquad \forall c \in \mathcal{C}. \tag{7.9}$$

The previous PDEs is supplemented by a system of algebraic equations imposing the volume conservation

$$\sum_{p \in \mathcal{P}} S_p = 1, \tag{7.10}$$

the conservation of the quantity of matter

$$\sum_{c \in \mathcal{C}_p} C_{p,c} = 1 \qquad \forall p \in \mathcal{P}, \tag{7.11}$$

and local thermodynamic equilibrium expressed by

$$\sum_{c \in \mathcal{C}} (|\mathcal{P}_c| - 1) = \sum_{p \in \mathcal{P}} |\mathcal{C}_p| - |\mathcal{C}| \tag{7.12}$$

equalities of fugacities, where we refer to [103] for details.

7.2 Space-time mesh

We suppose that the time simulation interval $[0, t_{\mathrm{F}}]$ is partitioned into time intervals $I^n := [t^{n-1}, t^n]$, $1 \leq n \leq N$, given by the discrete times $0 = t^0 < \ldots < t^n < \ldots < t^N = t_{\mathrm{F}}$. The discrete time steps are then $\tau^n := t^n - t^{n-1}$. Then, we let $(\mathcal{T}_H^n)_{0 \leq n \leq N}$ be a family of polytopal meshes of the space domain Ω defined in the sense of Section 3.6; we suppose that $\mathcal{T}_H^n$ has been obtained from $\mathcal{T}_H^{n-1}$ by refinement of some elements and by coarsening of some other. For all $0 \leq n \leq N$, we define $\mathcal{F}_H^n$, $\mathcal{F}_h^n$, $\mathcal{F}_{H,h}^n$, $\mathcal{F}_{K,h}^{n,\mathrm{ext}}$, and $\mathcal{F}_{K,h}^{n,\mathrm{int}}$ as in Section 3.6.

7.3 Weak solution and its properties

At this stage, we need to characterize a weak solution for the multiphase compositional model (7.1)–(7.12). We define

$$X := L^2((0, t_{\mathrm{F}}); H^1(\Omega)), \tag{7.13a}$$

$$Y := H^1((0, t_{\mathrm{F}}); L^2(\Omega)). \tag{7.13b}$$

We equip the space X with the norm

$$\|\varphi\|_X := \left\{ \sum_{n=1}^N \|\varphi\|_{X^n}^2 \right\}^{\frac{1}{2}}, \qquad \|\varphi\|_{X^n}^2 := \int_{I^n} \sum_{K \in \mathcal{T}_H^n} \left\{ h_K^{-2} \|\varphi\|_K^2 + \|\underline{\boldsymbol{K}}^{\frac{1}{2}} \nabla \varphi\|_K^2 \right\} \mathrm{d}t \qquad \varphi \in X, \tag{7.14}$$

where, we recall, h_K is the diameter of the cell $K \in \mathcal{T}_H^n$. Below, it will be convenient to use $X^n := L^2(I^n; H^1(\Omega))$. We suppose that there exists a unique weak solution of (7.1)–(7.12) in the following sense:

Assumption 7.2 (Weak solution). *There exists a weak solution $\boldsymbol{X}$ of (7.1)–(7.12) which can be characterized as follows:*

$$l_c \in Y \qquad \forall c \in \mathcal{C}, \tag{7.15a}$$

$$P_p \in X \qquad \forall p \in \mathcal{P}, \tag{7.15b}$$

$$\boldsymbol{\theta}_c \in [L^2((0, t_{\mathrm{F}}); L^2(\Omega))]^d \qquad \forall c \in \mathcal{C}, \tag{7.15c}$$

$$\int_0^{t_{\mathrm{F}}} \left\{ (\partial_t l_c, \varphi)(t) - (\boldsymbol{\theta}_c, \nabla \varphi)(t) \right\} \mathrm{d}t = \int_0^{t_{\mathrm{F}}} (q_c, \varphi)(t) \mathrm{d}t \qquad \forall \varphi \in X, \forall c \in \mathcal{C}, \tag{7.15d}$$

the initial condition (7.9) holds, \tag{7.15e}

and the algebraic closure equations (7.10)–(7.12) hold, \tag{7.15f}

where P_p, l_c, and $\boldsymbol{\theta}_c$ are defined, respectively, by (7.1), (7.4), and (7.5).

We have formulated the existence and uniqueness of a weak solution as an assumption, since, in contrast to the model problems of Sections 4–6, the problem (7.1)–(7.12) is too abstract and complicated to prove it. In simplified settings, with possibly only two phases present and each phase composed of a single component, such results can be found in [169, 78, 23, 81, 69, 161, 19] and the references therein.

Remark 7.3 (Component fluxes). *Since we assume q_c piecewise constant and thus $q_c \in L^2((0, t_F);$ $L^2(\Omega))$, it follows from (7.15) that actually*

$$\boldsymbol{\theta}_c \in L^2((0, t_F); \mathbf{H}(\mathrm{div}, \Omega)) \qquad \forall c \in \mathcal{C}, \tag{7.16a}$$

$$\nabla \cdot \boldsymbol{\theta}_c = q_c - \partial_t l_c \qquad \forall c \in \mathcal{C}, \tag{7.16b}$$

$$\boldsymbol{\theta}_c \cdot \boldsymbol{n}_\Omega = 0 \qquad \text{on } \partial\Omega \times (0, t_F) \qquad \forall c \in \mathcal{C}, \tag{7.16c}$$

so that the component fluxes $\boldsymbol{\theta}_c$ have the normal trace continuous in a proper sense, the governing equation (7.3) is satisfied with a weak divergence, and the boundary conditions (7.8) hold in the normal trace sense. This extends Proposition 4.1, (5.4), and (6.9) to the multiphase compositional model (7.1)–(7.12).

7.4 Total error and its spatial discretization, temporal discretization, linearization, and algebraic components

Suppose that Assumption 7.2 holds, so that we dispose of a weak solution. Then, we describe the distance of the current approximate solution to the exact solution by the following intrinsic error measure:

$$\mathcal{N}^{n,k,i} := \left\{ \sum_{c \in \mathcal{C}} \left(\mathcal{N}_c^{n,k,i}\right)^2 \right\}^{\frac{1}{2}} + \left\{ \sum_{p \in \mathcal{P}} \left(\mathcal{N}_p^{n,k,i}\right)^2 \right\}^{\frac{1}{2}}, \tag{7.17a}$$

where

$$\mathcal{N}_c^{n,k,i} := \sup_{\varphi \in X^n, \|\varphi\|_{X^n} = 1} \int_{I^n} \left\{ (q_c, \varphi) - (\partial_t l_{c,h\tau}^{n,k,i}, \varphi) + \left(\boldsymbol{\theta}_{c,h\tau}^{n,k,i}, \nabla\varphi\right) \right\} \mathrm{d}t \tag{7.17b}$$

is the *dual norm of the residual* of the weak formulation for each component flux $\boldsymbol{\theta}_{c,h\tau}^{n,k,i}$ on the time interval I^n and where

$$\mathcal{N}_p^{n,k,i} := \inf_{\varsigma_p \in X^n} \left\{ \sum_{c \in \mathcal{C}_p} \int_{I^n} \left\{ \sum_{K \in \mathcal{T}_H^n} \left(\nu_{p,K}^{n,k,i} C_{p,c,K}^{n,k,i}\right)^2 \|\boldsymbol{u}_{p,h\tau}^{n,k,i} + \underline{\boldsymbol{K}}\nabla\varsigma_p\|_{\underline{\boldsymbol{K}}^{-\frac{1}{2}}, K}^2 \right\} \mathrm{d}t \right\}^{\frac{1}{2}} \tag{7.17c}$$

evaluates the *nonconformity* (the distance of $\boldsymbol{u}_{p,h\tau}^{n,k,i}$ to $\underline{\boldsymbol{K}}\nabla H_0^1(\Omega)$) for each Darcy phase flux $\boldsymbol{u}_{p,h\tau}^{n,k,i}$. Indeed, from (7.16b), the first term should be zero and from (7.7), $\boldsymbol{u}_p(P_p)$ should be minus a gradient of a scalar-valued function, which is not necessarily the case for the discrete $\boldsymbol{u}_{p,h\tau}^{n,k,i}$ given at time t^n.

Remark 7.4 (Error measure $\mathcal{N}^{n,k,i}$). *The error measure $\mathcal{N}^{n,k,i}$ is a natural extension of the energy error from the simpler cases. Indeed, component by component and phase by phase, it extends Theorem 4.9 from the Poisson case as well as Remark 6.6 from the steady nonlinear case. The sum $\sum_{n=1}^N (\mathcal{N}^{n,k,i})^2$ (if the initial condition is satisfied exactly) then corresponds to the square of the energy error for the heat equation, see [120, Theorem 2.1 and equation (2.7)].*

In (7.17), we have denoted the approximate component fluxes by $\boldsymbol{\theta}_{c,h\tau}^{n,k,i}$ and the approximate Darcy phase fluxes by $\boldsymbol{u}_{p,h\tau}^{n,k,i}$, anticipating that our numerical approximation will here consist in *temporal discretization* (with characteristic time step τ and discrete times t^n), *spatial discretization* (with characteristic mesh size h), *iterative linearization* (with iteration index k), and *iterative algebraic resolution* (with iteration index i). Our goal will be to *distinguish the error components* as in (2.14) in Section 2.4.6, to design an a posteriori error estimate of the form (2.23b), and to use them in the fully adaptive Algorithm 7.1.

7.5 Backward Euler time discretization and locally conservative space discretization on polytopal meshes

To discretize problem (7.1)–(7.12), we consider a fully implicit backward Euler time stepping and an abstract locally conservative space discretization on a polytopal mesh of the form of Assumption 5.1, using phase-upwinding. For all $1 \leq n \leq N$, we let $\boldsymbol{X}_{\mathcal{T}_H}^n := (\boldsymbol{X}_K^n)_{K \in \mathcal{T}_H^n}$, with, following (7.2),

$$\boldsymbol{X}_K^n := (P_K^n, (S_{p,K}^n)_{p \in \mathcal{P}}, (C_{p,c,K}^n)_{p \in \mathcal{P}, c \in \mathcal{C}_p}) \tag{7.18}$$

be the algebraic vector of discrete unknowns on an element K. System (7.1)–(7.12) is then discretized as follows: for all time steps $1 \leq n \leq N$, all polytopal cells $K \in \mathcal{T}_H^n$, and each component $c \in \mathcal{C}$, we require

$$\boxed{\frac{|K|}{\tau^n}\left(l_{c,K}(\boldsymbol{X}_K^n) - l_{c,K}(\boldsymbol{X}_K^{n-1})\right) + \sum_{\sigma \in \mathcal{F}_K \cap \mathcal{F}_H^{\mathrm{int}}} \theta_{c,K,\sigma,\uparrow}(\boldsymbol{X}_{\mathcal{T}_H}^n) = |K|q_{c,K}^n.} \tag{7.19}$$

This equation expresses the mass balance for the element K. Here $q^n_{c,K}$ is the value of the source or sink q_c on the element K and time interval I^n – recall that we consider q piecewise constant. Recall also that $|K|$ stands for the volume of the element K.

In a discrete version of (7.4), we set

$$l_{c,K}(\boldsymbol{X}^n_K) := \phi_K \sum_{p\in\mathcal{P}_c} \varsigma^n_{p,K} S^n_{p,K} C^n_{p,c,K}. \tag{7.20}$$

Similarly, for each component $c \in \mathcal{C}$, following (7.5), the total flux across the face σ results from the sum of the corresponding fluxes for each phase $p \in \mathcal{P}_c$, i.e., for all elements $K \in \mathcal{T}^n_H$ and all faces $\sigma \in \mathcal{F}_K \cap \mathcal{F}^{\text{int}}_H$ with $\sigma \subset \partial K \cap \partial L$,

$$\theta_{c,K,\sigma,\uparrow}(\boldsymbol{X}^n_{\mathcal{T}_H}) := \sum_{p\in\mathcal{P}_c} \nu^n_{p,K^\uparrow_p} C^n_{p,c,K^\uparrow_p} v_{p,K,\sigma}(\boldsymbol{X}^n_{\mathcal{T}_H}), \qquad K^\uparrow_p := \begin{cases} K & \text{if } v_{p,K,\sigma}(\boldsymbol{X}^n_{\mathcal{T}_H}) \geq 0, \\ L & \text{otherwise,} \end{cases} \tag{7.21}$$

with $C^n_{p,c,K^\uparrow_p}$ and $\nu^n_{p,K^\uparrow_p}$ denoting, respectively, the upstream molar fraction and the upstream mobility. Let us also introduce

$$\theta_{c,K,\sigma}(\boldsymbol{X}^n_{\mathcal{T}_H}) := \sum_{p\in\mathcal{P}_c} \nu^n_{p,K} C^n_{p,c,K} v_{p,K,\sigma}(\boldsymbol{X}^n_{\mathcal{T}_H}), \tag{7.22}$$

where no upwinding is used. In (7.21)–(7.22), we have employed a finite volume approximation of the normal component of the average phase velocity on face σ given by

$$v_{p,K,\sigma}(\boldsymbol{X}^n_{\mathcal{T}_H}) := v_{p,K,\sigma}(\{\boldsymbol{X}^n_{K'}\}_{K'\in\mathcal{S}_\sigma}) := \sum_{K'\in\mathcal{S}_\sigma} \tau^\sigma_{K'}(P^n_{p,K'} + \rho^n_{p,\sigma} g z_{K'}), \tag{7.23}$$

where, following (7.1),

$$P^n_{p,K} := P^n_K + P_{c_p}(S^n_{p,K}), \tag{7.24}$$

and, for all $\sigma \in \mathcal{F}_K \cap \mathcal{F}^{\text{int}}_H$, $\mathcal{S}_\sigma$ is the flux stencil collecting the elements in $\mathcal{T}^n_H$ with nonzero flux contribution ($\mathcal{S}_\sigma = \{K, L\}$ for σ such that $\sigma = \sigma_{K,L} = \partial K \cap \partial L \in \mathcal{F}^{\text{int}}_H$) in the two-point scheme (4.19b). Moreover, for all $L \in \mathcal{S}_\sigma$, $\tau^\sigma_L \in \mathbb{R}$ is the transmissibility coefficient of the face σ (if $\underline{\boldsymbol{K}} = \boldsymbol{I}$, $\tau^\sigma_K = \frac{|\sigma_{K,L}|}{d_{K,L}}$ and $\tau^\sigma_L = -\frac{|\sigma_{K,L}|}{d_{K,L}}$ for the two-point scheme (4.19b)). Please remark that the boundary fluxes are set to zero to account for the homogeneous no-flow boundary conditions (7.8).

Finally, the initial condition comes from (7.9), and the algebraic constraints (7.10)–(7.12) are trivially discretized separately on all $K \in \mathcal{T}^n_H$, namely

$$\sum_{p\in\mathcal{P}} S^n_{p,K} = 1, \tag{7.25}$$

$$\sum_{c\in\mathcal{C}_p} C^n_{p,c,K} = 1 \qquad \forall p \in \mathcal{P}. \tag{7.26}$$

7.6 Iterative linearization

At this stage, we need to solve, at each time step, the system of nonlinear algebraic equations resulting from the discretization (7.19)–(7.26). To this purpose, for all times $1 \leq n \leq N$, we apply an iterative linearization, in extension of (6.12). We consider the *Newton linearization*, as in (6.13). This generates, for an initial guess $\boldsymbol{X}^{n,0}_{\mathcal{T}_H}$ (typically given by the previous time step/initial condition), a sequence $(\boldsymbol{X}^{n,k}_{\mathcal{T}_H})_{k\geq 1}$ with $\boldsymbol{X}^{n,k}_{\mathcal{T}_H}$ solution to the following *system of linear algebraic equations*: for all components $c \in \mathcal{C}$ and all mesh elements $K \in \mathcal{T}^n_H$,

$$\sum_{K'\in\mathcal{T}^n_H} \frac{\partial R_{c,K}}{\partial \boldsymbol{X}^n_{K'}}(\boldsymbol{X}^{n,k-1}_{\mathcal{T}_H}) \cdot (\boldsymbol{X}^{n,k}_{K'} - \boldsymbol{X}^{n,k-1}_{K'}) + R_{c,K}(\boldsymbol{X}^{n,k-1}_{\mathcal{T}_H}) = 0, \tag{7.27}$$

with, for all $c \in \mathcal{C}$ and all $K \in \mathcal{T}^n_H$,

$$R_{c,K}(\boldsymbol{X}^n_{\mathcal{T}_H}) := \frac{|K|}{\tau^n}(l_{c,K}(\boldsymbol{X}^n_K) - l_{c,K}(\boldsymbol{X}^{n-1}_K)) + \sum_{\sigma\in\mathcal{F}_K\cap\mathcal{F}^{\text{int}}_H} \theta_{c,K,\sigma,\uparrow}(\boldsymbol{X}^n_{\mathcal{T}_H}) - |K|q^n_{c,K}.$$

Developing the derivatives in (7.27), we can write

$$\boxed{\frac{|K|}{\tau^n}\left(l_{c,K}^{n,k} - l_{c,K}(\boldsymbol{X}_K^{n-1})\right) + \sum_{\sigma \in \mathcal{F}_K \cap \mathcal{F}_H^{\mathrm{int}}} \theta_{c,K,\sigma,\uparrow}^{n,k} - |K|q_{c,K}^n = 0,} \tag{7.28}$$

where the linearization of the amount of component $l_{c,K}(\boldsymbol{X}_K^{n,k})$ of (7.20) is given by

$$l_{c,K}^{n,k} := l_{c,K}(\boldsymbol{X}_K^{n,k-1}) + \frac{\partial l_{c,K}}{\partial \boldsymbol{X}_K^n}(\boldsymbol{X}_K^{n,k-1}) \cdot (\boldsymbol{X}_K^{n,k} - \boldsymbol{X}_K^{n,k-1}),$$

and where $\theta_{c,K,\sigma,\uparrow}^{n,k}$ are appropriate linearizations of the component fluxes $\theta_{c,K,\sigma,\uparrow}(\boldsymbol{X}_{\mathcal{T}_H}^{n,k})$ of (7.21).

Remark 7.5 (Blending Newton and fixed-point linearizations). *In numerical experiments below in Section 7.11, we will actually define $\theta_{c,K,\sigma,\uparrow}^{n,k}$ by evaluation of the mobility $\nu_{p,K_p^\uparrow}^n$ and the molar fraction $C_{p,c,K_p^\uparrow}^n$ at the previous linearization iterate $k-1$ as*

$$\theta_{c,K,\sigma,\uparrow}^{n,k} := \sum_{p\in\mathcal{P}_c}\left\{\nu_{p,K_p^\uparrow}^{n,k-1}C_{p,c,K_p^\uparrow}^{n,k-1}\left[v_{p,K,\sigma}(\boldsymbol{X}_{\mathcal{T}_H}^{n,k-1}) + \sum_{K'\in\mathcal{T}_H^n}\frac{\partial v_{p,K,\sigma}}{\partial \boldsymbol{X}_{K'}^n}(\boldsymbol{X}_{\mathcal{T}_H}^{n,k-1}) \cdot (\boldsymbol{X}_{K'}^{n,k} - \boldsymbol{X}_{K'}^{n,k-1})\right]\right\},$$

blending thus the Newton and the fixed-point linearizations.

7.7 Iterative algebraic system solution

Now, for a given time step $1 \le n \le N$ and a given linearization iteration $k \ge 1$, (7.28) is a system of linear algebraic equations. To find its approximate solution, we use an *iterative algebraic solver*. Let $\boldsymbol{X}_{\mathcal{T}_H}^{n,k,0}$ be an initial guess. Typically, we take $\boldsymbol{X}_{\mathcal{T}_H}^{n,k,0} = \boldsymbol{X}_{\mathcal{T}_H}^{n,k-1}$. This gives a sequence $(\boldsymbol{X}_{\mathcal{T}_H}^{n,k,i})_{i\ge 1}$ solving (7.28) up to the residuals $R_{c,K}^{n,k,i}$ for all $c \in \mathcal{C}$ and all $K \in \mathcal{T}_H^n$, i.e.,

$$\boxed{\frac{|K|}{\tau^n}\left(l_{c,K}^{n,k,i} - l_{c,K}(\boldsymbol{X}_K^{n-1})\right) + \sum_{\sigma \in \mathcal{F}_K \cap \mathcal{F}_H^{\mathrm{int}}} \theta_{c,K,\sigma,\uparrow}^{n,k,i} - |K|q_{c,K}^n =: R_{c,K}^{n,k,i},} \tag{7.29}$$

cf. (6.14)–(6.15). Here

$$l_{c,K}^{n,k,i} := l_{c,K}(\boldsymbol{X}_K^{n,k-1}) + \frac{\partial l_{c,K}}{\partial \boldsymbol{X}_K^n}(\boldsymbol{X}_K^{n,k-1}) \cdot (\boldsymbol{X}_K^{n,k,i} - \boldsymbol{X}_K^{n,k-1}), \tag{7.30}$$

and, in the case of Remark 7.5,

$$\theta_{c,K,\sigma,\uparrow}^{n,k,i} := \sum_{p\in\mathcal{P}_c}\left\{\nu_{p,K_p^\uparrow}^{n,k-1}C_{p,c,K_p^\uparrow}^{n,k-1}\left[v_{p,K,\sigma}(\boldsymbol{X}_{\mathcal{T}_H}^{n,k-1}) + \sum_{K'\in\mathcal{T}_H^n}\frac{\partial v_{p,K,\sigma}}{\partial \boldsymbol{X}_{K'}^n}(\boldsymbol{X}_{\mathcal{T}_H}^{n,k-1}) \cdot (\boldsymbol{X}_{K'}^{n,k,i} - \boldsymbol{X}_{K'}^{n,k-1})\right]\right\}. \tag{7.31}$$

We will also consider $j \ge 1$ additional algebraic solver steps in (7.29), similarly as we have done it in (6.16).

7.8 Face normal fluxes and potential point values

We proceed here as in Section 6.5. In particular, in extension of (6.19), we rely on the finite volume *face normal fluxes*. These employ the approximate phase normal velocities (7.23), the total normal flux nonlinear expressions (7.22) and their upwindings (7.21), as well as the linearizations (7.31) of the latter. For all phases $p \in \mathcal{P}$, components $c \in \mathcal{C}$, time levels $n \ge 1$, mesh elements $K \in \mathcal{T}_H^n$ and their faces not lying on the boundary of Ω, $\sigma \in \mathcal{F}_K \cap \mathcal{F}_H^{\mathrm{int}}$, linearization steps $k \ge 1$, and algebraic iterations $i \ge 1$, at times $t \in I^n$ where necessary, we define

$$(\mathsf{U}_{K,p}^{t,n,k,i})_\sigma := \frac{t - t^{n-1}}{\tau^n}\sum_{K'\in\mathcal{S}_\sigma}\tau_{K'}^\sigma P_{p,K'}^{n,k,i} + \frac{t^n - t}{\tau^n}\sum_{K'\in\mathcal{S}_\sigma}\tau_{K'}^\sigma P_{p,K'}^{n-1}, \tag{7.32a}$$

$$(\Theta_{\mathrm{tm},K,c}^{t,n,k,i})_\sigma := \frac{t^n - t}{\tau^n}\left(\theta_{c,K,\sigma}(\boldsymbol{X}_{\mathcal{T}_H}^{n,k,i}) - \theta_{c,K,\sigma}(\boldsymbol{X}_{\mathcal{T}_H}^{n-1})\right), \tag{7.32b}$$

$$(\Theta_{\mathrm{upw},K,c}^{n,k,i})_\sigma := \theta_{c,K,\sigma,\uparrow}(\boldsymbol{X}_{\mathcal{T}_H}^{n,k,i}) - \theta_{c,K,\sigma}(\boldsymbol{X}_{\mathcal{T}_H}^{n,k,i}), \tag{7.32c}$$

$$(\Theta_{\mathrm{lin},K,c}^{n,k,i})_\sigma := \theta_{c,K,\sigma,\uparrow}^{n,k,i} - \theta_{c,K,\sigma,\uparrow}(\boldsymbol{X}_{\mathcal{T}_H}^{n,k,i}), \tag{7.32d}$$

$$(\Theta_{\mathrm{alg},K,c}^{n,k,i})_\sigma := \theta_{c,K,\sigma,\uparrow}^{n,k,i+j} - \theta_{c,K,\sigma,\uparrow}^{n,k,i}. \tag{7.32e}$$

All these face normal fluxes are set to zero on faces σ located on the boundary of Ω, in accordance with the no-flow boundary condition (7.8), with the exception of $(\mathsf{U}_{K,p}^{t,n,k,i})_\sigma$ that is set to $\langle \underline{\boldsymbol{K}} \rho_p g \nabla z \cdot \boldsymbol{n}_\Omega, 1 \rangle_\sigma$, in accordance with (7.6)–(7.7). Note that (7.32a) is an affine-in-time interpolation of the previous time step $P_{p,K}^{n-1}$ and the current $P_{p,K}^{n,k,i}$ values.

We also need to define the phase pressure point values $\mathsf{Z}_{p,K}^{t,n,k,i}$ and $\mathsf{Z}_{p,K}^{\mathrm{ext},t,n,k,i}$. We again proceed as in Section 6.5, using directly the cells pressure values in Definition 5.11 or 5.13. More precisely, the starting values are the phase pressure values obtained from the finite volume scheme, the current $P_{p,K}^{n,k,i}$ and the previous time step $P_{p,K}^{n-1}$, that we combine for each time $t \in I^n$ to get an affine-in-time interpolation

$$\frac{t - t^{n-1}}{\tau^n} P_{p,K}^{n,k,i} + \frac{t^n - t}{\tau^n} P_{p,K}^{n-1}. \tag{7.33}$$

Note that all quantities of (7.32) and (7.33) are *readily available* from the *finite volume discretization* of Section 7.5, more precisely on each linearization step $k \geq 1$ and each algebraic solver step $i \geq 1$ as described in Sections 7.6–7.7.

7.9 Fictitious flux and potential reconstructions

We now identify the fictitious flux and potential reconstructions, following Section 6.6.

Let a phase $p \in \mathcal{P}$, a component $c \in \mathcal{C}$, a time step $1 \leq n \leq N$, a mesh element $K \in \mathcal{T}_H^n$, a linearization iteration $k \geq 1$, and an algebraic solver iteration $i \geq 1$ be fixed. For the fictitious flux reconstructions, we lift the face normal fluxes from (7.32), using them to define the spaces $\boldsymbol{V}_{h,\mathrm{N}}^K$ in (5.16) of Definition 5.5. We define:

- The *Darcy phase fluxes reconstructions* $\boldsymbol{u}_{p,h}^{n,k,i}|_K \in \boldsymbol{\mathcal{RT}}_0(\mathcal{T}_K) \cap \boldsymbol{H}(\mathrm{div}, K)$, as a discrete counterpart of (7.7) at the discrete time t^n, using the face fluxes $\mathsf{U}_{K,p}^{t,n,k,i}$ of (7.32a) at t^n.

- The *component fluxes reconstructions* $\boldsymbol{\theta}_{c,h}^{n,k,i}|_K \in \boldsymbol{\mathcal{RT}}_0(\mathcal{T}_K) \cap \boldsymbol{H}(\mathrm{div}, K)$, as a discrete counterpart of (7.5) at the discrete time t^n, prescribed by $\theta_{c,K,\sigma}(\boldsymbol{X}_{\mathcal{T}_H}^{n,k,i})$ appearing in (7.32b).

- The *upwinding error component fluxes reconstructions* $\boldsymbol{\theta}_{\mathrm{upw},c,h}^{n,k,i}|_K \in \boldsymbol{\mathcal{RT}}_0(\mathcal{T}_K) \cap \boldsymbol{H}(\mathrm{div}, K)$ prescribed by $\Theta_{\mathrm{upw},K,c}^{n,k,i}$ of (7.32c).

- The *linearization error flux reconstruction* $\boldsymbol{\theta}_{\mathrm{lin},c,h}^{n,k,i}|_K \in \boldsymbol{\mathcal{RT}}_0(\mathcal{T}_K) \cap \boldsymbol{H}(\mathrm{div}, K)$ prescribed by $\Theta_{\mathrm{lin},K,c}^{n,k,i}$ of (7.32d).

- The *algebraic error flux reconstruction* $\boldsymbol{\theta}_{\mathrm{alg},c,h}^{n,k,i}|_K \in \boldsymbol{\mathcal{RT}}_0(\mathcal{T}_K) \cap \boldsymbol{H}(\mathrm{div}, K)$ prescribed by $\Theta_{\mathrm{alg},K,c}^{n,k,i}$ of (7.32e).

The error measure $\mathcal{N}^{n,k,i}$ of (7.17) works with approximate Darcy phase fluxes $\boldsymbol{u}_{p,h\tau}^{n,k,i}$ together with approximate component fluxes $\boldsymbol{\theta}_{c,h\tau}^{n,k,i}$. We define them as affine-in-time on the time interval I^n, given respectively by $\boldsymbol{u}_{p,h}^{n,k,i}$ and $\boldsymbol{\theta}_{c,h}^{n,k,i}$ at the current time t^n, and by $\boldsymbol{u}_{p,h}^{n-1}$ and $\boldsymbol{\theta}_{c,h}^{n-1}$ obtained in the same way but on time previous time t^{n-1}, once the linearization and algebraic iterations have been stopped. All $\boldsymbol{u}_{p,h\tau}^{n,k,i}$, $\boldsymbol{\theta}_{c,h\tau}^{n,k,i}$, and the other fluxes $\boldsymbol{\theta}_{\mathrm{upw},c,h}^{n,k,i}$, $\boldsymbol{\theta}_{\mathrm{lin},c,h}^{n,k,i}$, and $\boldsymbol{\theta}_{\mathrm{alg},c,h}^{n,k,i}$ are only used for the proof of Theorem 7.6 below and *need not be constructed in practice* (whence, recall, the naming fictitious). It namely follows from (7.29), on algebraic step $i + j$, the above flux reconstructions, and the Green theorem that there holds

$$q_{c,K}^n - \frac{l_{c,K}^{n,k,i+j} - l_{c,K}(\boldsymbol{X}_K^{n-1})}{\tau^n} - \left(\nabla \cdot (\boldsymbol{\theta}_{c,h}^{n,k,i} + \boldsymbol{\theta}_{\mathrm{upw},c,h}^{n,k,i} + \boldsymbol{\theta}_{\mathrm{lin},c,h}^{n,k,i} + \boldsymbol{\theta}_{\mathrm{alg},c,h}^{n,k,i}) \right)|_K = -|K|^{-1} R_{c,K}^{n,k,i+j} \quad \forall K \in \mathcal{T}_H^n, \tag{7.34}$$

similarly as in (6.22).

The fictitious reconstruction of the phase pressures is again achieved as in Definition 5.14, i.e., setting

$$\zeta_{p,h\tau}^{n,k,i}(\boldsymbol{a}) := \mathsf{Z}_{p,\boldsymbol{a}}^{t,n,k,i} \qquad \forall \boldsymbol{a} \in \mathcal{V}_h$$

from the potential point values identified in Section 7.8. It is piecewise affine with respect to the simplicial submeshes $\mathcal{T}_h^n$ and $H^1(\Omega)$-conforming and piecewise affine in time. In the error measure $\mathcal{N}^{n,k,i}$ of (7.17),

we also need $l_{c,h\tau}^{n,k,i}$. It is defined as affine-in-time on each time interval I^n, given respectively by the values $l_{c,K}(\boldsymbol{X}_K^{n-1})$ and $l_{c,K}(\boldsymbol{X}_K^{n,k,i})$ at times t^{n-1} and t^n, for all $K \in \mathcal{T}_H^n$. This in particular gives

$$\partial_t l_{c,h\tau}^{n,k,i}|_K = \frac{l_{c,K}(\boldsymbol{X}_K^{n,k,i}) - l_{c,K}(\boldsymbol{X}_K^{n-1})}{\tau^n} \tag{7.35}$$

on each time interval I^n.

7.10 A guaranteed a posteriori error estimate distinguishing the error components with inexpensive implementation and evaluation

Proceeding as for the steady linear problem in Section 5.9 and the steady nonlinear problem in Section 6.7 (under Assumption 5.1), we obtain a posteriori error estimates for the problem (7.1)–(7.12). We first treat the simpler case where there is no coarsening of the space meshes between the time steps and postpone the general case to Remark 7.8 below.

Theorem 7.6 (A guaranteed a posteriori error estimate distinguishing the error components with inexpensive implementation and evaluation).

For the multiphase compositional Darcy flow (7.1)–(7.12), let the weak solution $\boldsymbol{X}$ satisfy (7.2). Consider any polytopal discretization of the form (7.19), an iterative linearization (7.27) on step $k \geq 1$, and any iterative algebraic solver (7.29) on step $i \geq 1$, for a given time step $1 \leq n \leq N$. Consider $j \geq 1$ additional algebraic solver steps. Let, for each polytopal element $K \in \mathcal{T}_H^n$, the element vectors of the face normal fluxes $\Theta_{\bullet,K,c}^{n,k,i}$, $\bullet = \mathrm{upw}, \mathrm{lin}, \mathrm{alg}$, together with $\Theta_{\mathrm{tm},K,c}^{t,n,k,i}$ and $\mathsf{U}_{K,p}^{t,n,k,i}$ be given by (7.32). Let the element vectors of potential point values $\mathsf{Z}_{p,K}^{t,n,k,i}$ and $\mathsf{Z}_{p,K}^{\mathrm{ext},t,n,k,i}$ be constructed from the time-interpolated values (7.33) of $P_{p,K}^{n,k,i}$ and $P_{p,K}^{n-1}$ following Section 7.8. Consider the approximate solution $\boldsymbol{u}_{p,h\tau}^{n,k,i}, \boldsymbol{\theta}_{c,h\tau}^{n,k,i}$ as defined in Section 7.9 and the intrinsic error measure $\mathcal{N}^{n,k,i}$ of (7.17). Let finally the element matrices $\widehat{\mathbb{A}}_{\mathrm{MFE},K}$, $\widehat{\mathbb{S}}_{\mathrm{FE},K}$, and $\widehat{\mathbb{M}}_{\mathrm{FE},K}$ be respectively defined by (5.20), (5.32), and (5.33). Let $\mathcal{T}_H^n \subset \mathcal{T}_H^{n-1}$, i.e., there is no coarsening. Then

$$\mathcal{N}^{n,k,i} \leq \left\{ \sum_{c \in \mathcal{C}} \left(\eta_{\mathrm{sp},c}^{n,k,i} + \eta_{\mathrm{tm},c}^{n,k,i} + \eta_{\mathrm{lin},c}^{n,k,i} + \eta_{\mathrm{alg},c}^{n,k,i} + \eta_{\mathrm{rem},c}^{n,k,i} \right)^2 \right\}^{\frac{1}{2}} \tag{7.36}$$

with

$$\eta_{\bullet,c}^{n,k,i} := \left\{ \delta_\bullet \int_{I^n} \sum_{K \in \mathcal{T}_H^n} \left(\eta_{\bullet,K,c}^{n,k,i} \right)^2 \mathrm{d}t \right\}^{\frac{1}{2}}, \tag{7.37}$$

where $\bullet = \mathrm{sp}, \mathrm{tm}, \mathrm{lin}, \mathrm{alg}, \mathrm{rem}$ and $\delta_\bullet := 2$, except for $\delta_{\mathrm{sp}} := 4$. Here, for $c \in \mathcal{C}$, we prescribe the elementwise *spatial estimators*

$$\eta_{\mathrm{sp},K,c}^{n,k,i} := \eta_{\mathrm{upw},K,c}^{n,k,i} + \left\{ \sum_{p \in \mathcal{P}_c} \left(\eta_{\mathrm{NC},K,c,p}^{t,n,k,i} \right)^2 \right\}^{\frac{1}{2}}, \tag{7.38a}$$

with

$$\left(\eta_{\mathrm{NC},K,c,p}^{t,n,k,i} \right)^2 := \left(\nu_{p,K}^{n,k,i} C_{p,c,K}^{n,k,i} \right)^2 \left[\left(\mathsf{U}_{K,p}^{t,n,k,i} \right)^{\mathrm{t}} \widehat{\mathbb{A}}_{\mathrm{MFE},K} \mathsf{U}_{K,p}^{t,n,k,i} + \left(\mathsf{Z}_{p,K}^{t,n,k,i} \right)^{\mathrm{t}} \widehat{\mathbb{S}}_{\mathrm{FE},K} \mathsf{Z}_{p,K}^{t,n,k,i} \right.$$
$$\left. + 2 \left(\mathsf{U}_{K,p}^{t,n,k,i} \right)^{\mathrm{t}} \mathsf{Z}_{p,K}^{\mathrm{ext},t,n,k,i} - 2 \sum_{\sigma \in \mathcal{F}_K} (\mathsf{U}_{K,p}^{t,n,k,i})_\sigma |K|^{-1} \mathbf{1}^{\mathrm{t}} \widehat{\mathbb{M}}_{\mathrm{FE},K} \mathsf{Z}_{p,K}^{t,n,k,i} \right], \tag{7.38b}$$

and the *upwinding estimators*

$$\left(\eta_{\mathrm{upw},K,c}^{n,k,i} \right)^2 := \left(\Theta_{\mathrm{upw},K,c}^{n,k,i} \right)^{\mathrm{t}} \widehat{\mathbb{A}}_{\mathrm{MFE},K} \left(\Theta_{\mathrm{upw},K,c}^{n,k,i} \right) \tag{7.38c}$$

together with the *temporal estimators*

$$\left(\eta_{\mathrm{tm},K,c}^{n,k,i} \right)^2 := \left(\Theta_{\mathrm{tm},K,c}^{t,n,k,i} \right)^{\mathrm{t}} \widehat{\mathbb{A}}_{\mathrm{MFE},K} \Theta_{\mathrm{tm},K,c}^{t,n,k,i}, \tag{7.38d}$$

the *linearization estimators*

$$\left(\eta_{\mathrm{lin},K,c}^{n,k,i}\right)^2 := (\Theta_{\mathrm{lin},K,c}^{n,k,i})^{\mathrm{t}}\widehat{\mathbb{A}}_{\mathrm{MFE},K}\Theta_{\mathrm{lin},K,c}^{n,k,i} + h_K(\tau^n)^{-1}\|l_{c,K}^{n,k,i} - l_{c,K}(\boldsymbol{X}_K^{n,k,i})\|_K, \tag{7.38e}$$

the *algebraic estimators*

$$\left(\eta_{\mathrm{alg},K,c}^{n,k,i}\right)^2 := (\Theta_{\mathrm{alg},K,c}^{n,k,i})^{\mathrm{t}}\widehat{\mathbb{A}}_{\mathrm{MFE},K}\Theta_{\mathrm{alg},K,c}^{n,k,i} + h_K(\tau^n)^{-1}\|l_{c,K}^{n,k,i+j} - l_{c,K}^{n,k,i}\|_K, \tag{7.38f}$$

and the *algebraic remainder estimators* by

$$\eta_{\mathrm{rem},K,c}^{n,k,i} := h_K|K|^{-1/2}|R_{c,K}^{n,k,i+j}|. \tag{7.38g}$$

Proof. Let the time step $1 \le n \le N$, the iterative linearization step $k \ge 1$, and the iterative algebraic solver $i \ge 1$ be fixed. We bound the two building terms $\mathcal{N}_c^{n,k,i}$ and $\mathcal{N}_p^{n,k,i}$ of the error measure $\mathcal{N}^{n,k,i}$ from (7.17) separately, for each component $c \in \mathcal{C}$ and each phase $p \in \mathcal{P}$. We rely on the fictitious flux and potential reconstructions of Section 7.9 and namely on (7.34).

We first bound $\mathcal{N}_c^{n,k,i}$. Let $\varphi \in X^n$ with $\|\varphi\|_{X^n} = 1$ be fixed. We have, using (7.35), (7.34), and the Green theorem

$$(q_c, \varphi) - (\partial_t l_{c,h\tau}^{n,k,i}, \varphi) + \left(\boldsymbol{\theta}_{c,h\tau}^{n,k,i}, \nabla\varphi\right)$$

$$= \sum_{K \in \mathcal{T}_H^n}\left(q_c - \frac{l_{c,K}^{n,k,i+j} - l_{c,K}(\boldsymbol{X}_K^{n-1})}{\tau^n} - \nabla\cdot(\boldsymbol{\theta}_{c,h}^{n,k,i} + \boldsymbol{\theta}_{\mathrm{upw},c,h}^{n,k,i} + \boldsymbol{\theta}_{\mathrm{lin},c,h}^{n,k,i} + \boldsymbol{\theta}_{\mathrm{alg},c,h}^{n,k,i}), \varphi\right)_K$$

$$+ \sum_{K \in \mathcal{T}_H^n}\left(\frac{l_{c,K}^{n,k,i+j} - l_{c,K}(\boldsymbol{X}_K^{n-1})}{\tau^n} - \frac{l_{c,K}(\boldsymbol{X}_K^{n,k,i}) - l_{c,K}(\boldsymbol{X}_K^{n-1})}{\tau^n}, \varphi\right)_K$$

$$+ (\nabla\cdot(\boldsymbol{\theta}_{c,h}^{n,k,i} + \boldsymbol{\theta}_{\mathrm{upw},c,h}^{n,k,i} + \boldsymbol{\theta}_{\mathrm{lin},c,h}^{n,k,i} + \boldsymbol{\theta}_{\mathrm{alg},c,h}^{n,k,i}), \varphi) + \left(\boldsymbol{\theta}_{c,h\tau}^{n,k,i}, \nabla\varphi\right)$$

$$= -\sum_{K \in \mathcal{T}_H^n}(|K|^{-1}R_{c,K}^{n,k,i+j}, \varphi)_K + \sum_{K \in \mathcal{T}_H^n}\left(\frac{l_{c,K}^{n,k,i+j} - l_{c,K}(\boldsymbol{X}_K^{n,k,i})}{\tau^n}, \varphi\right)_K$$

$$+ \left(\boldsymbol{\theta}_{c,h\tau}^{n,k,i} - \boldsymbol{\theta}_{c,h}^{n,k,i} - \boldsymbol{\theta}_{\mathrm{upw},c,h}^{n,k,i} - \boldsymbol{\theta}_{\mathrm{lin},c,h}^{n,k,i} - \boldsymbol{\theta}_{\mathrm{alg},c,h}^{n,k,i}, \nabla\varphi\right)$$

We now treat the terms herein separately. First, the triangle inequality gives

$$\left(\boldsymbol{\theta}_{c,h\tau}^{n,k,i} - \boldsymbol{\theta}_{c,h}^{n,k,i} - \boldsymbol{\theta}_{\mathrm{upw},c,h}^{n,k,i} - \boldsymbol{\theta}_{\mathrm{lin},c,h}^{n,k,i} - \boldsymbol{\theta}_{\mathrm{alg},c,h}^{n,k,i}, \nabla\varphi\right)$$

$$\le \sum_{K \in \mathcal{T}_H^n}\left\{\|\boldsymbol{\theta}_{c,h\tau}^{n,k,i} - \boldsymbol{\theta}_{c,h}^{n,k,i}\|_{\boldsymbol{K}^{-\frac{1}{2}},K} + \|\boldsymbol{\theta}_{\mathrm{upw},c,h}^{n,k,i}\|_{\boldsymbol{K}^{-\frac{1}{2}},K}\right.$$

$$\left. + \|\boldsymbol{\theta}_{\mathrm{lin},c,h}^{n,k,i}\|_{\boldsymbol{K}^{-\frac{1}{2}},K} + \|\boldsymbol{\theta}_{\mathrm{alg},c,h}^{n,k,i}\|_{\boldsymbol{K}^{-\frac{1}{2}},K}\right\}\|\boldsymbol{K}^{\frac{1}{2}}\nabla\varphi\|_K.$$

Second, the Cauchy–Schwarz inequality gives

$$\sum_{K \in \mathcal{T}_H^n}\left(\frac{l_{c,K}^{n,k,i+j} - l_{c,K}(\boldsymbol{X}_K^{n,k,i})}{\tau^n}, \varphi\right)_K \le \sum_{K \in \mathcal{T}_H^n} h_K(\tau^n)^{-1}\|l_{c,K}^{n,k,i+j} - l_{c,K}(\boldsymbol{X}_K^{n,k,i})\|_K h_K^{-1}\|\varphi\|_K.$$

And third,

$$-\sum_{K \in \mathcal{T}_H^n}(|K|^{-1}R_{c,K}^{n,k,i+j}, \varphi)_K \le \sum_{K \in \mathcal{T}_H^n} h_K|K|^{-1/2}|R_{c,K}^{n,k,i+j}|h_K^{-1}\|\varphi\|_K.$$

Thus, regrouping these estimates and using the Cauchy–Schwarz inequality together with (7.14), we have

$$\mathcal{N}_c^{n,k,i} \le \sup_{\varphi \in X^n, \|\varphi\|_{X^n}=1}\int_{I^n}\sum_{K \in \mathcal{T}_H^n}(\eta_{\mathrm{tm},K,c}^{n,k,i} + \eta_{\mathrm{upw},K,c}^{n,k,i} + \eta_{\mathrm{lin},K,c}^{n,k,i} + \eta_{\mathrm{alg},K,c}^{n,k,i} + \eta_{\mathrm{rem},K,c}^{n,k,i})$$

$$\left\{h_K^{-2}\|\varphi\|_K^2 + \|\boldsymbol{K}^{\frac{1}{2}}\nabla\varphi\|_K^2\right\}^{\frac{1}{2}}\mathrm{d}t$$

$$\le \sup_{\varphi \in X^n, \|\varphi\|_{X^n}=1}\left\{\int_{I^n}\sum_{K \in \mathcal{T}_H^n}(\eta_{\mathrm{tm},K,c}^{n,k,i} + \eta_{\mathrm{upw},K,c}^{n,k,i} + \eta_{\mathrm{lin},K,c}^{n,k,i} + \eta_{\mathrm{alg},K,c}^{n,k,i} + \eta_{\mathrm{rem},K,c}^{n,k,i})^2\,\mathrm{d}t\right\}^{\frac{1}{2}}\underbrace{\|\varphi\|_{X^n}}_{=1}, \tag{7.39}$$

where, for all components $c \in \mathcal{C}$ and all $K \in \mathcal{T}_H^n$,

$$\eta_{\mathrm{tm},K,c}^{n,k,i} := \|\boldsymbol{\theta}_{c,h\tau}^{n,k,i} - \boldsymbol{\theta}_{c,h}^{n,k,i}\|_{\underline{\boldsymbol{K}}^{-\frac{1}{2}},K}, \tag{7.40a}$$

$$\eta_{\mathrm{upw},K,c}^{n,k,i} := \|\boldsymbol{\theta}_{\mathrm{upw},c,h}^{n,k,i}\|_{\underline{\boldsymbol{K}}^{-\frac{1}{2}},K}, \tag{7.40b}$$

$$\eta_{\mathrm{lin},K,c}^{n,k,i} := \|\boldsymbol{\theta}_{\mathrm{lin},c,h}^{n,k,i}\|_{\underline{\boldsymbol{K}}^{-\frac{1}{2}},K} + h_K(\tau^n)^{-1}\|l_{c,K}^{n,k,i} - l_{c,K}(\boldsymbol{X}_K^{n,k,i})\|_K, \tag{7.40c}$$

$$\eta_{\mathrm{alg},K,c}^{n,k,i} := \|\boldsymbol{\theta}_{\mathrm{alg},c,h}^{n,k,i}\|_{\underline{\boldsymbol{K}}^{-\frac{1}{2}},K} + h_K(\tau^n)^{-1}\|l_{c,K}^{n,k,i+j} - l_{c,K}^{n,k,i}\|_K, \tag{7.40d}$$

$$\eta_{\mathrm{rem},K,c}^{n,k,i} := h_K|K|^{-1/2}|R_{c,K}^{n,k,i+j}|. \tag{7.40e}$$

We now bound $\mathcal{N}_p^{n,k,i}$. Let

$$\eta_{\mathrm{NC},K,c,p}^{t,n,k,i} := \nu_{p,K}^{n,k,i} C_{p,c,K}^{n,k,i}\|\boldsymbol{u}_{p,h\tau}^{n,k,i} + \underline{\boldsymbol{K}}\nabla\zeta_{p,h\tau}^{n,k,i}\|_{\underline{\boldsymbol{K}}^{-\frac{1}{2}},K}. \tag{7.40f}$$

We have

$$\mathcal{N}_p^{n,k,i} \leq \left\{ \sum_{c \in \mathcal{C}_p} \int_{I^n} \left\{ \sum_{K \in \mathcal{T}_H^n} \left(\eta_{\mathrm{NC},K,c,p}^{t,n,k,i}\right)^2 \right\} \mathrm{d}t \right\}^{\frac{1}{2}}. \tag{7.41}$$

We now remark that the estimators in (7.40a)–(7.40f) and (7.38b)–(7.38g) are indeed the same, proceeding as in Theorem 5.18 to evaluate the nonconformity estimator given by (7.40f) and by applying Lemma 5.8 to evaluate the upwinding, temporal, linearization, and algebraic estimators. Regrouping the spatial discretization estimators

$$\eta_{\mathrm{sp},K,c}^{n,k,i} := \eta_{\mathrm{upw},K,c}^{n,k,i} + \left\{ \sum_{p \in \mathcal{P}_c} \left(\eta_{\mathrm{NC},K,c,p}^{t,n,k,i}\right)^2 \right\}^{\frac{1}{2}} \tag{7.42}$$

and all the estimators per component $c \in \mathcal{C}$, using the Cauchy–Schwarz inequality, and recalling (7.17), we obtain (7.36). □

Remark 7.7 (Sharper estimates). *Sharper estimates, namely avoiding the factors $\delta_\bullet$, (but not identifying/regrouping the error component estimators) immediately follow from (7.39) and (7.41).*

Remark 7.8 (Coarsening). *If a mesh element $K \in \mathcal{T}_H^n$ has been obtained by coarsening of elements of the previous mesh $\mathcal{T}_H^{n-1}$ collected in the set $\mathcal{T}_{H,K}^{n-1}$, then the estimators $\eta_{\mathrm{NC},K,c,p}^{t,n,k,i}$ of (7.38b) and $\eta_{\mathrm{tm},K,c}^{n,k,i}$ of (7.38d) need to be evaluated using the matrices $\widehat{\mathbb{A}}_{\mathrm{MFE},K}$, $\widehat{\mathbb{S}}_{\mathrm{FE},K}$, and $\widehat{\mathbb{M}}_{\mathrm{FE},K}$ stemming from a common refinement of the element K by the simplicial meshes $\mathcal{T}_h^{n-1}|_{\mathcal{T}_{H,K}^{n-1}}$ and $\mathcal{T}_K$ and not $\mathcal{T}_K$ solely. In this sense, Theorem 7.6 also holds when $\mathcal{T}_H^n \not\subset \mathcal{T}_H^{n-1}$.*

7.11 Numerical experiments

The proposed a posteriori error estimate framework summarized in Theorem 7.6 and Algorithm 7.1 has been implemented in a reservoir prototype simulator [223], a thermal multi-purpose simulator written in C++, which is a part of the next generation IFPEn research simulators based on the *Arcane* framework [141]. The execution platform is a public computer Intel Core i7, 8 cores, 3.7Ghz with 16GB of memory.

7.11.1 Setup

We study two different test cases. The first one is taken from [89], relying on the tenth SPE comparative solution project model. It is an incompressible water-oil two-phase flow problem built on a Cartesian regular geometry. This test corresponds to the layer 85 of SPE10. We choose here first the Cartesian regular mesh so as to compare our new approach with approaches already validated on this type of meshes. The second case is a simulation of a black-oil model. For this test, the problem is built on a three-dimensional corner-point geometry (distorted grids), well-known and most often used in reservoir simulation due to the flexibility allowing a good representation of reservoir description, see [105] and the references therein.

In our computational experiments presented below, we perform a nonconforming-type mesh refinement and allow one hanging node per cell face for the space mesh adaptation. The resulting mesh is a polytopal
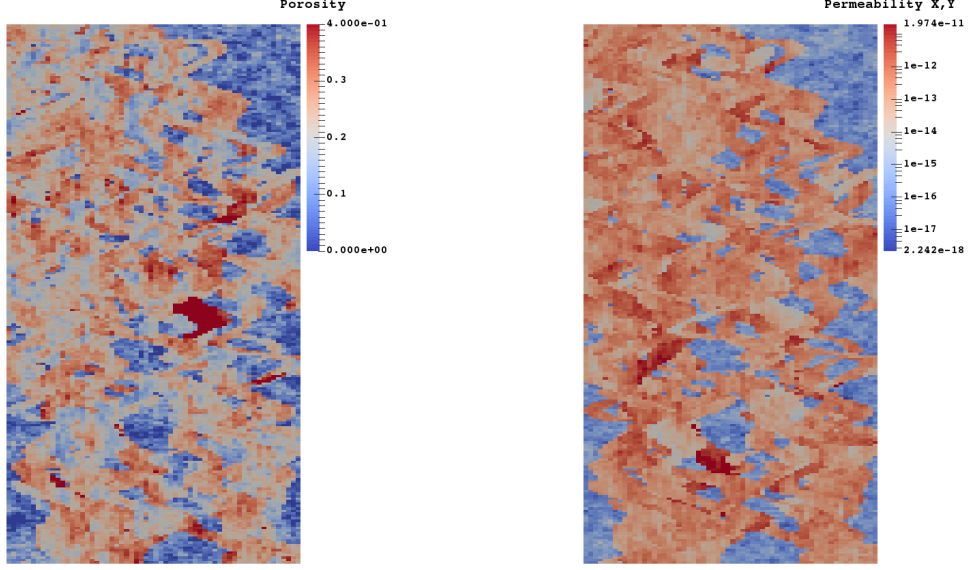
Figure 30: [Section 7.11.2] Porosity (*left*) and permeability (*right*), layer 85 of the 10th SPE case

mesh in the sense of Section 3.6. Coarsening is then obtained by retrogressive cell agglomeration. During the coarsening and refinement phases, we use some upscaling and interpolating mechanisms to ensure the mass balance. Geological data (rock properties like porosity and permeability) are stored at the finest mesh level as data on a cloud of space points. These data are used during mesh adaptation by standard upscaling algorithms to compute the values of the rock model properties on the newly created cells following [185].

7.11.2 Two-phase Darcy flow

We consider a two-dimensional spatial domain discretized by a grid of 60×220 rectangular cells of size 6.096m in the x direction and 3.048m in the y direction. We choose the initial time step as $\tau^0 = 4.32 \times 10^4$s, which equals to 0.5 days, and the process is simulated to $t_F = 2000$ days. The reservoir is initially saturated with hydrocarbons and we consider the injection of water by a well located at the center of the grid. Four production wells are placed at the four corners of the domain. Therefore, we have a water component W and an oil component O collected in the set of components $\mathcal{C} = \{W, O\}$ and two phases $\mathcal{P} = \{w, o\}$ corresponding to water and oil. The model is actually simplified in that the components can be identified with the phases, so that $\mathcal{P}_W = \{w\}$ and $\mathcal{P}_O = \{o\}$, $C_{w,W} = C_{o,O} = 1$, $C_{o,W} = C_{w,O} = 0$, and the vector of unknowns $\boldsymbol{X}$ reduces to $(P, (S_p)_{p \in \mathcal{P}})$. Note that then the accumulation term l_c becomes linear in the only unknown S_p, and one particular consequence is that the second term in the definition (7.38e) of $\eta_{\text{lin},K,c}^{n,k,i}$ vanishes. The porosity ϕ and the permeability field $\underline{\boldsymbol{K}}$ (scalar coefficient times an identity matrix) are shown in Figure 30. The other parameters of Section 7.1 are given by (see [89] for more details):

- $\mu_o = 10^{-3}$ Pa·s and $\mu_w = 0.3 \cdot 10^{-3}$ Pa·s,

- $\varsigma_o = \varsigma_w = 1$ mole·m^{-3},

- $S_{wr} = S_{or} = 0.2$,

-
$$k_{r,w}(S_w) = \left(\frac{S_w - S_{wr}}{1 - S_{wr} - S_{or}} \right)^2 \qquad \text{and} \qquad k_{r,o}(S_o) = \left(\frac{S_o - S_{or}}{1 - S_{wr} - S_{or}} \right)^2,$$

- $P_{c_p}(S_p) = 0$,

- there is no gravitational force, $z = 0$, so that the mass densities ρ_p need not be specified.

We consider a discretization by the implicit multi-point finite volume scheme of Section 7.5 with the linearization detailed in Section 7.6, using the blending of Newton and fixed-point of Remark 7.5. For

(a) Water saturation   (b) Cartesian estimate of [262]   (c) Polygonal simplified estimate (7.38a)
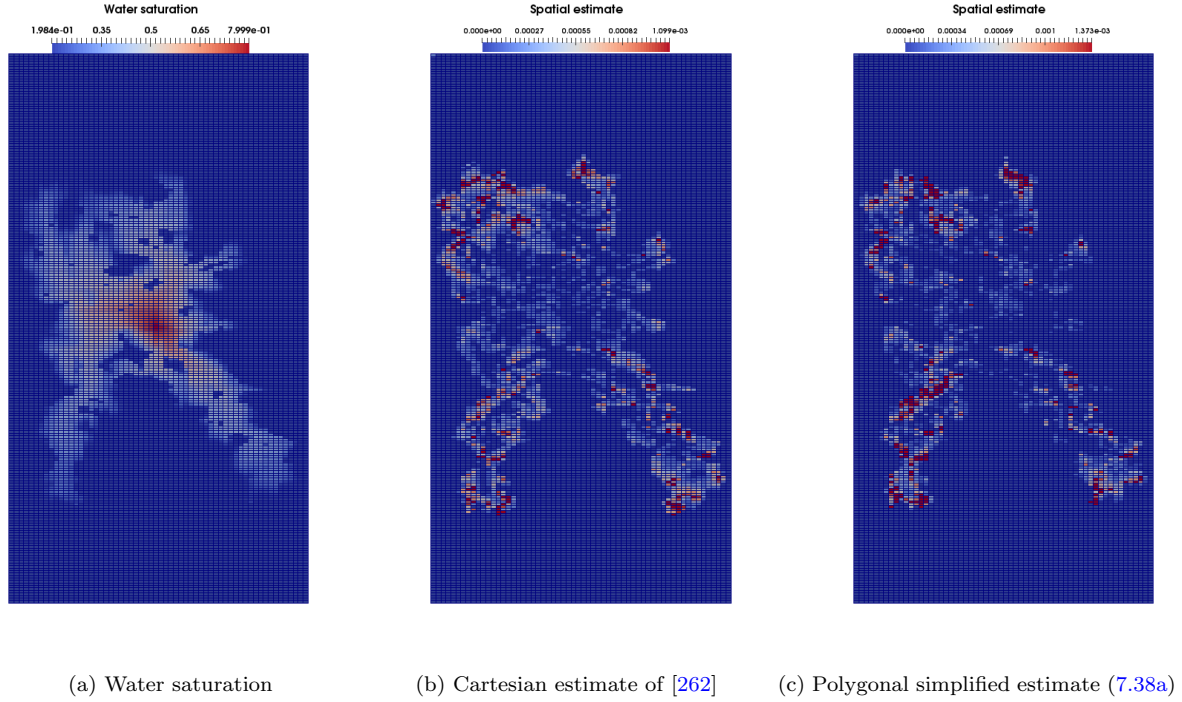
Figure 31: [Section 7.11.2] Results on the fine mesh at 500 days

the linear solver needed in Section 7.7, we use the Bi-Conjugated Gradient Stabilized (BiCGStab) [239] with an ILU{0} preconditioner.

Figure 31 shows, at 500 days of the simulation, the evolution of the approximate water saturation, the spatial estimator computed by the formula on Cartesian grid proposed in [262, Chapter 4], and the spatial estimator given by (7.38a). Similar behavior is observed for the two a posteriori estimators. Furthermore, we see that they both detect well the error following the saturation front despite the strong heterogeneity of the domain.

The previous result motivates an adaptive mesh refinement/coarsening (AMR) strategy based on these estimators. We now verify that the simple a posteriori estimators on polyhedral mesh of Theorem 7.6 give the expected results, while comparing it with the already validated case in [262, Chapter 4]. We focus on the mesh space adaptivity and we do not act on the time step. Actually, in our simulator, the time step size is controlled by the behavior of the Newton/fixed-point algorithm (increased systematically in general, by multiplying by 2, and divided by 2 if the Newton/fixed-point algorithm diverges), so the space-mesh adaptation only will enable an easy comparison with the non-adaptive code. We first apply Algorithm 7.1 with $\delta_{\mathrm{ref}} = 0.7, \delta_{\mathrm{deref}} = 0.2$, and "exact" (i.e., negligible-error) algebraic and linearization solvers. On the coarse scale, the domain is discretized by a grid of 30×110 cells and we allow one refinement level. Figure 32 shows the evolution of the approximate water saturation and of the meshes at two different simulation times. We remark that the refinement follows the saturation front as times evolves. Additionally, since we have a model with highly heterogeneous permeability, we are lead to perform a slow derefinement process in the zone abandoned by the front of water saturation. Figure 33 depicts the cumulated oil rate (left) and the water-cut[2] (right) during the simulation. We compare there the results on the fine grid, the results on the coarse grid, and the results of the two AMR strategies. We remark that the accuracy of the results on the fine grid is almost recovered by the AMR strategy, and appears much better than the coarse-grid result.

The details of the efficiency of the AMR strategy based on the a posteriori error estimator can finally be appreciated in Table 1. We compare the global CPU time of the simulation for the different strategies. We detail the CPU time spent on the evaluation of the estimators and the mesh adaptation. We remark a low cost of the estimators evaluation compared with the total computation CPU time, thanks to the use of the simple and fast-to-evaluate form of a posteriori error estimates on polygonal meshes of Theorem 7.6. In Table 1, when applying the mesh adaptation, the overall CPU time is the sum of the resolution time, AMR time, and estimators evaluation time. We remark that applying the AMR strategy on this two-

---

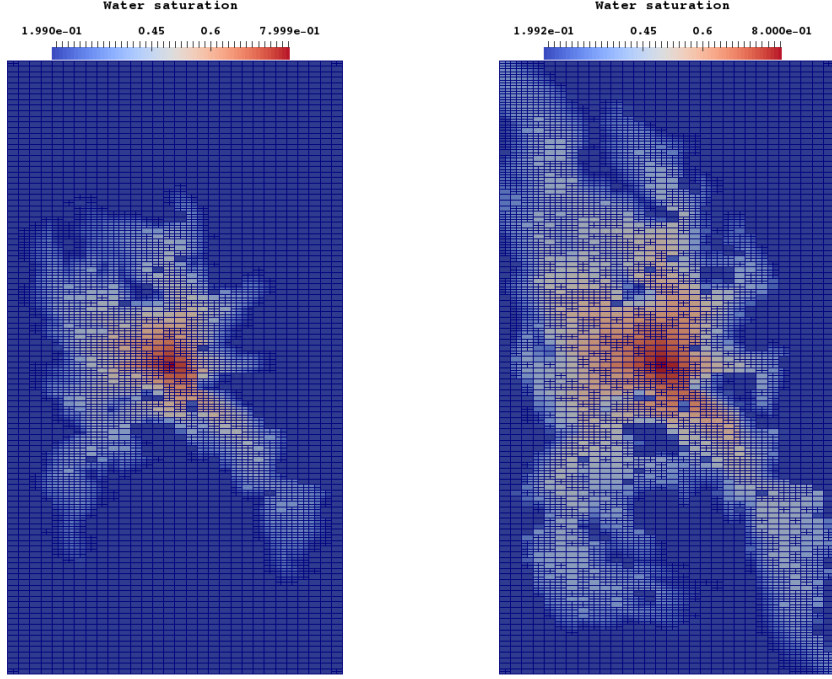[2]The ratio of water produced compared to the volume of total liquids produced.

Figure 32: [Section 7.11.2] Results on the adaptive mesh at 400 days and 1100 days
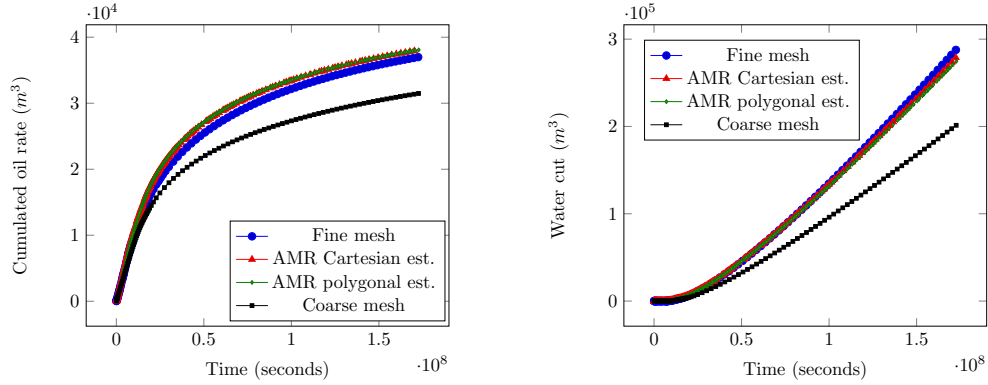

Figure 33: [Section 7.11.2] Fine mesh, coarse mesh, and adaptive meshes: cumulated oil rate (*left*) and water cut (*right*)

dimensional test case leads to a gain factor in the overall CPU time at around 2, for both the polygonal estimate proposed in this paper and the Cartesian estimate already validated on this type of meshes. Note that, however, that the simulation is slightly faster when adapting using the Cartesian estimate, due to the fact that we can here directly compute the $\mathcal{RT}$ basis functions on the rectangular cells.

-	Resolution	AMR	Estimators evaluation	Gain factor
Fine	603s	-	-	-
AMR Cartesian est.	229s	39s	19s	2.1
AMR polygonal est.	242s	46s	27s	1.9

Table 1: [Section 7.11.2] Fine grid vs. adaptive mesh refinements
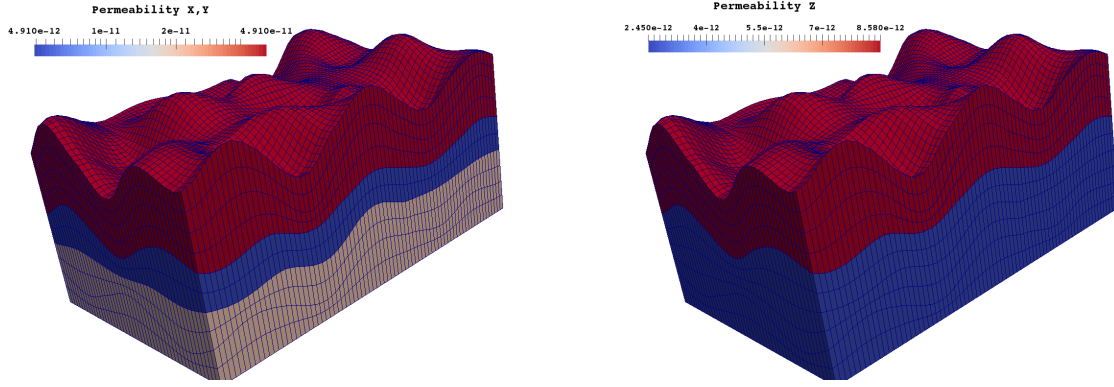
Figure 34: [Section 7.11.3] Permeability

7.11.3 Three-phases, three-components Darcy flow

In this section, we present a simulation of a black-oil model. Here, we have three phases constituted by water, oil, and gas, represented by lowercase letters w, o, g as indices, respectively. The oil phase contains two types of components: nonvolatile oil and volatile oil, which we call here oil component and gas component, respectively. This is due to the fact that in this model, the hydrocarbon components are divided into light and heavy components. The light component can dissolve into the liquid oil phase or volatilize in the gas phase according to the pressure and temperature. The gas phase only contains the gas components and the water phase only contains the water component. The components are represented by uppercase letters W for the water component, O for the oil component, and G for the gas component. Therefore, we have a problem with three phases $\mathcal{P} = \{w, o, g\}$ and three components $\mathcal{C} = \{W, O, G\}$.

Model setting

The reservoir considered in this test case is a 3-dimensional domain $\Omega := 4750\text{m} \times 3000\text{m} \times 114\text{m}$ discretized by a corner-point geometry grid. We consider a heterogeneous anisotropic reservoir with porosity 0.3 and permeability $\boldsymbol{K}$ in a form of a diagonal matrix with the x and y components identical and forming three horizontal layers, and the z component forming two horizontal layers, see Figure 34. We consider a gas injection in a reservoir initially unsaturated. A vertical gas injection well perforates a corner of the reservoir in the z direction and a production well is located in the opposite corner. On the fine scale, the domain is discretized by a grid of $76 \times 48 \times 10$ elements and on the coarse scale by a grid of $38 \times 24 \times 5$ cells, leading to one refinement level. The process is simulated to $t_F = 2000$ days with initial time step $\tau^0 = 4.32 \times 10^4$s, which equals to 0.5 days. Data, constraints, and pressure-volume properties are adapted from the first SPE comparative solution project model (SPE1) designed to simulate a three-dimensional black-oil reservoir, given in [201, Tables 1,2, and 3]. We consider a discretization by the implicit multi-point finite volume scheme of Section 7.5 with the blended Newton and fixed-point linearization detailed in Section 7.6. For the linear solver of Section 7.7, we again use the BiCGStab [239] with an ILU{0} preconditioner. Figure 35 shows the evolution of the gas saturation and of the spatial estimator at 1000 days. Note that, for the spatial estimator, the data are normalized by max value in order to have a $[0, 1]$ range. We observe that the spatial estimator follows the saturation front though the heterogeneous anisotropic medium with time evolution.

Adaptive space mesh refinement and stopping criteria for the linear solver

In the standard resolution of our reservoir prototype simulator, the chosen grid is the fine-scale one and the initial time step, as mentioned before, is chosen as $\tau^0 = 4.32 \times 10^4$s. As in the test of Section 7.11.2, the time step is increased systematically, by multiplying by 2, but also controlled by the convergence of the linearization loop in such a way that we divide it by 2 if a divergence of the linearization algorithm occurs. We thus stick to this setting and only focus on the stopping criteria for the linear solver (and not for the linearization one) and on mesh adaptivity in space (and not in time). For the adaptive resolution, we start on the coarse-level grid allowing to one refinement level and we fix $\delta_{\text{ref}} = 0.7, \delta_{\text{deref}} = 0.2$ in Algorithm 7.1.
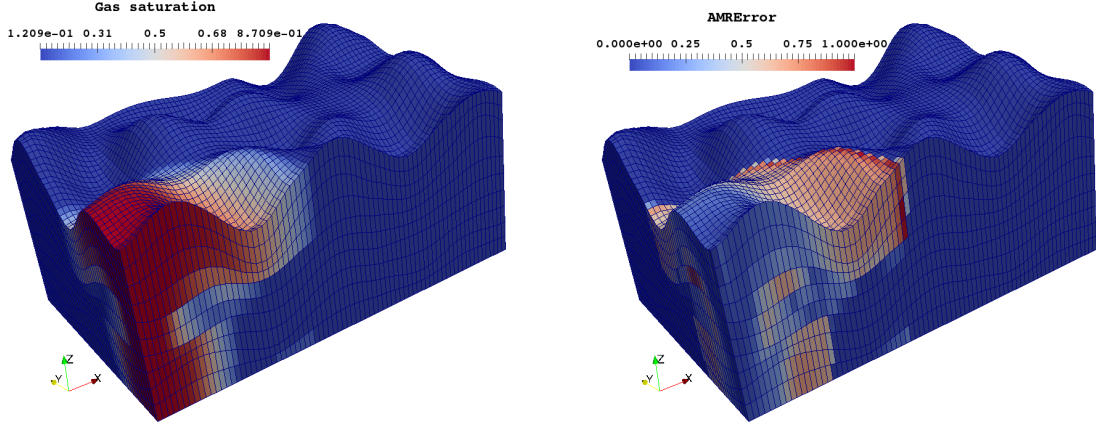
Figure 35: [Section 7.11.3] Results on the fine mesh at 1000 days, gas saturation (*left*) and normalized polygonal estimate (*right*)
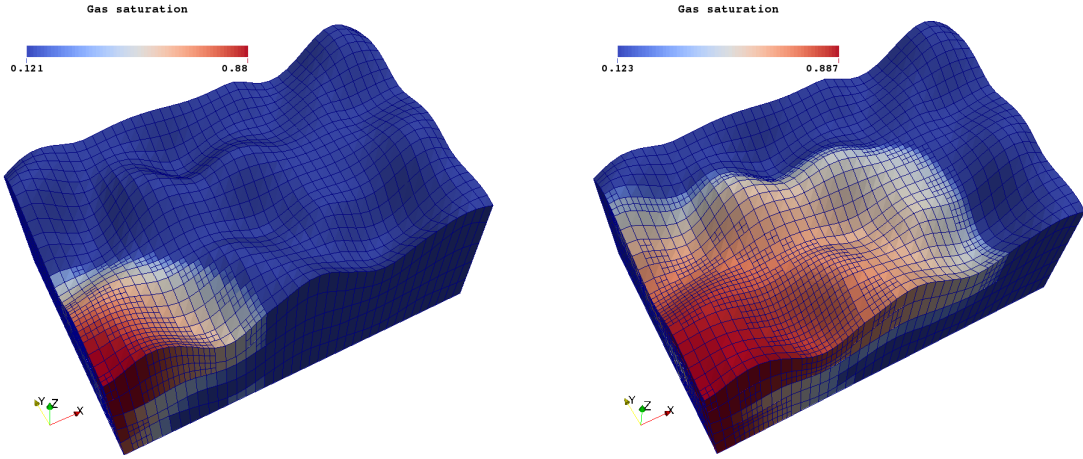

Figure 36: [Section 7.11.3] Gas saturation: results on the adaptive mesh at 500 days (*left*) and 1500 days (*right*)

Figure 36 illustrates the evolution of the approximate gas saturation at two different time steps. We remark that the refinement follows the saturation front as times evolves. Additionally, the fact that the light component G (gas component) can dissolve into the liquid oil phase or volatilize in the gas phase, we are lead to perform some localized refinement in zones abandoned by the front of gas saturation.

We show in the left part of Figure 37, at a fixed time of 500 days and for the first linearization iteration, the evolution of the total estimator $\left(\eta_{\mathrm{sp},c}^{n,k,i} + \eta_{\mathrm{tm},c}^{n,k,i} + \eta_{\mathrm{lin},c}^{n,k,i} + \eta_{\mathrm{alg},c}^{n,k,i} + \eta_{\mathrm{rem},c}^{n,k,i}\right)$ with $c = \mathrm{G}$ the gas component and the estimators given in (7.37)–(7.38), the algebraic estimator $\eta_{\mathrm{alg},c}^{n,k,i}$ of (7.37), (7.38f) (with $j = 20$ additional steps), and the relative algebraic residual given by

$$\mathrm{err}_{\mathrm{alg}}^{n,k,i} := \frac{\|\mathbb{A}^{n,k-1}\mathsf{X}^{n,k,i} - \mathsf{B}^{n,k-1}\|}{\|\mathsf{B}^{n,k-1}\|}$$

with $\mathbb{A}^{n,k-1}\mathsf{X}^{n,k} = \mathsf{B}^{n,k-1}$ being the linear system resulting from the k-th iteration of the blended Newton/fixed-point linearization at time step t^n, see Section 7.6. For the standard resolution, we stop the algebraic iteration using a fixed threshold $\mathrm{err}_{\mathrm{alg}}^{n,k,i} \leq 10^{-6}$, following our usual practice. In the adaptive resolution based on Algorithm 7.1, we fix $\gamma_{\mathrm{alg}} = 10^{-2}$.
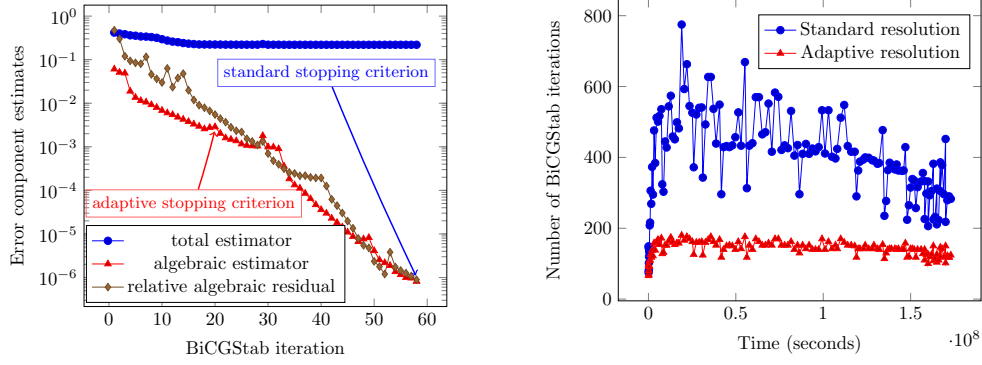
Figure 37: [Section 7.11.3] Standard resolution vs. adaptive resolution: total estimator and its algebraic component (*left*) and number of BiCGStab iterations per time step (*right*)
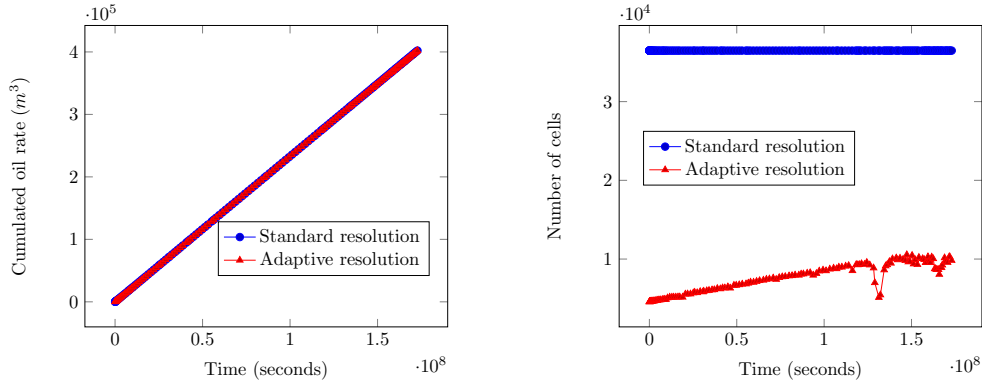

Figure 38: [Section 7.11.3] Standard resolution vs. adaptive resolution: cumulated oil rate (*left*) and number of cells (*right*)

We remark, in the left part of Figure 37, that the algebraic estimator steadily decreases, while the total estimator almost stagnates after about a third of total number of iterations necessary to converge using the standard stopping criterion. In the right part of Figure 37, we depict the cumulated number of BiCGStab iterations at each time step (the sum of the necessary number of BiCGStab iterations at each linearization iteration of the time step). We observe a significant gain with the adaptive stopping criterion.

We compare in Figure 38 the number of cells and the cumulated rate of oil production resulting from both the standard and the adaptive resolution. We can observe in the left part of Figure 38 that the adaptive algorithm does not have any significant influence on the accuracy of production. The right part of Figure 38 shows an important reduction in the number of cells via the adaptive resolution compared with the standard one.

The necessary linear solver steps and the different CPU times of the adaptive and standard resolutions are collected in Table 2. It shows that the total number of linear solver steps is reduced by 70%, which is an important gain. An important gain is also observed in the CPU time. We in particular remark that the time spent for the adaptation (AMR and estimators evaluation) only represents a small part of the actual resolution time. Globally, a reduction factor of 3.8 of the overall CPU time is obtained by comparing the adaptive resolution with the standard one. Important increase of this reduction factor is still to be expected when also the linearization solver stopping criteria and time step adaptation are used.

	Linear solver steps	Resolution time	AMR time	Estimators evaluation	Gain factor
Standard resolution	66386	1023s	-	-	-
Adaptive resolution	20184	201s	42s	26s	3.8

Table 2: [Section 7.11.3] Comparison between standard and adaptive resolutions

7.12 Bibliographic resources

Problem (7.1)–(7.12) is highly involved. For this reason, the existence and uniqueness of the weak solution is supposed in Assumption 7.2. For simplified problems, existence and uniqueness of the weak solution together with existence, uniqueness, and a priori error estimates for numerical discretizations have been derived in Alt and Luckhaus [16], Alt *et al.* [17], Kröner and Luckhaus [169], Otto [207, 208], Knabner and Otto [166], Chen [81]–[82] and Chen and Ewing [84], Smaï [231], Caro *et al.* [72], Cao and Pop [71], Ruiz-Baier and Lunati [226], Feireisl *et al.* [133], Kou *et al.* [167], Murphy and Walkington [194], Beneš [46], and Brenner *et al.* [59, 60], see also the references therein. Some classical books on the subject are those of Bear and Bachmat [35], Chavent and Jaffré [78], and Chen *et al.* [86].

For model (unsteady nonlinear) advection–diffusion–reaction problems, a posteriori analyses have been initiated in Eriksson and Johnson [117], Süli [234, 235], Angermann *et al.* [21], Verfürth [241, 240, 244], Kröner and Ohlberger [170], Ohlberger [203, 202], Ohlberger and Rohde [204], Kröner *et al.* [168], Nicaise [197], and Ern *et al.* [122], see also the references therein. A detailed attention has been paid to time-evolution problems of parabolic type in order to derive a posteriori error estimates robust with respect to the final time. This has been achieved in Verfürth [242, 243, 238], see also Bergam *et al.* [47], Ern and Vohralík [123], Hilhorst and Vohralík [151], Georgoulis *et al.* [137, 138, 139]. Lately, locally space–time efficient a posteriori error estimates robust with respect to the final time were obtained in Ern *et al.* [120], see also Ern *et al.* [121] and Mitra and Vohralík [190].

Finally, for results on a posteriori error estimates and adaptivity for numerical discretizations of porous media flows and namely multiphase compositional flows, we refer to Saad and Zhang [228], Chen and Ewing [85], Jenny *et al.* [153], Chen and Liu [80], Klieber and Rivière [165], Chueh *et al.* [91, 90], Vohralík and Wheeler [253], El Hassouni and Mghazli [115], Bernardi *et al.* [51], Cancès *et al.* [70], Di Pietro *et al.* [103, 104], Faigle *et al.* [131, 130], Henning *et al.* [148], Buhr *et al.* [67], Ganis *et al.* [135], Beaude *et al.* [36], Ahmed *et al.* [10, 9], and Li *et al.* [179], Li and Wheeler [181], Ben Gharbia *et al.* [44, 45], Deucher *et al.* [98], see also the references therein.

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** (1998), 1700–1716.

[2] Aavatsmark, I., Eigestad, G. T., Klausen, R. A., Wheeler, M. F., and Yotov, I. Convergence of a symmetric MPFA method on quadrilateral grids. *Comput. Geosci.* **11** (2007), 333–345. http://dx.doi.org/10.1007/s10596-007-9056-8.

[3] Achdou, Y., Bernardi, C., and Coquel, F. A priori and a posteriori analysis of finite volume discretizations of Darcy's equations. *Numer. Math.* **96** (2003), 17–42. http://dx.doi.org/10.1007/s00211-002-0436-7.

[4] Ács, G., Doleschall, S., and Farkas, É. General purpose compositional model. *Society of Petroleum Engineers Journal* **25** (1985), 543–553. https://doi.org/10.2118/10515-PA.

[5] Adams, R. A. Pure and Applied Mathematics, Vol. 65. *Sobolev spaces.* Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], New York-London, 1975.

[6] Afif, M., Bergam, A., Mghazli, Z., and Verfürth, R. A posteriori estimators for the finite volume discretization of an elliptic problem. *Numer. Algorithms* **34** (2003), 127–136. International Conference on Numerical Algorithms, Vol. II (Marrakesh, 2001).

[7] Agouzal, A., Baranger, J., Maître, J.-F., and Oudin, F. Connection between finite volume and mixed finite element methods for a diffusion problem with nonconstant coefficients. Application to a convection diffusion problem. *East-West J. Numer. Math.* **3** (1995), 237–254.

[8] Agouzal, A., and Oudin, F. A posteriori error estimator for finite volume methods. *Appl. Math. Comput.* **110** (2000), 239–250.

[9] Ahmed, E., Nordbotten, J. M., and Radu, F. A. Adaptive asynchronous time-stepping, stopping criteria, and a posteriori error estimates for fixed-stress iterative schemes for coupled poromechanics problems. *J. Comput. Appl. Math.* **364** (2020), 112312, 25. https://doi.org/10.1016/j.cam.2019.06.028.

[10] Ahmed, E., Radu, F. A., and Nordbotten, J. M. Adaptive poromechanics computations based on a posteriori error estimates for fully mixed formulations of Biot's consolidation model. *Comput. Methods Appl. Mech. Engrg.* **347** (2019), 264–294. https://doi.org/10.1016/j.cma.2018.12.016.

[11] Ahusborde, E., Azaïez, M., Ben Belgacem, F., and Bernardi, C. Automatic simplification of Darcy's equations with pressure dependent permeability. *ESAIM Math. Model. Numer. Anal.* **47** (2013), 1797–1820. http://dx.doi.org/10.1051/m2an/2013089.

[12] Ainsworth, M. Robust a posteriori error estimation for nonconforming finite element approximation. *SIAM J. Numer. Anal.* **42** (2005), 2320–2341. https://doi.org/10.1137/S0036142903425112.

[13] Ainsworth, M. A posteriori error estimation for lowest order Raviart–Thomas mixed finite elements. *SIAM J. Sci. Comput.* **30** (2007), 189–204. http://link.aip.org/link/?SCE/30/189/1.

[14] Ainsworth, M., and Oden, J. T. *A posteriori error estimation in finite element analysis.* Pure and Applied Mathematics (New York). Wiley-Interscience [John Wiley & Sons], New York, 2000.

[15] Allaire, G. *Analyse numérique et optimisation.* Editions Ecole Polytechnique, Palaiseau, France, 2005.

[16] Alt, H. W., and Luckhaus, S. Quasilinear elliptic-parabolic differential equations. *Math. Z.* **183** (1983), 311–341. http://dx.doi.org/10.1007/BF01176474.

[17] Alt, H. W., Luckhaus, S., and Visintin, A. On nonstationary flow through porous media. *Ann. Mat. Pura Appl. (4)* **136** (1984), 303–316. http://dx.doi.org/10.1007/BF01773387.

[18] Amaziane, B., Bergam, A., El Ossmani, M., and Mghazli, Z. A posteriori estimators for vertex centred finite volume discretization of a convection-diffusion-reaction equation arising in flow in porous media. *Internat. J. Numer. Methods Fluids* **59** (2009), 259–284.

[19] Amaziane, B., Jurak, M., and Žgaljić Keko, A. An existence result for a coupled system modeling a fully equivalent global pressure formulation for immiscible compressible two-phase flow in porous media. *J. Differential Equations* **250** (2011), 1685–1718. http://dx.doi.org/10.1016/j.jde.2010.09.008.

[20] Angermann, L. Balanced a posteriori error estimates for finite-volume type discretizations of convection-dominated elliptic problems. *Computing* **55** (1995), 305–323. https://doi.org/10.1007/BF02238485.

[21] Angermann, L., Knabner, P., and Thiele, K. An error estimator for a finite volume discretization of density driven flow in porous media. *Appl. Numer. Math.* **26** (1998), 179–191. Proceedings of the International Centre for Mathematical Sciences Conference on Grid Adaptation in Computational PDEs: Theory and Applications (Edinburgh, 1996).

[22] Antonietti, P. F., Beirão da Veiga, L., Lovadina, C., and Verani, M. Hierarchical a posteriori error estimators for the mimetic discretization of elliptic problems. *SIAM J. Numer. Anal.* **51** (2013), 654–675. http://dx.doi.org/10.1137/120873157.

[23] Antontsev, S. N., Kazhikhov, A. V., and Monakhov, V. N. Translated from the Russian. *Boundary value problems in mechanics of nonhomogeneous fluids*, vol. **22** of *Studies in Mathematics and its Applications.* North-Holland Publishing Co., Amsterdam, 1990.

[24] Arioli, M. A stopping criterion for the conjugate gradient algorithms in a finite element method framework. *Numer. Math.* **97** (2004), 1–24. http://dx.doi.org/10.1007/s00211-003-0500-y.

[25] Arioli, M., Georgoulis, E. H., and Loghin, D. Stopping criteria for adaptive finite element solvers. *SIAM J. Sci. Comput.* **35** (2013), A1537–A1559. http://dx.doi.org/10.1137/120867421.

[26] Arioli, M., Liesen, J., Miȩdlar, A., and Strakoš, Z. Interplay between discretization and algebraic computation in adaptive numerical solution of elliptic PDE problems. *GAMM-Mitt.* **36** (2013), 102–129. http://dx.doi.org/10.1002/gamm.201310006.

[27] Axelsson, O., and Kaporin, I. Error norm estimation and stopping criteria in preconditioned conjugate gradient iterations. *Numer. Linear Algebra Appl.* **8** (2001), 265–286.

[28] Aziz, K., and Settari, A. *Petroleum Reservoir Simulation.* Applied Science Publishers, Ltd, London, 1979.

[29] Babuška, I., and Strouboulis, T. *The finite element method and its reliability.* Numerical Mathematics and Scientific Computation. The Clarendon Press Oxford University Press, New York, 2001.

[30] Bakushinsky, A., and Smirnova, A. A posteriori stopping rule for regularized fixed point iterations. *Nonlinear Anal.* **64** (2006), 1255–1261. http://dx.doi.org/10.1016/j.na.2005.06.031.

[31] Bangerth, W., and Rannacher, R. *Adaptive finite element methods for differential equations.* Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 2003. https://doi.org/10.1007/978-3-0348-7605-6.

[32] Baranger, J., Maître, J.-F., and Oudin, F. Connection between finite volume and mixed finite element methods. *RAIRO Modél. Math. Anal. Numér.* **30** (1996), 445–465.

[33] Bartels, S., and Kaltenbach, A. Exact a posteriori error control for variational problems *via* convex duality and explicit flux reconstruction. In *Error control, adaptive discretizations, and applications. Part 1*, vol. **58** of *Adv. Appl. Mech.* Academic Press, San Diego, CA, 2024, pp. 295–375.

[34] Bastian, P., and Rivière, B. Superconvergence and $H(\mathrm{div})$ projection for discontinuous Galerkin methods. *Internat. J. Numer. Methods Fluids* **42** (2003), 1043–1057.

[35] Bear, J., and Bachmat, Y. *Introduction to Modeling of Transport Phenomena in Porous Media*, vol. **4** of *Theory and Applications of Transport in Porous Media.* Kluwer Academic Publishers, Dordrecht, Holland, 1990.

[36] Beaude, L., Brenner, K., Lopez, S., Masson, R., and Smai, F. Non-isothermal compositional liquid gas Darcy flow: formulation, soil-atmosphere boundary condition and application to high-energy geothermal simulations. *Comput. Geosci.* **23** (2019), 443–470. https://doi.org/10.1007/s10596-018-9794-9.

[37] Bebendorf, M. A note on the Poincaré inequality for convex domains. *Z. Anal. Anwendungen* **22** (2003), 751–756. http://dx.doi.org/10.4171/ZAA/1170.

[38] Becker, R., Capatina, D., and Luce, R. Stopping criteria based on locally reconstructed fluxes. In *Numerical mathematics and advanced applications—ENUMATH 2013*, vol. **103** of *Lect. Notes Comput. Sci. Eng.* Springer, Cham, 2015, pp. 243–251.

[39] Becker, R., Johnson, C., and Rannacher, R. Adaptive error control for multigrid finite element methods. *Computing* **55** (1995), 271–288. http://dx.doi.org/10.1007/BF02238483.

[40] Beirão da Veiga, L., Brezzi, F., Marini, L. D., and Russo, A. Mixed virtual element methods for general second order elliptic problems on polygonal meshes. *ESAIM Math. Model. Numer. Anal.* **50** (2016), 727–747. http://dx.doi.org/10.1051/m2an/2015067.

[41] Beirão da Veiga, L., Lipnikov, K., and Manzini, G. *The mimetic finite difference method for elliptic problems*, vol. **11** of *MS&A. Modeling, Simulation and Applications.* Springer, Cham, 2014. https://doi.org/10.1007/978-3-319-02663-3.

[42] Beirão da Veiga, L. A residual based error estimator for the mimetic finite difference method. *Numer. Math.* **108** (2008), 387–406. http://dx.doi.org/10.1007/s00211-007-0126-6.

[43] Beirão da Veiga, L., and Manzini, G. An a posteriori error estimator for the mimetic finite difference approximation of elliptic problems. *Internat. J. Numer. Methods Engrg.* **76** (2008), 1696–1723. http://dx.doi.org/10.1002/nme.2377.

[44] Ben Gharbia, I., Dabaghi, J., Martin, V., and Vohralík, M. A posteriori error estimates for a compositional two-phase flow with nonlinear complementarity constraints. *Comput. Geosci.* **24** (2020), 1031–1055. https://doi.org/10.1007/s10596-019-09909-5.

[45] Ben Gharbia, I., Ferzly, J., Vohralík, M., and Yousef, S. Semismooth and smoothing Newton methods for nonlinear systems with complementarity constraints: Adaptivity and inexact resolution. *J. Comput. Appl. Math.* **420** (2023), 114765. https://doi.org/10.1016/j.cam.2022.114765.

[46] Beneš, M. Analysis of non-isothermal multiphase flows in porous media. *Math. Methods Appl. Sci.* **45** (2022), 9653–9677. https://doi.org/10.1002/mma.8328.

[47] Bergam, A., Bernardi, C., and Mghazli, Z. A posteriori analysis of the finite element discretization of some parabolic equations. *Math. Comp.* **74** (2005), 1117–1138.

[48] Bergam, A., Mghazli, Z., and Verfürth, R. Estimations a posteriori d'un schéma de volumes finis pour un problème non linéaire. *Numer. Math.* **95** (2003), 599–624.

[49] Bernardi, C., Chacón Rebollo, T., Chacón Vera, E., and Franco Coronil, D. A posteriori error analysis for two-overlapping domain decomposition techniques. *Appl. Numer. Math.* **59** (2009), 1214–1236. http://dx.doi.org/10.1016/j.apnum.2008.06.004.

[50] Bernardi, C., Dakroub, J., Mansour, G., and Sayah, T. A posteriori analysis of iterative algorithms for a nonlinear problem. *J. Sci. Comput.* **65** (2015), 672–697. https://doi.org/10.1007/s10915-014-9980-4.

[51] Bernardi, C., El Alaoui, L., and Mghazli, Z. A posteriori analysis of a space and time discretization of a nonlinear model for the flow in partially saturated porous media. *IMA J. Numer. Anal.* **34** (2014), 1002–1036. http://dx.doi.org/10.1093/imanum/drt014.

[52] Bochev, P. B., and Hyman, J. M. Principles of mimetic discretizations of differential operators. In *Compatible spatial discretizations*, vol. **142** of *IMA Vol. Math. Appl.* Springer, New York, 2006, pp. 89–119. http://dx.doi.org/10.1007/0-387-38034-5_5.

[53] Boffi, D., Brezzi, F., and Fortin, M. *Mixed finite element methods and applications*, vol. **44** of *Springer Series in Computational Mathematics*. Springer, Heidelberg, 2013. https://doi.org/10.1007/978-3-642-36519-5.

[54] Boffi, D., and Di Pietro, D. A. Unified formulation and analysis of mixed and primal discontinuous skeletal methods on polytopal meshes. *ESAIM Math. Model. Numer. Anal.* **52** (2018), 1–28. https://doi.org/10.1051/m2an/2017036.

[55] Bonelle, J., and Ern, A. Analysis of compatible discrete operator schemes for elliptic problems on polyhedral meshes. *ESAIM Math. Model. Numer. Anal.* **48** (2014), 553–581. http://dx.doi.org/10.1051/m2an/2013104.

[56] Bossavit, A. "Hypercercle" et majorations d'erreur a posteriori par "corrections locales". In *Équations aux dérivées partielles et applications*. Gauthier-Villars, Éd. Sci. Méd. Elsevier, Paris, 1998, pp. 221–237.

[57] Braess, D., and Schöberl, J. Equilibrated residual error estimator for edge elements. *Math. Comp.* **77** (2008), 651–672. http://dx.doi.org/10.1090/S0025-5718-07-02080-7.

[58] Breil, J., and Maire, P.-H. A cell-centered diffusion scheme on two-dimensional unstructured meshes. *J. Comput. Phys.* **224** (2007), 785–823. http://dx.doi.org/10.1016/j.jcp.2006.10.025.

[59] Brenner, K., Chorfi, N., and Masson, R. Sequential implicit vertex approximate gradient discretization of incompressible two-phase Darcy flows with discontinuous capillary pressure. *Comput. Geosci.* **26** (2022), 147–169. https://doi.org/10.1007/s10596-021-10113-7.

[60] Brenner, K., Masson, R., Quenjel, E. H., and Droniou, J. Total velocity-based finite volume discretization of two-phase Darcy flow in highly heterogeneous media with discontinuous capillary pressure. *IMA J. Numer. Anal.* **42** (2022), 1231–1272. https://doi.org/10.1093/imanum/drab018.

[61] Brenner, S. C., and Scott, L. R. *The mathematical theory of finite element methods*, third ed., vol. **15** of *Texts in Applied Mathematics*. Springer, New York, 2008. http://dx.doi.org/10.1007/978-0-387-75934-0.

[62] Brezzi, F., Falk, R. S., and Marini, L. D. Basic principles of mixed virtual element methods. *ESAIM Math. Model. Numer. Anal.* **48** (2014), 1227–1240. https://doi.org/10.1051/m2an/2013138.

[63] Brezzi, F., and Fortin, M. *Mixed and hybrid finite element methods*, vol. **15** of *Springer Series in Computational Mathematics*. Springer-Verlag, New York, 1991. http://dx.doi.org/10.1007/978-1-4612-3172-1.

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

[65] Brezzi, F., Lipnikov, K., and Simoncini, V. A family of mimetic finite difference methods on polygonal and polyhedral meshes. *Math. Models Methods Appl. Sci.* **15** (2005), 1533–1551. `http://dx.doi.org/10.1142/S0218202505000832`.

[66] Bringmann, P., Feischl, M., Miraçi, A., Praetorius, D., and Streitberger, J. On full linear convergence and optimal complexity of adaptive FEM with inexact solver. *Comput. Math. Appl.* **180** (2025), 102–129. `https://doi.org/10.1016/j.camwa.2024.12.013`.

[67] Buhr, A., Engwer, C., Ohlberger, M., and Rave, S. ArbiLoMod, a simulation technique designed for arbitrary local modifications. *SIAM J. Sci. Comput.* **39** (2017), A1435–A1465. `https://doi.org/10.1137/15M1054213`.

[68] Burman, E., and Ern, A. Continuous interior penalty *hp*-finite element methods for advection and advection-diffusion equations. *Math. Comp.* **76** (2007), 1119–1140.

[69] Cancès, C., Gallouët, T., and Porretta, A. Two-phase flows involving capillary barriers in heterogeneous porous media. *Interfaces Free Bound.* **11** (2009), 239–258. `http://dx.doi.org/10.4171/IFB/210`.

[70] Cancès, C., Pop, I. S., and Vohralík, M. An a posteriori error estimate for vertex-centered finite volume discretizations of immiscible incompressible two-phase flow. *Math. Comp.* **83** (2014), 153–188. `http://dx.doi.org/10.1090/S0025-5718-2013-02723-8`.

[71] Cao, X., and Pop, I. S. Two-phase porous media flows with dynamic capillary effects and hysteresis: uniqueness of weak solutions. *Comput. Math. Appl.* **69** (2015), 688–695. `https://doi.org/10.1016/j.camwa.2015.02.009`.

[72] Caro, F., Saad, B., and Saad, M. Study of degenerate parabolic system modeling the hydrogen displacement in a nuclear waste repository. *Discrete Contin. Dyn. Syst. Ser. S* **7** (2014), 191–205. `https://doi.org/10.3934/dcdss.2014.7.191`.

[73] Carstensen, C., Lazarov, R., and Tomov, S. Explicit and averaging a posteriori error estimates for adaptive finite volume methods. *SIAM J. Numer. Anal.* **42** (2005), 2496–2521.

[74] Chaillou, A., and Suri, M. A posteriori estimation of the linearization error for strongly monotone nonlinear operators. *J. Comput. Appl. Math.* **205** (2007), 72–87. `http://dx.doi.org/10.1016/j.cam.2006.04.041`.

[75] Chaillou, A. L., and Suri, M. Computable error estimators for the approximation of nonlinear problems by linearized models. *Comput. Methods Appl. Mech. Engrg.* **196** (2006), 210–224. `http://dx.doi.org/10.1016/j.cma.2006.03.008`.

[76] Chamoin, L., and Legoll, F. An introductory review on a posteriori error estimation in finite element computations. *SIAM Rev.* **65** (2023), 963–1028. `https://doi.org/10.1137/21M1464841`.

[77] Chaudhry, J. H., Estep, D., and Tavener, S. J. A posteriori error analysis for Schwarz overlapping domain decomposition methods. *BIT* **61** (2021), 1153–1191. `https://doi.org/10.1007/s10543-021-00864-1`.

[78] Chavent, G., and Jaffré, J. Studies in Mathematics and Its Applications, Vol. 17. *Mathematical models and finite elements for reservoir simulation.* North-Holland, Amsterdam, 1986.

[79] 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** (2003), 655–682.

[80] Chen, Y., and Liu, W. A posteriori error estimates of mixed methods for miscible displacement problems. *Internat. J. Numer. Methods Engrg.* **73** (2008), 331–343. `http://dx.doi.org/10.1002/nme.2075`.

[81] Chen, Z. Degenerate two-phase incompressible flow. I. Existence, uniqueness and regularity of a weak solution. *J. Differential Equations* **171** (2001), 203–232. http://dx.doi.org/10.1006/jdeq.2000.3848.

[82] Chen, Z. Degenerate two-phase incompressible flow. II. Regularity, stability and stabilization. *J. Differential Equations* **186** (2002), 345–376. http://dx.doi.org/10.1016/S0022-0396(02)00027-X.

[83] Chen, Z. Mathematical techniques in oil recovery. *Reservoir simulation*, vol. **77** of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2007. https://doi.org/10.1137/1.9780898717075.

[84] Chen, Z., and Ewing, R. E. Degenerate two-phase incompressible flow. III. Sharp error estimates. *Numer. Math.* **90** (2001), 215–240. http://dx.doi.org/10.1007/s002110100291.

[85] Chen, Z., and Ewing, R. E. Degenerate two-phase incompressible flow. IV. Local refinement and domain decomposition. *J. Sci. Comput.* **18** (2003), 329–360. http://dx.doi.org/10.1023/A:1022673427893.

[86] Chen, Z., Huan, G., and Ma, Y. *Computational methods for multiphase flows in porous media*. Computational Science & Engineering. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2006. http://dx.doi.org/10.1137/1.9780898718942.

[87] Christiansen, S. H. A construction of spaces of compatible differential forms on cellular complexes. *Math. Models Methods Appl. Sci.* **18** (2008), 739–757. http://dx.doi.org/10.1142/S021820250800284X.

[88] Christiansen, S. H. Éléments finis mixtes minimaux sur les polyèdres. *C. R. Math. Acad. Sci. Paris* **348** (2010), 217–221. http://dx.doi.org/10.1016/j.crma.2010.01.017.

[89] Christie, M. A., and Blunt, M. J. Tenth SPE comparative solution project: a comparison of upscaling techniques. *SPE Reservoir Eval. Eng.* **4** (2001), 308–317. https://doi.org/10.2118/72469-PA.

[90] Chueh, C.-C., Djilali, N., and Bangerth, W. An *h*-adaptive operator splitting method for two-phase flow in 3D heterogeneous porous media. *SIAM J. Sci. Comput.* **35** (2013), B149–B175. https://doi.org/10.1137/120866208.

[91] Chueh, C. C., Secanell, M., Bangerth, W., and Djilali, N. Multi-level adaptive simulation of transient two-phase flow in heterogeneous porous media. *Comput. & Fluids* **39** (2010), 1585–1596. https://doi.org/10.1016/j.compfluid.2010.05.011.

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

[93] Coats, K. H. An equation of state compositional model. *Society of Petroleum Engineers Journal* **20** (1980), 363–376. https://doi.org/10.2118/8284-PA.

[94] Coats, K. H. Implicit compositional simulation of single-porosity and dual-porosity reservoirs. In *SPE Symposium on Reservoir Simulation* (1989). Paper SPE-18427-MS, https://doi.org/10.2118/18427-MS.

[95] Cockburn, B., Di Pietro, D. A., and Ern, A. Bridging the hybrid high-order and hybridizable discontinuous Galerkin methods. *ESAIM Math. Model. Numer. Anal.* **50** (2016), 635–650. http://dx.doi.org/10.1051/m2an/2015051.

[96] Cottereau, R., Díez, P., and Huerta, A. Strict error bounds for linear solid mechanics problems using a subdomain-based flux-free method. *Comput. Mech.* **44** (2009), 533–547. https://doi.org/10.1007/s00466-009-0388-1.

[97] Destuynder, P., and Métivet, B. Explicit error bounds in a conforming finite element method. *Math. Comp.* **68** (1999), 1379–1396. http://dx.doi.org/10.1090/S0025-5718-99-01093-5.

[98] Deucher, R. H., Franc, J., Mø yner, O., and Tchelepi, H. A. Compositional reservoir simulation with a high-resolution compact stencil adaptive implicit method. *J. Comput. Phys.* **521** (2025), Paper No. 113558, 17. https://doi.org/10.1016/j.jcp.2024.113558.

[99] Deuflhard, P., and Weiser, M. *Adaptive numerical solution of PDEs.* de Gruyter Textbook. Walter de Gruyter & Co., Berlin, 2012. http://dx.doi.org/10.1515/9783110283112.

[100] Di Pietro, D. A., and Ern, A. *Mathematical aspects of discontinuous Galerkin methods*, vol. **69** of *Mathématiques & Applications (Berlin) [Mathematics & Applications]*. Springer, Heidelberg, 2012. http://dx.doi.org/10.1007/978-3-642-22980-0.

[101] Di Pietro, D. A., and Ern, A. A hybrid high-order locking-free method for linear elasticity on general meshes. *Comput. Methods Appl. Mech. Engrg.* **283** (2015), 1–21. http://dx.doi.org/10.1016/j.cma.2014.09.009.

[102] Di Pietro, D. A., Ern, A., and Lemaire, S. An arbitrary-order and compact-stencil discretization of diffusion on general meshes based on local reconstruction operators. *Comput. Methods Appl. Math.* **14** (2014), 461–472. https://doi.org/10.1515/cmam-2014-0018.

[103] Di Pietro, D. A., Flauraud, E., Vohralík, M., and Yousef, S. A posteriori error estimates, stopping criteria, and adaptivity for multiphase compositional Darcy flows in porous media. *J. Comput. Phys.* **276** (2014), 163–187. http://dx.doi.org/10.1016/j.jcp.2014.06.061.

[104] Di Pietro, D. A., Vohralík, M., and Yousef, S. An a posteriori-based, fully adaptive algorithm with adaptive stopping criteria and mesh refinement for thermal multiphase compositional flows in porous media. *Comput. Math. Appl.* **68** (2014), 2331–2347. https://doi.org/10.1016/j.camwa.2014.08.008.

[105] Ding, Y., and Lemonnier, P. Use of corner point geometry in reservoir simulation. In *International Meeting on Petroleum Engineering* (1995). Paper SPE-29933-MS, https://doi.org/10.2118/29933-MS.

[106] Dolgov, S., and Vejchodský, T. Guaranteed *a posteriori* error bounds for low-rank tensor approximate solutions. *IMA J. Numer. Anal.* **41** (2021), 1240–1266. https://doi.org/10.1093/imanum/draa010.

[107] Droniou, J. Finite volume schemes for diffusion equations: introduction to and review of modern methods. *Math. Models Methods Appl. Sci.* **24** (2014), 1575–1619. http://dx.doi.org/10.1142/S0218202514400041.

[108] Droniou, J., and Eymard, R. A mixed finite volume scheme for anisotropic diffusion problems on any grid. *Numer. Math.* **105** (2006), 35–71.

[109] Droniou, J., Eymard, R., Gallouët, T., Guichard, C., and Herbin, R. *The gradient discretisation method*, vol. **82** of *Mathématiques & Applications (Berlin) [Mathematics & Applications]*. Springer, Cham, 2018. https://doi.org/10.1007/978-3-319-79042-8.

[110] Droniou, J., Eymard, R., Gallouët, T., and Herbin, R. A unified approach to mimetic finite difference, hybrid finite volume and mixed finite volume methods. *Math. Models Methods Appl. Sci.* **20** (2010), 265–295. http://dx.doi.org/10.1142/S0218202510004222.

[111] Droniou, J., Eymard, R., and Herbin, R. Gradient schemes: generic tools for the numerical analysis of diffusion equations. *ESAIM Math. Model. Numer. Anal.* **50** (2016), 749–781. http://dx.doi.org/10.1051/m2an/2015079.

[112] Durán, R., and Padra, C. An error estimator for nonconforming approximations of a nonlinear problem. In *Finite element methods (Jyväskylä, 1993)*, vol. **164** of *Lecture Notes in Pure and Appl. Math.* Dekker, New York, 1994, pp. 201–205.

[113] Edwards, M. G. Unstructured, control-volume distributed, full-tensor finite-volume schemes with flow based grids. *Comput. Geosci.* **6** (2002), 433–452. Locally conservative numerical methods for flow in porous media, http://dx.doi.org/10.1023/A:1021243231313.

[114] El Alaoui, L., Ern, A., and Vohralík, M. Guaranteed and robust a posteriori error estimates and balancing discretization and linearization errors for monotone nonlinear problems. *Comput. Methods Appl. Mech. Engrg.* **200** (2011), 2782–2795. http://dx.doi.org/10.1016/j.cma.2010.03.024.

[115] El Hassouni, S., and Mghazli, Z. Mesh adaptation for two-phase flow. *Int. J. Math. Stat.* **13** (2013), 56–68. https://doi.org/10.1002/pamm.201310023.

[116] Erath, C., and Praetorius, D. A posteriori error estimate and adaptive mesh refinement for the cell-centered finite volume method for elliptic boundary value problems. *SIAM J. Numer. Anal.* **47** (2008/09), 109–135. http://dx.doi.org/10.1137/070702126.

[117] Eriksson, K., and Johnson, C. Adaptive streamline diffusion finite element methods for stationary convection-diffusion problems. *Math. Comp.* **60** (1993), 167–188, S1–S2.

[118] Ern, A., and Guermond, J.-L. *Finite Elements I. Approximation and Interpolation*, vol. **72** of *Texts in Applied Mathematics*. Springer International Publishing, Springer Nature Switzerland AG, 2021. https://doi-org/10.1007/978-3-030-56341-7.

[119] Ern, A., Nicaise, S., and Vohralík, M. An accurate **H**(div) flux reconstruction for discontinuous Galerkin approximations of elliptic problems. *C. R. Math. Acad. Sci. Paris* **345** (2007), 709–712. http://dx.doi.org/10.1016/j.crma.2007.10.036.

[120] Ern, A., Smears, I., and Vohralík, M. Guaranteed, locally space-time efficient, and polynomial-degree robust a posteriori error estimates for high-order discretizations of parabolic problems. *SIAM J. Numer. Anal.* **55** (2017), 2811–2834. https://doi.org/10.1137/16M1097626.

[121] Ern, A., Smears, I., and Vohralík, M. Equilibrated flux a posteriori error estimates in $L^2(H^1)$-norms for high-order discretizations of parabolic problems. *IMA J. Numer. Anal.* **39** (2019), 1158–1179. https://doi.org/10.1093/imanum/dry035.

[122] Ern, A., Stephansen, A. F., and Vohralík, M. Guaranteed and robust discontinuous Galerkin a posteriori error estimates for convection-diffusion-reaction problems. *J. Comput. Appl. Math.* **234** (2010), 114–130. http://dx.doi.org/10.1016/j.cam.2009.12.009.

[123] Ern, A., and Vohralík, M. A posteriori error estimation based on potential and flux reconstruction for the heat equation. *SIAM J. Numer. Anal.* **48** (2010), 198–223. http://dx.doi.org/10.1137/090759008.

[124] Ern, A., and Vohralík, M. Adaptive inexact Newton methods with a posteriori stopping criteria for nonlinear diffusion PDEs. *SIAM J. Sci. Comput.* **35** (2013), A1761–A1791. http://dx.doi.org/10.1137/120896918.

[125] Ern, A., and Vohralík, M. Polynomial-degree-robust a posteriori estimates in a unified setting for conforming, nonconforming, discontinuous Galerkin, and mixed discretizations. *SIAM J. Numer. Anal.* **53** (2015), 1058–1081. http://dx.doi.org/10.1137/130950100.

[126] Eymard, R., Gallouët, T., and Herbin, R. Finite volume methods. In *Handbook of Numerical Analysis, Vol. VII*. North-Holland, Amsterdam, 2000, pp. 713–1020.

[127] Eymard, R., Gallouët, T., and Herbin, R. Finite volume approximation of elliptic problems and convergence of an approximate gradient. *Appl. Numer. Math.* **37** (2001), 31–53. http://dx.doi.org/10.1016/S0168-9274(00)00024-6.

[128] Eymard, R., Gallouët, T., and Herbin, R. Discretization of heterogeneous and anisotropic diffusion problems on general nonconforming meshes SUSHI: a scheme using stabilization and hybrid interfaces. *IMA J. Numer. Anal.* **30** (2010), 1009–1043. http://dx.doi.org/10.1093/imanum/drn084.

[129] Eymard, R., Guichard, C., Herbin, R., and Masson, R. Vertex-centred discretization of multiphase compositional Darcy flows on general meshes. *Comput. Geosci.* **16** (2012), 987–1005.

[130] Faigle, B., Elfeel, M. A., Helmig, R., Becker, B., Flemisch, B., and Geiger, S. Multi-physics modeling of non-isothermal compositional flow on adaptive grids. *Comput. Methods Appl. Mech. Engrg.* **292** (2015), 16–34. https://doi.org/10.1016/j.cma.2014.11.030.

[131] Faigle, B., Helmig, R., Aavatsmark, I., and Flemisch, B. Efficient multiphysics modelling with adaptive grid refinement using a MPFA method. *Comput. Geosci.* **18** (2014), 625–636. https://doi.org/10.1007/s10596-014-9407-1.

[132] Faille, I. A control volume method to solve an elliptic equation on a two-dimensional irregular mesh. *Comput. Methods Appl. Mech. Engrg.* **100** (1992), 275–290.

[133] Feireisl, E., Hilhorst, D., Petzeltová, H., and Takáč, P. Mathematical analysis of variable density flows in porous media. *J. Evol. Equ.* **16** (2016), 1–19. https://doi.org/10.1007/s00028-015-0290-6.

[134] Févotte, F., Rappaport, A., and Vohralík, M. Adaptive regularization, discretization, and linearization for nonsmooth problems based on primal-dual gap estimators. *Comput. Methods Appl. Mech. Engrg.* **418** (2024), 116558. https://doi.org/10.1016/j.cma.2023.116558.

[135] Ganis, B., Pencheva, G., and Wheeler, M. F. Adaptive mesh refinement with an enhanced velocity mixed finite element method on semi-structured grids using a fully coupled solver. *Comput. Geosci.* **23** (2019), 149–168. https://doi.org/10.1007/s10596-018-9789-6.

[136] Gantner, G., Haberl, A., Praetorius, D., and Schimanko, S. Rate optimality of adaptive finite element methods with respect to overall computational costs. *Math. Comp.* **90** (2021), 2011–2040. https://doi.org/10.1090/mcom/3654.

[137] Georgoulis, E. H., Lakkis, O., and Virtanen, J. M. A posteriori error control for discontinuous Galerkin methods for parabolic problems. *SIAM J. Numer. Anal.* **49** (2011), 427–458. http://dx.doi.org/10.1137/080722461.

[138] Georgoulis, E. H., Lakkis, O., and Wihler, T. P. A posteriori error bounds for fully-discrete hp-discontinuous Galerkin timestepping methods for parabolic problems. *Numer. Math.* **148** (2021), 363–386. https://doi.org/10.1007/s00211-021-01187-7.

[139] Georgoulis, E. H., and Makridakis, C. G. Lower bounds, elliptic reconstruction and *a posteriori* error control of parabolic problems. *IMA J. Numer. Anal.* **43** (2023), 3212–3242. https://doi.org/10.1093/imanum/drac080.

[140] Golub, G. H., and Strakoš, Z. Estimates in quadratic formulas. *Numer. Algorithms* **8** (1994), 241–268. http://dx.doi.org/10.1007/BF02142693.

[141] Grospellier, G., and Lelandais, B. The Arcane development framework. In *Proceedings of the 8th workshop on Parallel/High-Performance Object-Oriented Scientific Computing* (New York, NY, USA, 2009), POOSC '09, ACM, pp. 4:1–4:11. http://doi.acm.org/10.1145/1595655.1595659.

[142] Haberl, A., Praetorius, D., Schimanko, S., and Vohralík, M. Convergence and quasi-optimal cost of adaptive algorithms for nonlinear operators including iterative linearization and algebraic solver. *Numer. Math.* **147** (2021), 679–725. https://doi.org/10.1007/s00211-021-01176-w.

[143] Han, W. A posteriori error analysis for linearization of nonlinear elliptic problems and their discretizations. *Math. Methods Appl. Sci.* **17** (1994), 487–508. http://dx.doi.org/10.1002/mma.1670170702.

[144] Han, W. With applications in modeling and numerical approximations. *A posteriori error analysis via duality theory*, vol. **8** of *Advances in Mechanics and Mathematics*. Springer-Verlag, New York, 2005.

[145] Harnist, A., Mitra, K., Rappaport, A., and Vohralík, M. Robust augmented energy a posteriori estimates for Lipschitz and strongly monotone elliptic problems. HAL Preprint 04033438, https://hal.inria.fr/hal-04033438, 2024.

[146] Heid, P., Praetorius, D., and Wihler, T. P. Energy contraction and optimal convergence of adaptive iterative linearized finite element methods. *Comput. Methods Appl. Math.* **21** (2021), 407–422. https://doi.org/10.1515/cmam-2021-0025.

[147] Heid, P., and Wihler, T. P. Adaptive iterative linearization Galerkin methods for nonlinear problems. *Math. Comp.* **89** (2020), 2707–2734. https://doi.org/10.1090/mcom/3545.

[148] Henning, P., Ohlberger, M., and Schweizer, B. Adaptive heterogeneous multiscale methods for immiscible two-phase flow in porous media. *Comput. Geosci.* **19** (2015), 99–114. https://doi.org/10.1007/s10596-014-9455-6.

[149] Herbin, R. An error estimate for a finite volume scheme for a diffusion-convection problem on a triangular mesh. *Numer. Methods Partial Differential Equations* **11** (1995), 165–173. https://doi.org/10.1002/num.1690110205.

[150] Herbin, R., and Hubert, F. Benchmark on discretization schemes for anisotropic diffusion problems on general grids. In *Finite volumes for complex applications V*. ISTE, London, 2008, pp. 659–692.

[151] Hilhorst, D., and Vohralík, M. A posteriori error estimates for combined finite volume–finite element discretizations of reactive transport equations on nonmatching grids. *Comput. Methods Appl. Mech. Engrg.* **200** (2011), 597–613. http://dx.doi.org/10.1016/j.cma.2010.08.017.

[152] Hlaváček, I., Haslinger, J., Nečas, J., and Lovíšek, J. Translated from the Slovak by J. Jarník. *Solution of variational inequalities in mechanics*, vol. **66** of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1988.

[153] Jenny, P., Lee, S. H., and Tchelep, H. A. Adaptive multiscale finite-volume method for multiphase flow and transport in porous media. *Multiscale Model. Simul.* **3** (2004/05), 50–64. http://dx.doi.org/10.1137/030600795.

[154] Jiránek, P., Strakoš, Z., and Vohralík, M. A posteriori error estimates including algebraic error and stopping criteria for iterative solvers. *SIAM J. Sci. Comput.* **32** (2010), 1567–1590. http://dx.doi.org/10.1137/08073706X.

[155] Jovanović, B. S., and Süli, E. For linear partial differential equations with generalized solutions. *Analysis of finite difference schemes*, vol. **46** of *Springer Series in Computational Mathematics*. Springer, London, 2014. https://doi.org/10.1007/978-1-4471-5460-0.

[156] Kaltenbacher, B. A posteriori parameter choice strategies for some Newton type methods for the regularization of nonlinear ill-posed problems. *Numer. Math.* **79** (1998), 501–528. http://dx.doi.org/10.1007/s002110050349.

[157] Karakashian, O. A., and Pascal, F. A posteriori error estimates for a discontinuous Galerkin approximation of second-order elliptic problems. *SIAM J. Numer. Anal.* **41** (2003), 2374–2399. http://dx.doi.org/10.1137/S0036142902405217.

[158] Keilegavlen, E., and Nordbotten, J. M. Conservative inexact solvers for porous media flow. In *Domain decomposition methods in science and engineering XXI*, vol. **98** of *Lect. Notes Comput. Sci. Eng.* Springer, Cham, 2014, pp. 333–340.

[159] Keilegavlen, E., and Nordbotten, J. M. Inexact linear solvers for control volume discretizations in porous media. *Comput. Geosci.* **19** (2015), 159–176. https://doi.org/10.1007/s10596-014-9453-8.

[160] Kelley, C. T. With separately available software. *Iterative methods for linear and nonlinear equations*, vol. **16** of *Frontiers in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1995. https://doi.org/10.1137/1.9781611970944.

[161] Khalil, Z., and Saad, M. Degenerate two-phase compressible immiscible flow in porous media: the case where the density of each phase depends on its own pressure. *Math. Comput. Simulation* **81** (2011), 2225–2233. http://dx.doi.org/10.1016/j.matcom.2010.12.012.

[162] Kim, K.-Y. A posteriori error analysis for locally conservative mixed methods. *Math. Comp.* **76** (2007), 43–66. http://dx.doi.org/10.1090/S0025-5718-06-01903-X.

[163] Kim, K.-Y. A posteriori error estimators for locally conservative methods of nonlinear elliptic problems. *Appl. Numer. Math.* **57** (2007), 1065–1080. http://dx.doi.org/10.1016/j.apnum.2006.09.010.

[164] Klausen, R. A., and Russell, T. F. Relationships among some locally conservative discretization methods which handle discontinuous coefficients. *Comput. Geosci.* **8** (2004), 341–377 (2005). http://dx.doi.org/10.1007/s10596-005-1815-9.

[165] Klieber, W., and Rivière, B. Adaptive simulations of two-phase flow by discontinuous Galerkin methods. *Comput. Methods Appl. Mech. Engrg.* **196** (2006), 404–419. http://dx.doi.org/10.1016/j.cma.2006.05.007.

[166] Knabner, P., and Otto, F. Solute transport in porous media with equilibrium and nonequilibrium multiple-site adsorption: uniqueness of weak solutions. *Nonlinear Anal.* **42** (2000), 381–403.

[167] Kou, J., Sun, S., and Wang, X. Linearly decoupled energy-stable numerical methods for multi-component two-phase compressible flow. *SIAM J. Numer. Anal.* **56** (2018), 3219–3248. https://doi.org/10.1137/17M1162287.

[168] Kröner, D., Küther, M., Ohlberger, M., and Rohde, C. A posteriori error estimates and adaptive methods for hyperbolic and convection dominated parabolic conservation laws. In *Trends in nonlinear analysis*. Springer, Berlin, 2003, pp. 289–306.

[169] Kröner, D., and Luckhaus, S. Flow of oil and water in a porous medium. *J. Differential Equations* **55** (1984), 276–288. http://dx.doi.org/10.1016/0022-0396(84)90084-6.

[170] Kröner, D., and Ohlberger, M. A posteriori error estimates for upwind finite volume schemes for nonlinear conservation laws in multidimensions. *Math. Comp.* **69** (2000), 25–39. http://dx.doi.org/10.1090/S0025-5718-99-01158-8.

[171] Kumar, K., Kyas, S., Nordbotten, J. M., and Repin, S. Guaranteed and computable error bounds for approximations constructed by an iterative decoupling of the Biot problem. *Comput. Math. Appl.* **91** (2021), 122–149. https://doi.org/10.1016/j.camwa.2020.05.005.

[172] Kuznetsov, Y., and Repin, S. New mixed finite element method on polygonal and polyhedral meshes. *Russian J. Numer. Anal. Math. Modelling* **18** (2003), 261–278. http://dx.doi.org/10.1163/156939803322380846.

[173] Kuznetsov, Y. A. Mixed finite element methods on polyhedral meshes for diffusion equations. In *Partial differential equations*, vol. **16** of *Comput. Methods Appl. Sci.* Springer, Dordrecht, 2008, pp. 27–41.

[174] Ladevèze, P. *Comparaison de modèles de milieux continus*. Ph.D. thesis, Université Pierre et Marie Curie (Paris 6), 1975.

[175] Ladevèze, P., and Leguillon, D. Error estimate procedure in the finite element method and applications. *SIAM J. Numer. Anal.* **20** (1983), 485–509.

[176] Ladevèze, P., and Pelle, J.-P. Translated from the 2001 French original by Theofanis Strouboulis. *Mastering calculations in linear and nonlinear mechanics*. Mechanical Engineering Series. Springer-Verlag, New York, 2005.

[177] Larson, M. G., and Niklasson, A. J. A conservative flux for the continuous Galerkin method based on discontinuous enrichment. *Calcolo* **41** (2004), 65–76. https://doi.org/10.1007/s10092-004-0084-7.

[178] Lazarov, R., and Tomov, S. A posteriori error estimates for finite volume element approximations of convection-diffusion-reaction equations. *Comput. Geosci.* **6** (2002), 483–503. Locally conservative numerical methods for flow in porous media, https://doi.org/10.1023/A:1021247300362.

[179] Li, H., Leung, W. T., and Wheeler, M. F. Sequential local mesh refinement solver with separate temporal and spatial adaptivity for non-linear two-phase flow problems. *J. Comput. Phys.* **403** (2020), 109074, 23. https://doi.org/10.1016/j.jcp.2019.109074.

[180] Li, H., Mu, L., and Ye, X. A posteriori error estimates for the weak Galerkin finite element methods on polytopal meshes. *Commun. Comput. Phys.* **26** (2019), 558–578. https://doi.org/10.4208/cicp.oa-2018-0058.

[181] Li, H., and Wheeler, M. F. Dynamic local coupling for multiphase flow: a compromise between efficiency and stability. *J. Comput. Phys.* **469** (2022), Paper No. 111535, 14. https://doi.org/10.1016/j.jcp.2022.111535.

[182] Liu, L., and Keyes, D. E. Approximate error bounds on solutions of nonlinearly preconditioned PDEs. *SIAM J. Sci. Comput.* **43** (2021), A2526–A2554. https://doi.org/10.1137/19M1285044.

[183] Luce, R., and Wohlmuth, B. I. A local a posteriori error estimator based on equilibrated fluxes. *SIAM J. Numer. Anal.* **42** (2004), 1394–1414. http://dx.doi.org/10.1137/S0036142903433790.

[184] Manteuffel, T. A., and White, Jr., A. B. The numerical solution of second-order boundary value problems on nonuniform meshes. *Math. Comp.* **47** (1986), 511–535, S53–S55. https://doi.org/10.2307/2008170.

[185] Mesri, Y., and Ricois, O. Construction process for an improved meshing for the simulation of a reservoir in an underground formation. Patent CA 28886110, http://brevets-patents.ic.gc.ca/opic-cipo/cpd/fra/brevet/2886110/, 23 03, 2015.

[186] Mghazli, Z. A posteriori error analysis for finite element approximation of some groundwater models. Part I: Linear case. In *Error control, adaptive discretizations, and applications. Part 3*, vol. **60** of *Adv. Appl. Mech.* Academic Press, San Diego, CA, 2025. In press.

[187] Mghazli, Z. A posteriori error analysis for finite element approximation of some groundwater models. Part II: Richards equation and mixed finite element. In *Error control, adaptive discretizations, and applications. Part 4*, vol. **61** of *Adv. Appl. Mech.* Academic Press, San Diego, CA, 2025. In press.

[188] Mitchell, W. F. A collection of 2D elliptic problems for testing adaptive grid refinement algorithms. *Appl. Math. Comput.* **220** (2013), 350–364. https://doi.org/10.1016/j.amc.2013.05.068.

[189] Mitra, K., and Vohralík, M. Guaranteed, locally efficient, and robust a posteriori estimates for nonlinear elliptic problems in iteration-dependent norms. An orthogonal decomposition result based on iterative linearization. HAL Preprint 04156711, submitted for publication, https://hal.inria.fr/hal-04156711, 2023.

[190] Mitra, K., and Vohralík, M. A posteriori error estimates for the Richards equation. *Math. Comp.* **93** (2024), 1053–1096. https://doi.org/10.1090/mcom/3932.

[191] Morton, K. W., and Süli, E. Finite volume methods and their analysis. *IMA J. Numer. Anal.* **11** (1991), 241–260. https://doi.org/10.1093/imanum/11.2.241.

[192] Mozolevski, I., and Prudhomme, S. Goal-oriented error estimation based on equilibrated-flux reconstruction for finite element approximations of elliptic problems. *Comput. Methods Appl. Mech. Engrg.* **288** (2015), 127–145. https://doi.org/10.1016/j.cma.2014.09.025.

[193] Munar, M., Cangiani, A., and Velásquez, I. Residual-based a posteriori error estimation for mixed virtual element methods. *Comput. Math. Appl.* **166** (2024), 182–197. https://doi.org/10.1016/j.camwa.2024.05.011.

[194] Murphy, T. J., and Walkington, N. J. Control volume approximation of degenerate two-phase porous flows. *SIAM J. Numer. Anal.* **57** (2019), 527–546.

[195] Nédélec, J.-C. Mixed finite elements in $\mathbb{R}^3$. *Numer. Math.* **35** (1980), 315–341.

[196] Nicaise, S. A posteriori error estimations of some cell-centered finite volume methods. *SIAM J. Numer. Anal.* **43** (2005), 1481–1503.

[197] Nicaise, S. A posteriori error estimations of some cell centered finite volume methods for diffusion-convection-reaction problems. *SIAM J. Numer. Anal.* **44** (2006), 949–978.

[198] Nochetto, R. H., Siebert, K. G., and Veeser, A. Theory of adaptive finite element methods: an introduction. In *Multiscale, nonlinear and adaptive approximation*. Springer, Berlin, 2009, pp. 409–542. http://dx.doi.org/10.1007/978-3-642-03413-8_12.

[199] Nordbotten, J. M., and Keilegavlen, E. An introduction to multi-point flux (MPFA) and stress (MPSA) finite volume methods for thermo-poroelasticity. In *Polyhedral methods in geosciences*, vol. **27** of *SEMA SIMAI Springer Ser.* Springer, Cham, [2021] ©2021, pp. 119–158. https://doi.org/10.1007/978-3-030-69363-3_4.

[200] Nordbotten, J. M., Keilegavlen, E., and Sandvin, A. Mass conservative domain decomposition for porous media flow. In *Finite Volume Method - Powerful Means of Engineering Design*, R. Petrova, Ed. IntechOpen, Rijeka, 2012, ch. 11. https://doi.org/10.5772/38700.

[201] Odeh, A. S. Comparison of solutions to a three-dimensional black-oil reservoir simulation problem (includes associated paper 9741). *J. Pet. Technol.* **33** (1981), 13–25. https://doi.org/10.2118/9723-PA.

[202] Ohlberger, M. A posteriori error estimate for finite volume approximations to singularly perturbed nonlinear convection–diffusion equations. *Numer. Math.* **87** (2001), 737–761.

[203] Ohlberger, M. A posteriori error estimates for vertex centered finite volume approximations of convection–diffusion–reaction equations. *M2AN Math. Model. Numer. Anal.* **35** (2001), 355–387.

[204] Ohlberger, M., and Rohde, C. Adaptive finite volume approximations for weakly coupled convection dominated parabolic systems. *IMA J. Numer. Anal.* **22** (2002), 253–280.

[205] Olshanskii, M. A., and Tyrtyshnikov, E. E. Theory and applications. *Iterative methods for linear systems*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2014. http://dx.doi.org/10.1137/1.9781611973464.

[206] Omnes, P., Penel, Y., and Rosenbaum, Y. A posteriori error estimation for the discrete duality finite volume discretization of the Laplace equation. *SIAM J. Numer. Anal.* **47** (2009), 2782–2807.

[207] Otto, F. L^1-contraction and uniqueness for quasilinear elliptic-parabolic equations. *J. Differential Equations* **131** (1996), 20–38. http://dx.doi.org/10.1006/jdeq.1996.0155.

[208] Otto, F. L^1-contraction and uniqueness for unstationary saturated-unsaturated porous media flow. *Adv. Math. Sci. Appl.* **7** (1997), 537–553.

[209] Papež, J. Algebraic error in numerical PDEs and its estimation. In *Error control, adaptive discretizations, and applications. Part 1*, vol. **58** of *Adv. Appl. Mech.* Academic Press, San Diego, CA, 2024, pp. 377–427.

[210] Papež, J., Rüde, U., Vohralík, M., and Wohlmuth, B. Sharp algebraic and total a posteriori error bounds for h and p finite elements via a multilevel approach. Recovering mass balance in any situation. *Comput. Methods Appl. Mech. Engrg.* **371** (2020), 113243. https://doi.org/10.1016/j.cma.2020.113243.

[211] Patera, A. T., and Rønquist, E. M. A general output bound result: application to discretization and iteration error estimation and control. *Math. Models Methods Appl. Sci.* **11** (2001), 685–712. http://dx.doi.org/10.1142/S0218202501001057.

[212] Payne, L. E., and Weinberger, H. F. An optimal Poincaré inequality for convex domains. *Arch. Rational Mech. Anal.* **5** (1960), 286–292.

[213] Pencheva, G. V., Vohralík, M., Wheeler, M. F., and Wildey, T. Robust a posteriori error control and adaptivity for multiscale, multinumerics, and mortar coupling. *SIAM J. Numer. Anal.* **51** (2013), 526–554. http://dx.doi.org/10.1137/110839047.

[214] Picasso, M. Adaptive finite elements for a linear parabolic problem. *Comput. Methods Appl. Mech. Engrg.* **167** (1998), 223–237. http://dx.doi.org/10.1016/S0045-7825(98)00121-2.

[215] Picasso, M. A stopping criterion for the conjugate gradient algorithm in the framework of anisotropic adaptive finite elements. *Comm. Numer. Methods Engrg.* **25** (2009), 339–355. https://doi.org/10.1002/cnm.1120.

[216] Prager, W., and Synge, J. L. Approximations in elasticity based on the concept of function space. *Quart. Appl. Math.* **5** (1947), 241–269.

[217] Rannacher, R., and Vihharev, J. Adaptive finite element analysis of nonlinear problems: balancing of discretization and iteration errors. *J. Numer. Math.* **21** (2013), 23–61. https://doi.org/10.1515/jnum-2013-0002.

[218] 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.

[219] Repin, S. *A posteriori estimates for partial differential equations*, vol. **4** of *Radon Series on Computational and Applied Mathematics*. Walter de Gruyter GmbH & Co. KG, Berlin, 2008. http://dx.doi.org/10.1515/9783110203042.

[220] Rey, V., Gosselet, P., and Rey, C. Strict bounding of quantities of interest in computations based on domain decomposition. *Comput. Methods Appl. Mech. Engrg.* **287** (2015), 212–228. http://dx.doi.org/10.1016/j.cma.2015.01.009.

[221] Rey, V., Gosselet, P., and Rey, C. Strict lower bounds with separation of sources of error in non-overlapping domain decomposition methods. *Internat. J. Numer. Methods Engrg.* **108** (2016), 1007–1029. http://dx.doi.org/10.1002/nme.5244.

[222] Rey, V., Rey, C., and Gosselet, P. A strict error bound with separated contributions of the discretization and of the iterative solver in non-overlapping domain decomposition methods. *Comput. Methods Appl. Mech. Engrg.* **270** (2014), 293–303. http://dx.doi.org/10.1016/j.cma.2013.12.001.

[223] Ricois, O. Vision générale du simulateur ARCEOR. Tech. rep., IFPEN, 2011. Number 62096.

[224] Roberts, J. E., and Thomas, J.-M. Mixed and hybrid methods. In *Handbook of Numerical Analysis, Vol. II*. North-Holland, Amsterdam, 1991, pp. 523–639.

[225] Rose, M. E. Compact finite volume methods for the diffusion equation. *J. Sci. Comput.* **4** (1989), 261–290. https://doi.org/10.1007/BF01061058.

[226] Ruiz-Baier, R., and Lunati, I. Mixed finite element–discontinuous finite volume element discretization of a general class of multicontinuum models. *J. Comput. Phys.* **322** (2016), 666–688. https://doi.org/10.1016/j.jcp.2016.06.054.

[227] Russell, T. F., and Wheeler, M. F. Finite element and finite difference methods for continuous flows in porous media. In *The Mathematics of Reservoir Simulation*. SIAM, Philadelphia, 1983, pp. 35–106.

[228] Saad, M., and Zhang, H. Front tracking for two-phase flow in reservoir simulation by adaptive mesh. *Numer. Methods Partial Differential Equations* **13** (1997), 673–697. http://dx.doi.org/10.1002/(SICI)1098-2426(199711)13:6<673::AID-NUM5>3.0.CO;2-O.

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

[230] Sboui, A., Jaffré, J., and Roberts, J. A composite mixed finite element for hexahedral grids. *SIAM J. Sci. Comput.* **31** (2009), 2623–2645. http://dx.doi.org/10.1137/070703703.

[231] Smaï, F. Existence of solutions for a model of multiphase flow in porous media applied to gas migration in underground nuclear waste repository. *Appl. Anal.* **88** (2009), 1609–1616. https://doi.org/10.1080/00036810902942226.

[232] Smears, I. An introduction to the a posteriori error analysis of parabolic partial differential equations. In *Error control, adaptive discretizations, and applications. Part 4*, vol. **61** of *Adv. Appl. Mech.* Academic Press, San Diego, CA, 2025. In press.

[233] Stokke, J. S., Mitra, K., Storvik, E., Both, J. W., and Radu, F. A. An adaptive solution strategy for Richards' equation. *Comput. Math. Appl.* **152** (2023), 155–167. https://doi.org/10.1016/j.camwa.2023.10.020.

[234] Süli, E. A posteriori error analysis and global error control for adaptive finite volume approximations of hyperbolic problems. In *Numerical analysis 1995 (Dundee, 1995)*, vol. **344** of *Pitman Res. Notes Math. Ser.* Longman, Harlow, 1996, pp. 169–190.

[235] Süli, E. A posteriori error analysis and adaptivity for finite element approximations of hyperbolic problems. In *An introduction to recent developments in theory and numerics for conservation laws (Freiburg/Littenweiler, 1997)*, vol. **5** of *Lect. Notes Comput. Sci. Eng.* Springer, Berlin, 1999, pp. 123–194. https://doi.org/10.1007/978-3-642-58535-7_4.

[236] Synge, J. L. *The hypercircle in mathematical physics: a method for the approximate solution of boundary value problems.* Cambridge University Press, New York, 1957.

[237] Thomas, J.-M. *Sur l'analyse numérique des méthodes d'éléments finis hybrides et mixtes.* Ph.D. dissertation, Université Pierre et Marie Curie (Paris 6), May 1977.

[238] Tobiska, L., and Verfürth, R. Robust *a posteriori* error estimates for stabilized finite element methods. *IMA J. Numer. Anal.* **35** (2015), 1652–1671. https://doi.org/10.1093/imanum/dru060.

[239] 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** (1992), 631–644.

[240] Verfürth, R. A posteriori error estimators for convection-diffusion equations. *Numer. Math.* **80** (1998), 641–663. https://doi.org/10.1007/s002110050381.

[241] Verfürth, R. Robust a posteriori error estimators for a singularly perturbed reaction-diffusion equation. *Numer. Math.* **78** (1998), 479–493. https://doi.org/10.1007/s002110050322.

[242] Verfürth, R. A posteriori error estimates for finite element discretizations of the heat equation. *Calcolo* **40** (2003), 195–212. http://dx.doi.org/10.1007/s10092-003-0073-2.

[243] Verfürth, R. Robust a posteriori error estimates for nonstationary convection-diffusion equations. *SIAM J. Numer. Anal.* **43** (2005), 1783–1802. http://dx.doi.org/10.1137/040604273.

[244] Verfürth, R. Robust a posteriori error estimates for stationary convection-diffusion equations. *SIAM J. Numer. Anal.* **43** (2005), 1766–1782. http://dx.doi.org/10.1137/040604261.

[245] Verfürth, R. *A posteriori error estimation techniques for finite element methods.* Numerical Mathematics and Scientific Computation. Oxford University Press, Oxford, 2013. http://dx.doi.org/10.1093/acprof:oso/9780199679423.001.0001.

[246] Vohralík, M. Equivalence between lowest-order mixed finite element and multi-point finite volume methods on simplicial meshes. *M2AN Math. Model. Numer. Anal.* **40** (2006), 367–391. http://dx.doi.org/10.1051/m2an:2006013.

[247] Vohralík, M. A posteriori error estimates for lowest-order mixed finite element discretizations of convection-diffusion-reaction equations. *SIAM J. Numer. Anal.* **45** (2007), 1570–1599. http://dx.doi.org/10.1137/060653184.

[248] Vohralík, M. A posteriori error estimation in the conforming finite element method based on its local conservativity and using local minimization. *C. R. Math. Acad. Sci. Paris* **346** (2008), 687–690. http://dx.doi.org/10.1016/j.crma.2008.03.006.

[249] Vohralík, M. Residual flux-based a posteriori error estimates for finite volume and related locally conservative methods. *Numer. Math.* **111** (2008), 121–158. http://dx.doi.org/10.1007/s00211-008-0168-4.

[250] Vohralík, M. Unified primal formulation-based a priori and a posteriori error analysis of mixed finite element methods. *Math. Comp.* **79** (2010), 2001–2032. http://dx.doi.org/10.1090/S0025-5718-2010-02375-0.

[251] Vohralík, M. Lecture notes. *A posteriori error estimates for efficiency and error control in numerical simulations.* Charles University, Prague, 2024. https://who.rocq.inria.fr/Martin.Vohralik/Enseig/APost/a_posteriori.pdf.

[252] Vohralík, M. Lecture notes. *Analysis and approximation of partial differential equations by finite elements.* ENSTA, Institut Polytechnique de Paris, 2025. https://who.rocq.inria.fr/Martin.Vohralik/Enseig/FEM/FEM.pdf.

[253] Vohralík, M., and Wheeler, M. F. A posteriori error estimates, stopping criteria, and adaptivity for two-phase flows. *Comput. Geosci.* **17** (2013), 789–812. http://dx.doi.org/10.1007/s10596-013-9356-0.

[254] 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.* **23** (2013), 803–838. https://doi.org/10.1142/S0218202512500613.

[255] Vohralík, M., and Yousef, S. A simple a posteriori estimate on general polytopal meshes with applications to complex porous media flows. *Comput. Methods Appl. Mech. Engrg.* **331** (2018), 728–760. https://doi.org/10.1016/j.cma.2017.11.027.

[256] Wheeler, M., Xue, G., and Yotov, I. A multipoint flux mixed finite element method on distorted quadrilaterals and hexahedra. *Numer. Math.* **121** (2012), 165–204. http://dx.doi.org/10.1007/s00211-011-0427-7.

[257] Wheeler, M. F., and Yotov, I. A multipoint flux mixed finite element method. *SIAM J. Numer. Anal.* **44** (2006), 2082–2106. http://dx.doi.org/10.1137/050638473.

[258] Whitney, H. *Geometric integration theory*. Princeton University Press, Princeton, NJ, 1957.

[259] Younès, 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** (2004), 365–388.

[260] Younès, A., Mosé, 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** (1999), 148–167.

[261] Young, L. C., and Stephenson, R. E. A generalized compositional approach for reservoir simulation. *Society of Petroleum Engineers Journal* **23** (1983), 727–742. https://doi.org/10.2118/10516-PA.

[262] Yousef, S. *A posteriori error estimates and adaptivity based on stopping criteria and adaptive mesh refinement for multiphase and thermal flows. Application to steam-assisted gravity drainage*. Ph.D. thesis, Université Pierre et Marie Curie - Paris VI, 2013. https://tel.archives-ouvertes.fr/tel-00918782.

[263] Zeidler, E. *Nonlinear functional analysis and its applications. II/B*. Springer-Verlag, New York, 1990. https://doi.org/10.1007/978-1-4612-0985-0.