Comparison of finite-volume schemes for diffusion problems

Martin Schneider*, Dennis Gläser, Bernd Flemisch, and Rainer Helmig

Institute for Modelling Hydraulic and Environmental Systems, University of Stuttgart, Pfaffenwaldring 61, 70569 Stuttgart, Germany

Received: 13 February 2018 / Accepted: 18 September 2018

Abstract. We present an abstract discretization framework and demonstrate that various cell-centered and hybrid finite-volume schemes fit into it. The different schemes considered in this work are then analyzed numerically for an elliptic model problem with respect to the properties consistency, coercivity, extremum principles, and sparsity. The test cases presented comprise of two- and three-dimensional setups, mildly and highly anisotropic tensors and grids of different complexities. The results show that all schemes show a similar convergence behavior, except for the two-point flux approximation scheme, and seem to be coercive. Furthermore, they confirm that linear schemes, in contrast to nonlinear schemes, are in general neither positivity-preserving nor satisfy discrete minimum or maximum principles.

1 Introduction

Numerical simulations play a crucial role in a wide range of geotechnical engineering applications, thus, it is of great importance that the schemes used to produce these results are reliable and robust. Additionally, efficient schemes are desirable due to the often very large spatial scales involved and the uncertain nature of the subsurface, which might require statistical analyses with a large number of model runs to be performed. Appropriate schemes have to be chosen application-dependent subject to the particular requirements. Besides consistency, efficiency, reliability, and robustness, local mass conservation is essential when performing subsurface flow simulations. This is why the most commonly used schemes for subsurface flow simulations are cell-centered finite-volume methods, such as cell-centered Galerkin methods [1] or Multi-Point Flux Approximation (MPFA) methods [2–7]; or hybrid, mixed, and mimetic (HMM) schemes, such as the Hybrid Finite-Volume (HFV) schemes [8, 9], the Mixed Finite-Element (MFE) [10, 11] or the Mimetic Finite-Difference (MFD) methods [12, 13]. All these schemes are closely related and they can be either presented within a finite-volume framework (which is done in this article) or within a finite-element framework. Such aspects have been presented in [8, 14–16]. The advantage of HMM methods is the fact that they can be applied to highly complex grids, e.g., corner-point grids. However, this comes with the cost of additional unknowns. Recently, monotone or discrete extremum-principles-preserving schemes have been developed in [17–22], and applied to highly complex porous media applications in [23–25].

In this work, we compare different finite-volume schemes for an elliptic model problem regarding convergence, consistency, and efficiency. The latter is measured via the sparsity of the resulting linear systems of equations and the number of nonlinear solver iterations in the case of nonlinear schemes, while consistency is estimated based on discrete extremum principles and the linearity preservation property. Definitions of these properties as well as the numerical schemes considered in this work, covering a number of cell-centered and hybrid schemes, are presented in Section 2. The convergence of the schemes is investigated in Section 3.1, while in Section 3.2 the linearity preservation property is studied. The satisfaction of discrete extremum principles is considered in Section 3.3, before in Section 3.4 all schemes are applied to a three-dimensional benchmark case. Finally, in Section 4 we summarize the results obtained in this article.

2 Discretization schemes

Let \( \Omega \subset \mathbb{R}^d \), \( d \in \mathbb{N}^* \), be an open bounded connected polygonal domain with boundary \( \partial \Omega \), and \( d \)-dimensional measure \( |\Omega| \). In the following, we consider the elliptic problem

$$
\begin{align*}
\nabla \cdot (-\Lambda \nabla u) &= f & & \text{in } \Omega, \\
\quad u &= 0 & & \text{on } \partial \Omega.
\end{align*}
$$

(H1)

Here, we assume that \( f \in L^2(\Omega) \) and \( \Lambda \) is a symmetric tensor-valued function such that (s.t.) the spectrum of \( \Lambda(x) \) is contained in \([\alpha_0, \beta_0]\), with \( 0 < \alpha_0 < \beta_0 < +\infty \), for

* Corresponding author:
martin.schneider@iws.uni-stuttgart.de

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
almost every (a.e.) \( x \in \Omega \). These assumptions guarantee the existence of a weak solution \( \tilde{u} \) (i.e. a solution of the variational formulation of problem (1)).

**Remark 1.** Within this section, homogeneous Dirichlet boundary conditions are assumed for ease of presentation. Other types of boundary conditions have for example been discussed within the gradient discretization framework [26].

In the following an admissible discretization is defined together with the notations that are used within this article.

**Definition 1** (Admissible discretization). An admissible discretization \( D \) is a triplet \( D = (T, E, P) \), where

(i) \( T \) are the cells s.t. \( \Omega = \bigcup_{K \in T} K \). For each cell \( K \in T \), \( |K| > 0 \) denotes the cell volume and \( \partial K \) \( \text{def} = \overline{K \setminus K} \) its boundary, \( h_D \text{def} = \sup_{K \in T} \text{diam}(K) \) defines the discretization length.

(ii) \( E \) are the faces such that each face \( \sigma \) is included in a hyperplane of \( \mathbb{R}^d \) with \((d-1)\)-dimensional measure \(|\sigma| > 0\). For each cell \( K \in T \), \( E_K \) is the subset of \( E \) such that \( \partial K = \bigcup_{\sigma \in E_K} \sigma \). \( T_\sigma \text{def} = \{ K \in T \mid \sigma \in E_K \} \) contains the cells that share the face \( \sigma \); the sets of inner and boundary faces are denoted by \( E_{\text{int}} \) and \( E_{\text{ext}} \), respectively.

(iii) \( P = \{ x_K \}_{K \in T} \) are the cell centers (not required to be the barycenters) s.t. \( x_K \in K \) and \( K \) is star-shaped with respect to \( x_K \). For all \( K \in T \), \( \sigma \in E_K \), \( d_{K,\sigma} \) is given as the Euclidean distance between \( x_K \) and \( \sigma \).

For a more detailed definition of an admissible discretization see [3, 4, 21]. Let \( M \) be any set, then the cardinality of this set is indicated with \( n_M \). With this definition, the number of cells, i.e. the cardinality of the set \( T \), is given by \( n_T \). The face evaluation points are denoted as \( \mathbf{x}_\sigma \), \( \sigma \in E \) (not required to be the barycenters). With \( \mathbf{A}_K \text{def} = (A_{K,\sigma} \text{def} = \frac{1}{|K|} \int_K A(x) dx) \) we define the averaged tensor on cell \( K \), where the integral is meant component-wise. In addition, it is assumed that \( A_{K,\sigma} \in \mathbb{C}_2^2(\Omega) \) for all \( K \in T \). Furthermore, \( n^{K,\sigma} \) denotes the unit vector that is normal to \( \sigma \) and outward to \( K \).

### 2.1 Abstract discretization framework

Integration of equation (1) over the control volume \( K \in T \) yields

\[
- \int_K (A \nabla u) \cdot n_K \, dx = \int_K f \, dx. \tag{2}
\]

Using Definition 1, the integral on the left of equation (2) can be split into integrals over the faces \( \sigma \), which results in

\[
- \sum_{\sigma \in E_K} \int_\sigma (A \nabla u) \cdot n_{K,\sigma} \, dx = \int_K f \, dx. \tag{3}
\]

To discretize equation (3) we need to define flux approximations \( F_{K,\sigma} \), for each cell \( K \) and face \( \sigma \in E_K \), such that

\[
F_{K,\sigma} \approx \int_\sigma (A \nabla u) \cdot n_{K,\sigma} \, dx, \tag{4}
\]

i.e. \( F_{K,\sigma} \) is an approximation of the exact flux.

For this purpose, we define the following discrete solution spaces

\[
X_T \text{def} = \{ v = (v_K)_{K \in T}, v_K \in \mathbb{R} \}, \tag{5}
\]

\[
X_\varepsilon \text{def} = \{ v = (v_\sigma)_{\sigma \in E}, v_\sigma \in \mathbb{R} \}, \tag{6}
\]

\[
X_D \text{def} = X_T \times X_\varepsilon. \tag{7}
\]

For any element \( v \in X_D \), the cell unknowns are denoted by \( v_K \in X_T \) and the face unknowns are denoted by \( v_\sigma \in X_\varepsilon \). The space that takes into account the homogeneous zero Dirichlet boundary conditions is accordingly given as

\[
X_{D,0} \text{def} = \{ v \in X_D \mid v_\sigma = 0, \forall \sigma \in E_{\text{ext}} \}. \tag{8}
\]

Additionally, the space of piecewise constant functions on \( T \) is defined as

\[
H_T \text{def} = \{ v \in L^2(\Omega) \mid \forall K \in T, \forall x \in K, v_I(x) = v_K, v_K \in \mathbb{R} \}. \tag{9}
\]

Therefore, for all \( v \in H_T \) and for all \( K \in T, v_K \) will denote the constant value of \( v \) on \( K \). Furthermore, the extraction operator is defined for all \( v \in X_D \) as

\[
\Pi_T : X_D \mapsto H_T, \text{s.t., } \forall K \in T, (\Pi_T v)_K = v_K. \tag{10}
\]

The presented finite-volume schemes differ in the choice of the fluxes \( F_{K,\sigma} \) and the choice of the solution space \( X_h \subseteq X_{D,0} \). Therefore, for all \( K \in T \), and for all \( \sigma \in E_K \), let \( F_{K,\sigma} : X_h \mapsto \mathbb{R} \) be a numerical flux function that approximates the flux flowing out of \( K \) through \( \sigma \) such that the finite-volume scheme reads: Find \( u \in X_h \) s.t.

\[
- \sum_{\sigma \in E_K} F_{K,\sigma}(u) = \int_K f \, dx, \forall K \in T, \tag{10}
\]

\[
F_{K,\sigma}(u) + F_{L,\sigma}(u) = 0, \forall \sigma \in E_{\text{int}}, T_\sigma = \{ K, L \}. \tag{11}
\]

**Cell-centered schemes:**

For cell-centered schemes the solution space is given as

\[
X_h = \{ v = (v_T, v_\varepsilon) \in X_{D,0} \mid v_\sigma = I_\varepsilon v_T, \forall \sigma \in E_{\text{int}} \}, \tag{12}
\]

whereby, for each face \( \sigma \), \( I_\varepsilon \in \mathcal{L}(X_T; \mathbb{R}) \) is a trace reconstruction operator. Here, \( \mathcal{L}(X_T; \mathbb{R}) \) is the space of linear operators that map some element \( u \in X_T \) to a constant value associated with the corresponding face. This means that for cell-centered schemes, the face unknowns are eliminated using some trace reconstruction operators \( I = \{ I_\sigma \}_{\sigma \in E} \).

Here, only locally mass-conservative cell-centered schemes are considered such that for all \( u \in X_h, \sigma \in E_{\text{int}} \), with \( T_\sigma = \{ K, L \} \)

\[
F_{K,\sigma}(u) + F_{L,\sigma}(u) = 0. \tag{13}
\]

This means that equation (11) is fulfilled by construction and holds not only for the discrete solution.
**Hybrid Mixed Mimetic (HMM) schemes:**

For HMM schemes (see [14] for a detailed description of such schemes) the solution space is given as

\[ X_h = X_{D,0}. \] 

(14)

This means that HMM schemes introduce additional face unknowns, in contrast to cell-centered schemes where these unknowns are eliminated.

For the locally mass-conservative cell-centered schemes with solution space (12) and for the HMM schemes with solution space (14), it can be shown that problem (10)–(11) is equivalent to the problem: Find \( u \in X_h \) s.t.

\[ a_D(u, v) = \int_{\Omega} f \Pi_T v \, dx, \quad \forall v \in X_h, \] 

(15)

with the form, for all \((u,v) \in [X_h]^2,

\[ a_D(u, v) \overset{\text{def}}{=} \sum_{K \in T} \sum_{\sigma \in \partial K} F_{K,\sigma}(u)(v_{\sigma} - v_K). \] 

(16)

Therefore, each solution of problem (10) and (11) also solves problem (15). Starting from the discrete problem (15), with \( a_D \) defined by (16), equation (11) is fulfilled for the cell-centered schemes by construction, whereas for the HMM schemes it is obtained by taking \( v_{\sigma} = 1, v_K = 0 \) for all \( \sigma \neq \sigma' \), and \( v_K = 0 \) for all \( K \).

Equation (10) is obtained by taking, for each \( K \in T, v_K = 1 \) and \( v_{\sigma'} = 0 \) for all \( \sigma' \neq \sigma \) and by using the flux continuity (11). This shows that the formulation (10) and (11) is equivalent to (15).

**Remark 2.** If the flux functions are linear, then the form (16) is linear in both arguments. However, in Section 2.3 we will also consider nonlinear flux functions such that the linearity is only given for the second argument.

The advantage of the formulation (15) is that properties like coercivity can easily be defined for the form \( a_D \). Such properties are briefly defined in the next section.

**2.2 Properties of discretization schemes**

Besides simplicity, parallelizability, computational efficiency, flexibility, and code integrability, mathematical and physical properties of discretization schemes, *e.g.* consistency, coercivity or monotonicity, can be defined. Unfortunately, to the author's knowledge, there is currently no scheme that satisfies all these properties. Therefore, appropriate schemes have to be chosen application dependent. In general, the convergence of schemes can be proven if the scheme is consistent and coercive, as it has been done in [4, 8, 12, 21]. In the following, we briefly describe such fundamental properties for cell-centered schemes (with discrete solution space (12)) and for HMM schemes (with discrete solution space (14)).

**Consistency**

There is no unique discretization-independent definition of consistency. For example, finite-volume schemes are in general not consistent in the finite-difference context [27].

In general, a scheme is consistent if the truncation error between discrete and continuous operators goes to zero for \( h_D \to 0 \). In the context of finite-volume schemes we say that a scheme is consistent if the numerical flux approximates the exact flux for regular functions, meaning that

\[ F_{K,\sigma}(\phi_D) = F_{K,\sigma}(\phi) + O(|\sigma|) \text{diam}(K), \quad \forall \phi \in D, \] 

(17)

where \( D \subset C_0(\Omega) \) is a test function space which is assumed to be dense in \( H^1(\Omega) \), and \( F_{K,\sigma}, F_{K,\sigma} \) are the discrete and exact flux functions, respectively. Furthermore, \( \phi_D = (\phi_T, \phi_E) \in X_D \) is defined as \( (\phi_T)_K = \phi(x_K) \) for all \( K \in T \), and \( (\phi_E)_\sigma = \phi(x_\sigma) \) for all \( \sigma \in E \). More details can for example be found in [21].

**Coercivity**

A scheme is called uniformly coercive if the form \( a_D \), defined in (16), satisfies the following estimate:

\[ a_D(u, u) \geq \gamma \| u \|_{X_h}^2, \quad \forall u \in X_h, \] 

(18)

with some appropriate discrete norm \( \| \cdot \|_{X_h} \), and \( \gamma > 0 \) that is independent of \( u \) and \( h_D \). When talking about coercivity in the following, we always refer to the uniform coercivity property.

**Minimum and maximum principles**

It is desirable that the discrete solution satisfies properties of the exact solution. An important property from a physical point of view is the minimum and maximum principle. Schemes that satisfy these principles prevent oscillations of the discrete solutions, such that the discrete solution remains within physical bounds. Such extremum principles can be found in [28]. For the definition of discrete minimum and maximum principles, we refer to the overview provided by [29]. Furthermore, a scheme is said to be monotone if the discretization matrix \( A \) is monotone, which means that all entries of its inverse are non-negative *i.e.* \( A^{-1} \geq 0 \). The monotonicity property is for example satisfied by M-matrices. M-matrices are monotone matrices with non-positive off-diagonal entries (see, *e.g.* [30]).

**Sparsity**

For solving large-scale systems, it is indispensable that the discretization matrices are sparse, which means that the number of entries (noe) in the matrix is significantly smaller than \( n_T^2 \). The stencil denotes the noe of each row. Therefore, schemes resulting in small stencils are important for large-scale problems.

**2.3 Cell-centered schemes**

Within this section, we present different cell-centered finite-volume schemes. Here, the discrete solution space is given as (12), and the face values are eliminated using some trace reconstruction operators \( I_{E,\sigma} \). As already mentioned,
only locally mass-conservative schemes are considered here such that equation (11) holds by construction.

**Family of weighted schemes**

An established idea to obtain monotone or extremum-principles-preserving schemes, as those developed in [17–20, 22, 23, 31, 32], is to combine, for each interior edge \( \sigma \in \mathcal{E}_{\text{int}} \), with \( T_{\sigma} = \{ K, L \} \), two consistent linear flux approximations \( \hat{F}_{K,\sigma}(u) \) and \( \hat{F}_{L,\sigma}(u) \) as a convex combination, using solution-dependent weights. Thus, the final flux \( F_{K,\sigma}(u) \) is given as

\[
F_{K,\sigma}(u) = \mu_{K,\sigma}(u) \hat{F}_{K,\sigma}(u) - \mu_{L,\sigma}(u) \hat{F}_{L,\sigma}(u),
\]

with \( \mu_{K,\sigma}(u) \geq 0, \mu_{L,\sigma}(u) \geq 0 \),

and \( \mu_{K,\sigma}(u) + \mu_{L,\sigma}(u) = 1 \). (19)

For any \( K \in T \) and \( \sigma \in \mathcal{E}_K \cap \mathcal{E}_{\text{int}} \), the linear flux \( \hat{F}_{K,\sigma}(u) \) is built in order to ensure a strong consistency property, such that the total flux \( F_{K,\sigma} \) satisfies a weak form of consistency, see [21] for detailed information. A general definition of these sub-fluxes is

\[
\hat{F}_{K,\sigma}(u) \overset{\text{def}}{=} \left| \sigma \right| \sum_{\sigma' \in \mathcal{E}_{K,\sigma}} x_{K,\sigma'} (I_{\sigma'} u - u_K),
\]

with face stencil \( S_{K,\sigma} \) and trace reconstruction operator \( I_{\sigma'} \).

The coefficients \( \mu_{K,\sigma} \) and \( \mu_{L,\sigma} \) are chosen to include some of the previously mentioned properties, e.g. monotonicity. These coefficients may nonlinearly depend on the unknown \( u \). Therefore, the form \( a_\sigma \) is also nonlinear in its first argument. This can be prevented by introducing an additional argument for the fluxes and correspondingly for the form \( a_\sigma \). The details are not presented here, for a detailed description see [21].

The final flux function (19) is locally mass-conservative, whereas the sub-fluxes \( \hat{F} \) are in general not mass-conservative. In the following, different schemes that fit into this family are shortly introduced.

**AvgMPFA scheme**

The easiest choice of coefficients is \( \mu_{K,\sigma} = \mu_{L,\sigma} = 0.5 \), which results in a linear finite-volume scheme that is in the following denoted as AvgMPFA scheme.

**NLTPFA scheme**

To derive a Nonlinear Two-Point Flux Approximation (NLTPFA), the different terms are reordered such that the flux is written as

\[
F_{K,\sigma}(u) = t_{L,\sigma}(u)u_L - t_{K,\sigma}(u)u_K - (\mu_{L,\sigma}(u)\hat{t}_{L,\sigma}(u) - \mu_{K,\sigma}(u)\hat{t}_{K,\sigma}(u)).
\]

The idea of the NLTPFA scheme is to choose the weights such that \( R_{K,\sigma}(u) = 0 \). From a numerical point of view, it is sufficient that \(|R_{K,\sigma}(u)| \leq \epsilon\). This is given, under the assumption that \( \hat{t}_{K,\sigma} t_{L,\sigma} \geq 0 \), for the choice

\[
\mu_{K,\sigma}(u) = \frac{|\hat{t}_{L,\sigma}(u)| + \epsilon}{|\hat{t}_{K,\sigma}(u)| + |\hat{t}_{L,\sigma}(u)| + 2\epsilon},
\]

\[
\mu_{L,\sigma}(u) = \frac{|\hat{t}_{K,\sigma}(u)| + \epsilon}{|\hat{t}_{K,\sigma}(u)| + |\hat{t}_{L,\sigma}(u)| + 2\epsilon}.
\]

Further details and the discussion of the case \( \hat{t}_{K,\sigma} t_{L,\sigma} < 0 \) can be found in [21], where the monotonicity of the NLTPFA scheme is also discussed. Here, \( \epsilon \) is set to \( 10^{-8} \) but any other small number could also be chosen.

**NLMPFA scheme**

A Nonlinear Multi-Point Flux Approximation (NLMPFA) is derived by splitting the sub-fluxes into a two-point part and a residual flux part

\[
\begin{align*}
F_{K,\sigma}(u) &= |\sigma| \beta_\sigma(u_L - u_K) + \tilde{F}_{K,\sigma}^{\text{res}}(u), \\
\tilde{F}_{L,\sigma}(u) &= |\sigma| \beta_\sigma(u_K - u_L) + \tilde{F}_{L,\sigma}^{\text{res}}(u),
\end{align*}
\]

with some appropriate \( \beta_\sigma \) and \( \tilde{F}_{K,\sigma}^{\text{res}}, \tilde{F}_{L,\sigma}^{\text{res}} \). Here, the weights are chosen as

\[
\begin{align*}
\mu_{K,\sigma}(u) &= \left( \frac{|\tilde{F}_{K,\sigma}^{\text{res}}(u)|}{|\tilde{F}_{K,\sigma}^{\text{res}}(u)| + |\tilde{F}_{L,\sigma}^{\text{res}}(u)|} \right) + \epsilon, \\
\mu_{L,\sigma}(u) &= \left( \frac{|\tilde{F}_{L,\sigma}^{\text{res}}(u)|}{|\tilde{F}_{K,\sigma}^{\text{res}}(u)| + |\tilde{F}_{L,\sigma}^{\text{res}}(u)|} \right) + \epsilon.
\end{align*}
\]

Again, \( \epsilon \) is added due to numerical reasons. The final flux is then given by (19) with weights (24). It can be shown that this scheme satisfies discrete extremum principles.

**Remark 3.** Without going into detail, we shortly summarize the main differences between the NLTPFA and the NLMPFA scheme. For a detailed description we refer to [21]. Both schemes mainly differ in the choice of the weights \( \mu_{K,\sigma} \) and \( \mu_{L,\sigma} \) (the coefficients \( x_{K,\sigma'} \) in the sub-fluxes (20) may also vary). The weights of the NLTPFA scheme are defined by \( \hat{t}_{K,\sigma} \) and \( \hat{t}_{L,\sigma} \), which are written as a sum of solution values, i.e. \( \hat{t}_{K,\sigma} = \sum \omega_{M,i} I_{\sigma'}(u) \). Whereas, those of the NLMPFA scheme are defined by the residual fluxes \( \tilde{F}_{K,\sigma}^{\text{res}} \) and \( \tilde{F}_{L,\sigma}^{\text{res}} \), which can be written as a sum of solution value differences, i.e. \( \tilde{F}_{K,\sigma}^{\text{res}} = \sum \omega_{M,i} (u_{M,i} - u_{M,j}) \).

**Linear TPFA/MPFA schemes**

Here, we shortly demonstrate that well-established schemes such as the Two-Point Flux Approximation (TPFA) or the Multi-Point Flux Approximation (MPFA) also fit into this family of schemes. The difference here is that the sub-fluxes are constructed to satisfy flux continuity, i.e. \( \hat{F}_{K,\sigma} + \hat{F}_{L,\sigma} = 0 \). Therefore, the final flux is given as \( F_{K,\sigma} = \hat{F}_{K,\sigma} \) (20), independent of the weights \( \mu_{K,\sigma}, \mu_{L,\sigma} \).

For the TPFA scheme it holds that \( S_{K,\sigma} = \{ \sigma \} \) and the coefficients are

\[
x_{K,\sigma} = \frac{\mathbf{n}_K^T \Lambda_{K}(x_K - x_L)}{||x_K - x_L||_2^2}, \quad x_{L,\sigma} = \frac{\mathbf{n}_L^T \Lambda_{L}(x_K - x_L)}{||x_K - x_L||_2^2}.
\]

- **M. Schneider et al.: Oil & Gas Science and Technology - Rev. IFP Energies nouvelles**
The trace reconstruction operator is
\[ I_{\sigma}u = \frac{\gamma_{K,\sigma}u_K + \gamma_{L,\alpha}u_L}{\gamma_{K,\sigma} + \gamma_{L,\alpha}}, \]  
(26)
such that the final flux reads
\[ F_{K,\sigma}(u) = |\sigma| \frac{\gamma_{K,\sigma}u_K + \gamma_{L,\alpha}u_L}{\gamma_{K,\sigma} + \gamma_{L,\alpha}} (u_L - u_K). \]  
(27)

This can be similarly shown for MPFA schemes but the details are not presented here.

### 2.4 Hybrid/mimetic schemes

Within this section, we present hybrid and mimetic schemes. Here, the discrete solution space is given as (14), such that additional face unknowns are introduced. Here, the fluxes are given, for all \( u \in \mathcal{X}_h, K \in \mathcal{T}, \sigma \in \mathcal{E}_k \), as
\[ F_{K,\sigma}(u) \equiv |\sigma| \sum_{\sigma' \in \mathcal{E}_k} \gamma_{K,\sigma'}(u_{\sigma'} - u_K), \]  
(28)
with the face unknowns \( u_{\sigma'} \), the cell unknowns \( u_K \), and coefficients \( \gamma_{K,\sigma'} \). These coefficients are chosen such that the resulting scheme is coercive and consistent.

Defining the matrices,
\[ \mathbf{u}_{\sigma_k} \equiv (u_\sigma)_{\sigma \in \mathcal{E}_k} = \begin{pmatrix} u_{\sigma_1} \\ \vdots \\ u_{\sigma_{|\mathcal{E}_k|}} \end{pmatrix}, \quad \mathbf{C}_K = \text{diag}(|\sigma_1|, \ldots, |\sigma_{n_{\mathcal{E}_k}}|), \]  
(29)
we obtain the cell flux vector as
\[ \mathbf{F}_{K,\mathcal{E}_k} \equiv \begin{pmatrix} F_{K,\sigma_1} \\ \vdots \\ F_{K,\sigma_{|\mathcal{E}_k|}} \end{pmatrix} = \mathbf{C}_K \mathbf{W}_K \mathbf{C}_K (\mathbf{u}_{\mathcal{E}_k} - u_K \mathbf{e}), \]  
(30)
with \( \mathbf{W}_K = \mathbf{A}_K \mathbf{C}_K^{-1}, \mathbf{A}_K,\nu_{\sigma'} = \gamma_{K,\sigma'}, \) and \( \mathbf{e} \) is the vector with entries equal to one. Additionally, let \( \mathbb{N}_K \) be the conormal matrix, and let \( \mathbb{R}_K \) be the matrix that contains the scaled distance vectors:
\[ \mathbb{N}_K = \begin{pmatrix} \mathbf{n}_{K,\sigma_1} \mathbf{A}_K \\ \vdots \\ \mathbf{n}_{K,\sigma_{n_{\mathcal{E}_k}}} \mathbf{A}_K \end{pmatrix}, \quad \mathbb{R}_K = \begin{pmatrix} |\sigma_1|(x_{\sigma_1} - x_K)^T \\ \vdots \\ |\sigma_{n_{\mathcal{E}_k}}|(x_{\sigma_{n_{\mathcal{E}_k}}} - x_K)^T \end{pmatrix}. \]  
(31)

Then, the consistency condition is given by
\[ \mathbb{N}_K = \mathbb{W}_K \mathbb{R}_K. \]  
(32)

From this consistency condition, we can derive a general form of the coefficient matrix \( \mathbb{W}_K \):
\[ \mathbb{W}_K = \mathbb{N}_K (\mathbb{N}_K^T \mathbb{R}_K)^{-1} \mathbb{N}_K^T + \mathbb{S}_K, \]  
(33)
with a stabilization matrix \( \mathbb{S}_K \) such that \( \mathbb{S}_K \mathbb{R}_K = \mathbb{0} \). Choosing \( x_\sigma \) as the barycenter of face \( \sigma \) and using the geometrical identity
\[ \sum_{\sigma \in \mathcal{E}_k} |\sigma| (n_{K,\sigma}(x_\sigma - x_K))^T = |K|, \quad \forall K \in \mathcal{T}, \]  
(34)
it can be seen that
\[ \mathbb{N}_K^T \mathbb{R}_K = |K| \mathbf{A}_K. \]  
(35)

Therefore, the matrix \( \mathbb{W}_K \) is symmetric if the stabilization matrix \( \mathbb{S}_K \) is chosen symmetric.

#### MFD scheme

A simple choice of \( \mathbb{S}_K \) is
\[ \mathbb{S}_K = v_K (1 - \mathbb{R}_K (\mathbb{R}_K^T \mathbb{R}_K)^{-1} \mathbb{R}_K^T), \]  
(36)
with \( v_K = \frac{1}{2} \text{trace}(\mathbb{N}_K (\mathbb{N}_K^T \mathbb{R}_K)^{-1} \mathbb{N}_K^T) \). This results in the final matrix
\[ \mathbb{W}_K = \mathbb{N}_K (\mathbb{N}_K^T \mathbb{R}_K)^{-1} \mathbb{N}_K^T + v_K (1 - \mathbb{R}_K (\mathbb{R}_K^T \mathbb{R}_K)^{-1} \mathbb{R}_K^T). \]  
(37)

For this matrix, the consistency condition (32) is fulfilled by construction, whereas the coercivity can be proven by using some kind of stability condition, see [33].

#### HFV scheme

Another interesting scheme that fits into this framework, and also fits into the recently developed gradient discretization framework [15], is the Hybrid Finite-Volume (HFV) scheme [8]. The main idea of this scheme is the construction of a consistent discrete gradient, which at the same time defines the form \( a_D \), such that the consistency is naturally fulfilled. In the following we shortly introduce the scheme and present the main ideas.

Let us recall the weak formulation of problem (1): Find \( \pi \in H_0^1(\Omega) \) such that
\[ \int_\Omega \mathbf{A} \nabla \pi \cdot \nabla v \, dx = \int_\Omega f v \, dx, \quad \forall v \in H_0^1(\Omega). \]

The idea of symmetric gradient discretization schemes is to replace the operator \( \nabla \) with a consistent approximation \( \mathbf{\nabla}_D \) such that the form of the discrete problem (15) is given by
\[ a_D(u, v) \equiv \int_\Omega \mathbf{\nabla}_D u \cdot \mathbf{\nabla}_D v \, dx. \]  
(38)

Therefore, the scheme is defined by the discrete gradient reconstruction operator \( \mathbf{\nabla}_D \). First, let us define a discrete gradient on each cell \( K \in \mathcal{T} \) as
\[ \mathbf{\nabla}_K u = \frac{1}{|K|} \sum_{\sigma \in \mathcal{E}_k} |\sigma| (u_\sigma - u_K) \mathbf{n}_{K,\sigma}, \]  
(39)
the consistency of this formula follows thanks to (34). However, to end up in a coercive form, we need an additional stabilization term, which is defined as
\[ \mathbb{S}_{K,\sigma} u = S_{K,\sigma} \mathbf{n}_{K,\sigma}, \]  
(40)
with
\[ S_{K,\sigma} \equiv \frac{\sqrt{2}}{\gamma_{K,\sigma}} (u_\sigma - u_K - \mathbf{\nabla}_K u \cdot (x_\sigma - x_K)). \]

This results in the final discrete gradient: For all \( K \in \mathcal{T}, \sigma \in \mathcal{E}_k \)
\[ \nabla_D u(x) \overset{\text{def}}{=} \nabla u + S_{K,\sigma} n_{K,\sigma}, \text{ for a.e. } x \in \Delta_{K,\sigma}, \] (41)

where \( \Delta_{K,\sigma} \) is the convex hull that includes the face \( \sigma \) and the cell center \( x_K \). Inserting these discrete gradients into the form (38) and reordering of terms lead to the form that is defined in (16). Thus, the fluxes and the coefficient matrix \( \mathbb{W}_K \) (30) can also be identified. Therefore, the hybrid finite-volume scheme also belongs to the family of hybrid mimetic schemes. The coefficient matrix, a detailed description, and the proof of consistency and coercivity can be found in [8].

Remark 4. For a more detailed summary of finite-volume schemes we refer to [34, 35].

3 Numerical experiments

In this chapter, the behavior of the different finite-volume schemes, that have been presented in the last section, is investigated. The AvgMPFA, NLTPFA, and NLMPFA schemes need the coefficients \( a_{K,\sigma} \), that occur in the sub-fluxes (20). These coefficients are calculated using optimization strategies together with a Primal-Dual Simplex Method provided by the open-source library GNU Linear Programming Kit\(^1\) (GLPK). Detailed explanations can be found in [21, 24]. Furthermore, the harmonic averaging interpolator [35] is used as reconstruction operator \( I_e \), which is needed to define the sub-fluxes (20).

The mimetic finite-difference scheme with fluxes (30) and local cell matrix (37) is denoted as MFD scheme. Finally, with TPFA we denote the scheme with fluxes (27) and with MPFA-O the scheme that is described in [37]. The Box method [38, 39] is a vertex-centered finite-volume scheme that uses finite-element basis functions on each cell to calculate fluxes over sub-control volume faces.

All simulations are performed using the open-source simulator DmUt\(^2\) [40], which comes in the form of an additional DUNE module [41]. Newton’s method is used to solve the nonlinear systems of equations occurring for the nonlinear finite-volume schemes, where the iteration loop is stopped if the absolute residual is below \( 10^{-5} \) with respect to the Euclidean norm.

In this section, more general boundary conditions are considered, where the Dirichlet conditions are taken into account using the function \( g \in H^1(\partial \Omega) \) such that the weak solution \( \pi \in H^1(\Omega) \) has to satisfy \( T\pi = g \) (where \( T : H^1(\Omega) \rightarrow L^2(\partial \Omega) \) denotes the trace operator).

A discrete seminorm on the space \( X_D \) is given by

\[ ||u||_D, \overset{\text{def}}{=} \left( \sum_{K \subseteq T} \left( \sum_{\sigma \in \Delta E_K \cap \partial \Omega} \frac{|\sigma|}{d_{K,\sigma}} (v_{\sigma} - v_K)^2 \right) \right)^{1/2}, \] (42)

which is a norm on the space \( X_{D,0} \). For the cell-centered schemes, we define the following discrete norm (that takes into account the Dirichlet data)

\[ ||u||_{T,\text{Dir}} = \left( \sum_{K \subseteq T} \left( \sum_{\sigma \in \Delta E_K \cap \partial \Omega} \frac{|\sigma|}{d_{K,\sigma}} (v_{\sigma} - v_K)^2 \right) \right)^{1/2}, \] (43)

where \( \langle g \rangle_\sigma \) denotes the average of \( g \) on the face \( \sigma \). Using the Cauchy-Schwarz inequality, one can show that (see [8] for more details)

\[ ||u||_T \leq ||u||_D, \forall u \in X_{D,g}, \] (44)

with

\[ X_{D,g} = \{ v \in X_D | v_\sigma = \langle g \rangle_\sigma, \forall \sigma \in \mathcal{E}_{\text{ext}} \}. \] (45)

Owing to inequality (44), the discrete norm (43) will also be used for the hybrid and mimetic schemes, although it is not a discrete norm for these schemes since it does not depend on the face unknowns. For measuring the coercivity of the schemes, the following estimate is defined

\[ e_D(u) = \frac{a_D(u, u)}{||u||_T^2}. \] (46)

In the next section the coercivity estimate is evaluated for the numerical solution. Please note that this is not sufficient to show the coercivity of the schemes, but it serves as a good indicator.

In the following test cases, the properties that have been mentioned in Section 2.2 are investigated numerically. Hereby, the numerical and exact solutions are denoted as \( u_n \) (where \( n \) indicates the grid refinement level) and \( u \), respectively.

Remark 5. In this section, hexahedral grids are considered (or quadrilateral grids for the case of a two-dimensional domain). Such grids are, in general, not admissible in the sense of Definition 1, because the faces are usually non-planar. Therefore, the averaged normal vectors are calculated which are still denoted by \( n_{K,\sigma} \). For more details, we refer to [24, 37, 42]. In [42], the authors suggest to introduce, for strongly curved faces, additional degrees of freedom accounting for tangential velocities. These additional velocities are related to the planar face which is defined through the averaged normal vector. Such additional degrees of freedom are not introduced here, because we did not observe any convergence rate reduction for the considered examples.

3.1 Convergence behavior

Within this section, the convergence rates of the different schemes are compared for three-dimensional test problems with smooth solutions. The convergence of the family of schemes (19) has been proven, under the assumption of coercivity, in [21]. Here, we demonstrate that the convergence rates are similar to well-established schemes. These examples are based on our previous work [24], but here we are using different norms and a more challenging tensor for the highly anisotropic test case. In addition, the convergence behavior of the MPFA-O, MFD, and HFV are investigated and compared.

\(^1\) http://www.gnu.org/software/glpk/glpk.html

\(^2\) http://www.dune-project.org
The convergence behavior is investigated for the meshes shown in Figure 1. The checkerboard and the random hexahedral meshes are taken from the FVCA6 benchmark [43], whereas the non-convex grid and the curved grid are generated with the Matlab Reservoir Simulation Toolbox (MRST) [44]. For all grids, we use the dune-alugrid module [45]. Except for the checkerboard mesh, all grids exhibit non-planar faces. Therefore, we calculate the integrated normal vectors, see [24]. For the MPFA-O scheme this is applied to the sub-faces of the control volumes, see [37].

Mildly anisotropic test case

The first test case is similar to test 1 from the FVCA6 benchmarks [43], here, the exact solution
\[
\bar{u}(x, y, z) = \sin(\pi x) \sin\left(\pi \left( y + \frac{1}{2} \right) \right) \sin\left(\pi \left( z + \frac{1}{3} \right) \right) + 1,
\]
(47)
is prescribed on the unit cube \( \Omega = [0, 1]^3 \). Dirichlet conditions are set on the boundary as \( g = \bar{u} \) on \( \partial \Omega \). The permeability tensor is
\[
\Lambda = \begin{pmatrix} 1 & 0.5 & 0 \\ 0.5 & 1 & 0.5 \\ 0 & 0.5 & 1 \end{pmatrix}. \tag{48}
\]

The discrete \( L^2 \)-errors and \( H^1 \)-errors are shown in Figure 2. Please note that the results of the AvgMPFA scheme are not shown since they only differ slightly from the NLTPFA results. Furthermore, the MPFA-O scheme cannot be applied to hanging nodes on three-dimensional grids, which is why there are no results of the MPFA-O scheme for the checkerboard mesh.

It can be seen that for increasing mesh refinement, the schemes converge with second order accuracy in the \( L^2 \)-norm and at least first order accuracy in the \( H^1 \)-norm. The convergence rates of all schemes are quite similar. We also observe that the MFD scheme produces the smallest errors. The errors of the NLMPFA scheme are the largest ones for most grids and refinement levels. Additionally, the behavior of the MPFA-O and HFV is similar, especially for the random and non-convex grids. In this example, the NLTPFA scheme produces smaller errors than the NLMPFA scheme. Considering Table 1, we observe that the coercivity estimates \( \varepsilon_D \) are bounded from below which indicates that all schemes are coercive for this test case on all grids. The NLTPFA scheme converges within two Newton iterations, whereas the Newton convergence of the NLMPFA is in general worse.

Highly anisotropic test case

The next test problem investigates a highly anisotropic permeability tensor
\[
\Lambda = \frac{1}{x^2 + y^2} \begin{pmatrix} \beta x^2 + y^2 & (\beta - 1)xy & 0 \\ (\beta - 1)yx & x^2 + \beta y^2 & 0 \\ 0 & 0 & (z^2 + \beta)(x^2 + y^2) \end{pmatrix},
\]
(49)
with \( \beta = 10^{-2} \) and the exact solution
\[
\bar{u}(x, y, z) = \sin(2\pi x) \sin(2\pi y) \sin(2\pi z) + 1. \tag{50}
\]

Again, Dirichlet conditions are set on the whole boundary corresponding to the exact solution. Figure 3 shows the errors for test case two, with solution (50). Again, the results of the AvgMPFA scheme are not shown since they only differ slightly from the NLTPFA results. As in the previous example, no results could be obtained for the MPFA-O scheme on the checkerboard mesh due to the non-conformity of the grid.

In Figure 3, we can see that the NLTPFA scheme produces now similar errors than the HFV and MPFA-O scheme, and results in the smallest \( H^1 \)-errors on the random grid. Considering Table 2, we observe that the coercivity estimates \( \varepsilon_D \) are bounded from below which indicates that all schemes are coercive also for this test case. Again, the NLTPFA scheme converges within two Newton iterations for all grids. However, the NLMPFA needs much more Newton iterations for the checkerboard mesh. The worse convergence behavior is probably caused by the fact that the weights \( \mu_{K,p} \), \( \mu_{l,p} \) are constructed differently for the NLMPFA and NLTPFA schemes, as explained in Remark 3. Sign changes for the functions \( \lambda_{K,p} \) are rare, whereas sign changes occur more often for the residual fluxes \( \tilde{F}_{K,p}^{res} \), which are used to define the weights of the NLMPFA scheme. This explains why the Newton method frequently passes points of non-differentiability of the NLMPFA weights. In addition, for the checkerboard mesh there are some cells and faces where some of the coefficients \( \lambda_{K,p} \) in the sub-flux definition (20) are negative. These two arguments might be the reason why the NLMPFA scheme needs 623 Newton iterations on the finest refinement level.

Fig. 1. Hexahedral meshes used for the convergence test cases (modified after [24]).
Fig. 2. Logarithm of $L^2$-error (left) and discrete $H^1$-error (right) for convergence test case one. From top to bottom: checkerboard, random, non-convex, curved mesh.
In the last two test cases, it has been observed that the NLTPFA, NLMPFA, and AvgMPFA that use the conormal decomposition behave similar to well-established linear schemes such as the MPFA-O, MFD, or HFV schemes. Especially for highly heterogeneous tensors, the convergence behavior of the NLTPFA scheme is similar to the behavior of the MPFA-O or HFV scheme. In addition, the Newton converges much better for the NLTPFA scheme than for the NLMPFA scheme.

**Remark 6.** For the above test cases, the classical linear TPFA method does not converge. This is well-known for non-K-orthogonal grids, see for example [6]. Therefore, the TPFA scheme has not been considered so far.

### 3.2 Linearity preservation

Within this section, the linearity preservation property of the schemes is investigated, which is a good indicator for consistency of schemes. This example is based on our recent work [21]. The considered domain and the grid are shown in Figure 4 (right). The domain consists of two sub-domains $\Omega_1$ and $\Omega_2$. The transition from $\Omega_1$ to $\Omega_2$ is located at $x = 0.6$, and the tensors are chosen as

$$\mathbf{A}_1 = \begin{pmatrix} 3 & 1 & 0 \\ 1 & 3 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{A}_2 = \begin{pmatrix} 10 & 3 & 0 \\ 3 & 10 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$  \hspace{1cm} (51)

The exact solutions in the sub-domains are

$$\bar{u}_1 = 14x + y + z, \quad \bar{u}_2 = 4x + y + z + 6.$$  \hspace{1cm} (52)

Figure 4 (left) depicts the exact solution. Please note that the exact solution and the corresponding flux function are globally continuous within the domain. It can also be seen that the grid is non-matching at the transition of the sub-domains. Such non-matching grids often occur in faulted geological environments. The grid in Figure 4 is defined by means of the standard corner-point grid format and has been generated with the *Matlab Reservoir Simulation Toolbox* (MRST) [44]. To read in the grid, the `opm-grid` module from the *Open Porous Media* (OPM) initiative\(^2\) is used, which supports the standard corner-point grid format, see [24, 46] for more information about corner-point grids.

Table 3 lists the discrete error norms, the number of entries in the Jacobian matrix (noe), and the number of Newton iterations needed for the simulation run. It can be seen that all schemes except the TPFA scheme reproduce the exact solution, because the errors are within the range of the nonlinear and linear solver tolerance, whereas the errors of the linear TPFA scheme are approximately five orders of magnitude higher. It is well-known that the errors of the linear TPFA scheme are in $O(1)$ for non-K-orthogonal grids. However, the improved accuracy of the other schemes

---

\(^2\) http://opm-project.org/
Fig. 3. Logarithm of $L^2$-error (left) and discrete $H^1$-error (right) for convergence test case one. From top to bottom: checkerboard, random, non-convex, curved mesh.
comes with the cost of a larger face flux stencil, which is the reason why the corresponding Jacobian matrices are denser than the one of the TPFA scheme. When using Picard’s method instead of Newton’s method, the number of entries is the same for the NLTPFA and TPFA schemes. It can also be seen that the MFD and HFV schemes result in a four times denser matrix than the NLTPFA, NLMPFA, or AvgMPFA schemes. This is due to the fact that the calculation of the coefficient matrix (37) for the MFD and HFV schemes, in general, include all face information, i.e. $S_{K,r} = E_K$. Additionally, flux conservation is weakly enforced using the additional equation (11) that is assembled into the global discretization matrix. The MPFA-O or Box scheme cannot be easily applied to this test case because of the non-matching interface between the sub-domains.

### 3.3 Discrete extremum principles

The following two examples investigate whether the schemes satisfy discrete extremum principles. As already mentioned before, in general, consistent linear schemes do not satisfy extremum principles, which motivates the construction of nonlinear schemes as introduced above. The setups for these test cases are taken from [21].

#### First test case

The first example investigates a test case without a source term. The domain and the grid are shown in Figure 5, with an inner and an outer boundary. The Dirichlet values

| Scheme     | $||u_n - \pi||_L^2$ | $||u_n - \pi||_T$ | noe | NIt |
|------------|---------------------|-------------------|-----|-----|
| NLTPFA     | 1.97e-8             | 8.11e-7           | 184111 | 4   |
| NLMPFA     | 1.99e-8             | 8.31e-7           | 184202 | 7   |
| AvgMPFA    | 1.99e-8             | 8.31e-7           | 184111 | 1   |
| TPFA       | 9.11e-3             | 3.92e-1           | 107600 | 1   |
| MFD        | 1.99e-8             | 7.91e-7           | 747800 | 1   |
| HFV        | 2.00e-8             | 8.17e-7           | 747800 | 1   |
$u = 10^5$ and $u = 0$ are set at the inner and outer boundaries, respectively. Therefore, the solution is expected to be within these bounds. The tensor $\Lambda$ is chosen homogeneous but anisotropic as

$$\Lambda = R\left(\frac{\pi}{6}\right) \left( \begin{array}{cc} 10^0 & 0 \\ 0 & 1 \end{array} \right) R^{-1}\left(\frac{\pi}{6}\right).$$

with $R(x) = \left( \begin{array}{cc} \cos x & -\sin x \\ \sin x & \cos x \end{array} \right).$ (53)

Figure 6 shows the numerical solutions of the different schemes on a three times refined grid. The TPFA scheme produces no under- or overshoots, and therefore satisfies extremum principles. This is well-known since the resulting discretization matrix is an M-matrix. However, the linear TPFA scheme is not consistent for this test case which explains why the anisotropy is not reflected by the TPFA solution. The minimum principle is violated by the MPFA-O, MFD, and HFV schemes, with overshoots of 8% for the MPFA-O scheme and up to 4% for the MFD and HFV schemes. The maximum principle is violated by all consistent linear schemes. The undershoots of the AvgMPFA scheme are above 4%, those of the Box scheme above 2%. The undershoots of the other schemes are below 1%. For this test case, both nonlinear schemes satisfy the discrete extremum principles. The small negative undershoots of the NLMPFA scheme are caused by Newton’s method. These undershoots can be prevented by using other nonlinear solvers such as Picard’s method or enhanced solvers [47]. Furthermore, we also point out that the NLTPFA scheme in general does not satisfy the minimum principle. A similar test case has been considered in [47, 48] with outer Dirichlet boundary conditions above zero. In that case undershoots can also be observed by the NLTPFA scheme.

Second test case

The previous example has demonstrated that the NLTPFA and NLMPFA schemes satisfy the minimum principle, whereas the NLTPFA scheme is positivity-preserving. In the following example, it is demonstrated that the NLMPFA scheme satisfies the minimum principle, in contrast to the NLTPFA scheme. This test case has been introduced in [49]. The boundary conditions are Neumann no-flow on the whole boundary, i.e., $(\nabla u) \cdot n = 0$ on $\partial \Omega$, and $\Omega = [0, 1]^2$ is discretized with a regular Cartesian grid, with 77 $\times$ 77 cells. The tensor $\Lambda$ is chosen as (53) with $\pi = 67.5^\circ$.

We add the constraints $u = 0$ and $u = 1$ at the cells with centers $(\frac{2}{15}, \frac{1}{2})$ and $(\frac{1}{2}, \frac{1}{15})$, respectively. Therefore, the solution is bounded, such that $0 \leq u \leq 1$.

Figure 7 depicts the numerical solutions of the different schemes as surface plots. The TPFA obviously again satisfies the minimum and maximum principle due to the M-matrix property but again the anisotropy is not correctly reproduced. The only consistent scheme that fulfills the maximum principle is the NLMPFA scheme, all the other schemes produce overshoots. The Box and AvgMPFA schemes produce the highest over- and undershoots. Those of the Box scheme are above 26%. The overshoots of the NLTTPFA scheme are higher than those of the MPFA-O, MFD, and HFV schemes. But the NLTTPFA scheme produces no undershoots since it is positivity-preserving. Furthermore, the undershoots of the MPFA-O, MFD, and HFV are below 3%.

The above test cases exhibit how nonlinear schemes are capable to reproduce physical solutions, whereas linear schemes can produce negative values. When solving highly complex partial differential equations, where secondary variables non-linearly depend on primary variables, such negative values can strongly influence the efficiency of the scheme, in terms of linear and nonlinear solver convergence. It should be mentioned again that the NLTTPFA is only positivity-preserving, whereas the NLMPFA satisfies discrete extremum principles.

3.4 Benchmark: Northeast German Basin

The following example is a synthetic model inspired by the 3D Northeast German Basin model presented in [50]. This test case has recently been published in [21], where the results of the AvgMPFA, NLTTPFA, NLMPFA, TPFA, and Box schemes have been presented. Here, we additionally present the results of the MPFA-O, MFD, and HFV schemes. However, the model setup is analogous to [21]. The data and the approximate geometry of the basin are provided by IFPEN. The different facies of the Northeast German Basin are depicted in Figure 8, where the domain is reflected such that $-z$ is oriented in depth direction.

Salt diapirs within this model create highly conductive regions, as shown in Figure 9, leading to thermal anomalies. A robust discretization with respect to the grid is required for this type of structure, in order to evaluate the temperature field and to perform thermohaline simulations.

Here, the stationary heat equation is solved, where $\Lambda$ corresponds to the thermal conductivity [W/(m·K)] and $u$ to the temperature $T$ [K]. The thermal conductivity is computed with the following law [51]

$$\Lambda = \Lambda_w^\phi \Lambda_s^{1-\phi} \mathbf{I},$$

where $\mathbf{I}$ is the identity matrix, $\Lambda_w$ and $\Lambda_s$ denote the water and rock conductivities, and $\phi$ the porosity. The conductivity of water is set to $\Lambda_w = 0.6$. The thermal properties
of the different facies are listed in Table 4. The porosity distribution and the thermal conductivities are shown in Figure 9. At the top and bottom boundaries, Dirichlet conditions are set to 281.15 K and 423.15 K, respectively, whereas Neumann no-flow conditions are set elsewhere. The grid consists of 864,435 cells.

Figure 10 shows the numerical solutions of the MPFA-O and TPFA schemes. Additionally, the absolute difference...
between the MPFA-O and the NLTPFA, TPFA, HFV, and Box schemes is depicted in this figure. It is observed that the TPFA scheme differs the most from the other schemes. The largest differences occur at the salt domes, where it seems that the TPFA scheme overestimates the temperature values. The NLTPFA and HFV schemes are in good agreement with the MPFA-O results. The largest difference between the NLTPFA and MPFA-O scheme is below 7.88 K, and for the HFV scheme below 2.29 K. This can also be seen in Table 5, which lists the relative discrete errors $\frac{||u_1 - u_2||_L^2}{V} \Omega$ between the schemes, with the total domain volume $V = 1.75 \times 10^{14} \text{ m}^3$ and the reference temperature $T_{\text{ref}} = 352.15 \text{ K}$. It can be observed that the AvgMPFA, NLMPFA, and NLTPFA schemes produce similar solutions. The same applies to the HFV and MFD schemes. Furthermore, the TPFA and the Box schemes differ the most from the other schemes, which is in agreement with the results shown in Figure 10. Again, the number of entries of the NLTPFA, NLMPFA, and AvgMPFA is approximately twice the number of the TPFA scheme. Moreover, the most dense matrices are those of the HFV and MFD schemes.

### Table 4. Thermal properties of facies of the Northeast German Basin.

| Facies                                      | $\Lambda$ [W/(m·K)] |
|---------------------------------------------|----------------------|
| Marl                                        | 2.76                 |
| Silt                                        | 2.5                  |
| Chalk                                       | 3.57                 |
| Sandstone                                   | 6.32                 |
| Sandstone and shale (mixture 50%-50%)       | 3.67                 |
| Limestone                                   | 3.57                 |
| Salt                                        | 6.1                  |
| Conglomerate                                | 3.27                 |

---

**Fig. 8.** Different facies of the Northeast German Basin, where the domain is scaled by a factor of five in $z$-direction and the different layers are shifted horizontally for better visibility of the features. The domain lengths in coordinate directions are approximately 169 km (in the $x$-direction), 165 km (in the $y$-direction), and 10.87 km (in the $z$-direction). The domain has been reflected at the $z$-plane such that $-z$ corresponds to the depth axis. At the top and bottom boundaries, Dirichlet conditions are set to 281.15 K and 423.15 K, respectively, whereas Neumann no-flow conditions are set elsewhere (modified after [21]).

**Fig. 9.** Thermal conductivity, calculated with the law (54) (choosing $\Lambda_0 = 0.6$) and porosity distribution of the Northeast German Basin, whereby the domain is scaled by a factor of five in $z$-direction. The salt domes correspond to the highly conductive regions (modified after [21]).

**Fig. 10.** Solution of MPFA-O and TPFA schemes (first row). Absolute difference between the MPFA-O and the NLTPFA, TPFA, HFV, and Box schemes (second and third row). The results are shown for a part of the domain.
Table 5. Discrete relative errors $\frac{|u - u_h|^2}{	au_n \sqrt{\|u\|}}$ between the different schemes. Number of entries (noe) and Newton iterations (NIt).

| Scheme    | NLTPFA | NLMPFA | AvgMPFA | TPFA   | Box    | MPFA-O | MFD   | HFV   |
|-----------|--------|--------|---------|--------|--------|--------|-------|-------|
| NLTPFA    | 0      | 9.45e-4 | 4.16e-4 | 1.31e-2 | 6.10e-3 | 1.84e-3 | 2.51e-3 | 2.28e-3 |
| NLMPFA    | 9.45e-4 | 6.61e-4 | 1.30e-2 | 5.90e-3 | 1.17e-3 | 2.12e-3 | 1.81e-3 |       |
| AvgMPFA   | 4.16e-4 | 6.61e-4 | 1.32e-2 | 6.05e-3 | 1.50e-3 | 2.30e-3 | 2.01e-3 |       |
| TPFA      | 1.31e-2 | 1.30e-2 | 1.32e-2 | 0.16e-2 | 1.33e-2 | 1.42e-2 | 1.40e-2 |       |
| Box       | 6.10e-3 | 5.90e-3 | 6.05e-3 | 1.60e-2 | 5.65e-3 | 4.23e-3 | 4.74e-3 |       |
| MPFA-O    | 1.84e-3 | 1.17e-3 | 1.50e-3 | 1.33e-2 | 5.65e-3 | 0      | 1.59e-3 | 1.15e-3 |
| MFD       | 2.51e-3 | 2.12e-3 | 2.30e-3 | 1.42e-2 | 4.23e-3 | 1.59e-3 | 0      | 6.68e-4 |
| HFV       | 2.28e-3 | 1.81e-3 | 2.01e-3 | 1.40e-2 | 4.74e-3 | 1.15e-3 | 6.68e-4 | 0      |
| noe       | 11886463 | 11884419 | 11886463 | 5974567 | 23684992 | 22655815 | 39802249 | 39802249 |
| NIt       | 2      | 7      | 1       | 1      | 1      | 1      | 1      | 1      |

4 Conclusion

Within this article, we have presented an abstract discretization framework for elliptic equations. Furthermore, it has been demonstrated that various finite-volume schemes fit into this framework, with the difference that hybrid schemes introduce additional face unknowns, whereas cell-centered schemes eliminate these face unknowns by using trace reconstruction operators. Properties like consistency, coercivity, extremum principles, and sparsity have been defined in Section 2.2 for schemes that belong to this discretization framework. Consistency and coercivity are particularly essential to prove the convergence of such schemes. Hybrid schemes are coercive by construction, whereas cell-centered schemes are generally not coercive.

In Section 2.3, besides the linear AvgMPFA scheme, a monotone Nonlinear Two-Point Flux Approximation (NLTPFA) and an extremum-principles-preserving Multi-Point Flux Approximation (NLMPFA) have been presented. In Section 2.4, we have briefly introduced the well-known mimetic [13, 52] and hybrid finite-volume [8] schemes. Note that all schemes have been implemented into the open-source simulator DuMu$^5$ [40], allowing for a comparison of the different methods within the same software framework. The schemes have been analyzed numerically in Section 3. For the two test cases considered, involving a mildly and highly anisotropic tensor, it has been demonstrated that all schemes, except for the TPFA, show a similar convergence behavior and that all schemes seem to be coercive. The consistency of the NLTPFA, NLMPFA, AvgMPFA, MF, and HFV schemes, for piecewise linear solutions, has been investigated in Section 3.2 for a non-matching grid. In addition, we have shown in Section 3.3 that linear schemes are in general neither positivity-preserving nor satisfy discrete minimum or maximum principles, in contrast to nonlinear schemes. In the last example the Northeast German Basin has been considered, showing that the NLTPFA and NLMPFA schemes result in similar solutions than other well-established consistent schemes such as the MPFA-O, MFD, or HFV schemes.

Acknowledgments. As previously mentioned, some of the presented test cases are based on our recent publication [21] that has been produced in collaboration with Léo Agélès and Guillaume Enchéry from IFPEN. Therefore, the authors would like to thank both of them for their support and the many fruitful conversations. Furthermore, we thank the German Research Foundation (DFG) for funding this work within SFB 1313.

References

1. Di Pietro D.A. (2012) Cell centered galerkin methods for diffusive problems, Math. Model. Numer. Anal. 46, 1, 111–144.
2. Aavatsmark I., Barkve T., Bøe Ø., Mannseth T. (1996) Discretization on non-orthogonal, quadrilateral grids for inhomogeneous, anisotropic media, J. Comput. Phys. 127, 1, 2–14.
3. Agélès L., Di Pietro D., Droniou J. (2010) The G method for heterogeneous anisotropic diffusion on general meshes, M2AN Math. Model. Numer. Anal. 44, 4, 597–625.
4. Agélès L., Guichard C., Masson R. (2010) Convergence of finite volume MPFA O type schemes for heterogeneous anisotropic diffusion problems on general meshes, Int. J. Finite Vol. 7, 2, 1–33.
5. Agélès L., Masson R. (2008) Convergence of the finite volume MPFA O scheme for heterogeneous anisotropic diffusion problems on general meshes, C. R. Math. 346, 17, 1007–1012.
6. Edwards M., Rogers C. (1998) Finite volume discretization with imposed flux continuity for the general tensor pressure equation, Comput. Geosci. 2, 4, 259–290.
7. Wolff M., Cao Y., Flemisch B., Helmig R., Wohlmuth B. (2013) Multi-point flux approximation L-method in 3d: numerical convergence and application to two-phase flow through porous media, Radon Ser. Comput. Appl. Math., De Gruyter 12, 39–80.
8. Eymard R., Gallouët T., Herbin R. (2010) Discretization of heterogeneous and anisotropic diffusion problems on general non-conforming meshes, SUSHI: a scheme using stabilization and hybrid interfaces, IAM J. Num. Anal. 30, 4, 1000–1043.
9. Eymard R., Herbin R., Guichard C., Masson R. (2012) Vertex-centred discretization of multiphase compositional Darcy flows on general meshes, Comput. Geosci. 16, 4, 987–1005.
48 Kapryn I., Nikitin K., Terekhov K., Vassilevski Y. (2014) Nonlinear monotone FV schemes for radionuclide geomigration and multiphase flow models, in: Finite volumes for complex applications VII-Elliptic, parabolic and hyperbolic problems, Springer, pp. 655–663.

49 Aavatsmark I., Eigestad G., Mallison B., Nordbotten J. (2008) A compact multipoint flux approximation method with improved robustness, Numer. Methods Partial Differ. Equ. 24, 5, 1329–1360.

50 Scheck M., Bayer U. (1999) Evolution of the Northeast German Basin – inferences from a 3D structural model and subsidence analysis, Tectonophysics 3, 3, 145–169.

51 Woodside W., Messmer J. (1961) Thermal conductivity of porous media. I. Unconsolidated sands, J. Appl. Phys. 32, 9, 1688–1699.

52 Lipnikov K., Manzini G., Svyatskiy D. (2011) Analysis of the monotonicity conditions in the mimetic finite difference method for elliptic problems, J. Comput. Phys. 230, 7, 2620–2642.