Nonlinear multigrid solvers exploiting AMGe coarse spaces with approximation properties

Christensen, Max la Cour; Vassilevski, Panayot S.; Villa, Umberto

Published in:
Journal of Computational and Applied Mathematics

Link to article, DOI:
10.1016/j.cam.2017.10.029

Publication date:
2018

Document Version
Peer reviewed version

Citation (APA):
Accepted Manuscript

Nonlinear multigrid solvers exploiting AMGe coarse spaces with approximation properties

Max la Cour Christensen, Panayot S. Vassilevski, Umberto Villa

PII: S0377-0427(17)30530-7
DOI: https://doi.org/10.1016/j.cam.2017.10.029
Reference: CAM 11359

To appear in: Journal of Computational and Applied Mathematics

Received date: 21 April 2017
Revised date: 20 October 2017


This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
Nonlinear Multigrid solvers exploiting AMGe Coarse Spaces with Approximation Properties

Max la Cour Christensen\textsuperscript{a}, Panayot S. Vassilevski\textsuperscript{b}, Umberto Villa\textsuperscript{c}

\textsuperscript{a}Technical University of Denmark, Center for Oil and Gas (max@maxlacour.com).
\textsuperscript{b}Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, California, USA, (vassilevski1@llnl.gov), and Fariborz Maseeh Department of Mathematics and Statistics, Portland State University, Portland, OR 97201, USA, (panayot@pdx.edu).
\textsuperscript{c}Institute for Computational Engineering and Sciences (ICES), The University of Texas at Austin, (uvilla@ices.utexas.edu).

Abstract
This paper introduces a nonlinear multigrid solver for mixed finite element discretizations based on the Full Approximation Scheme (FAS) and element-based Algebraic Multigrid (AMGe). The AMGe coarse spaces with approximation properties used in this work enable us to overcome the difficulties in evaluating the nonlinear coarse operators and the degradation in convergence rates that characterized previous attempts to extend FAS to algebraic multilevel hierarchies on general unstructured grids. Specifically, the AMGe technique employed in this paper allows to derive stable and accurate coarse discretizations on general unstructured grids for a large class of nonlinear partial differential equations, including saddle point problems. The approximation properties of the coarse spaces ensure that our FAS approach for general unstructured meshes leads to optimal mesh-independent convergence rates similar to those achieved by geometric FAS on a nested hierarchy of refined meshes. In the numerical results, Newton’s method and Picard iterations with state-of-the-art inner linear solvers are compared to our FAS algorithm for the solution of a nonlinear saddle point problem arising from porous media flow applications. Our approach outperforms — both in terms of number of iterations and computational time — traditional methods in all the experiments.

*This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344 supported in part by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Applied Mathematics program.

Keywords: Nonlinear multigrid, Full Approximation Schemes (FAS), Element-based Algebraic Multigrid (AMGe), Nonlinear Saddle Point Problems, Multilevel Numerical Upscaling, Mixed Finite Element Method (MFEM), Flow in Porous Media
1. Introduction

Efficient solvers for nonlinear partial differential equations (PDEs) are an essential component of modern large-scale simulations. Newton-Krylov methods [1] represent the state-of-the-art for this kind of problems. They are generally robust and allow for fast (quadratic) convergence provided that a good initial guess is given. In addition, they leverage on existing scalable linear solvers and preconditioners, including black-box algebraic multigrid (see, e.g., [2, 3]) and domain decomposition methods (see, e.g., [4, 5]), to efficiently solve the large sparse linear systems arising from global linearization of the nonlinear operator (i.e. the Jacobian matrix). However, Newton-Krylov methods have a number of drawbacks as stand-alone solvers (see, e.g., [6] where composable nonlinear solver are discussed), both in terms of performance and of robustness with respect to the initial guess. In terms of performance, the repeated construction and solution of the globally defined Jacobian operator may lead to memory bandwidth and communication bottlenecks in modern multi-core multi-processor computer architectures. In terms of robustness, Newton solution strategies may fail when the initial guess is outside the basin of attraction of the method, and globalization techniques (such as line search, trust region, pseudo-time continuation) may be necessary to ensure a convergent method.

The need of new algorithms that can better exploit the complex architecture of modern super computers and the lack of global convergence of these methods motivate our interest in nonlinear multilevel solvers for PDEs. Specifically, we are interested in Full Approximation Scheme (FAS) type algorithms [7, 8, 9, 10]. The most widespread use of FAS is in geometric multigrid on structured grids due to difficulties associated with defining a coarse nonlinear operator on unstructured meshes. FAS offers potential benefits with respect to traditional Newton-Krylov methods, such as larger basin of attraction, faster initial convergence, more data locality and lower memory footprint.

Several papers in the last two decades have addressed the applicability of FAS to unstructured grids; however, to the best of our knowledge, our work is the first to describe a nonlinear multilevel solver with algebraic coarse spaces that offers optimal convergence properties — both in terms of iterations and computational time — on general unstructured meshes.

In [11, 12], FAS based on agglomeration multigrid was compared to Newton-Multigrid methods for the solution of problems of practical relevance in computational fluid dynamics. In these papers, coarse grid control-volumes were formed by merging together finer grid control-volumes. Based on this agglomeration of control-volumes, the associated interpolators between grids were defined as simple injection operators (piecewise constant on agglomerated control-volumes). In a multilevel context, piecewise constant interpolation between grids was insufficient and resulted in loss of accuracy and therefore loss of performance in the overall multigrid scheme, [13]. An improvement was suggested in [13], where an implicit prolongation operator was introduced to achieve mesh independent convergence rates. However, the gain in convergence rate was canceled out by a drastic increase in computational cost, undermining the overall
efficiency of the method.

In [14, 15], FAS was combined with element-based Algebraic Multigrid (AMGe) to obtain a nonlinear solver for lowest order nodal finite elements. Their method was based on the AMGe technique introduced in [16, 17]. This resulted in coarse spaces where only one degree of freedom could be used for each agglomerate. Consequently, it was difficult to maintain accuracy on very coarse agglomerated meshes, resulting in a degradation of the FAS solver performance. In fact, their numerical results showed mesh-independent convergence only for an elliptic 2D model problem, while convergence rates were sub-optimal in 3D.

This paper can be seen as an extension and improvement upon the work in [14, 15]. Our main contributions are: i) the underlying AMGe method providing the multigrid components — namely the restriction, prolongation and nonlinear coarse operators — that allows for optimal scalability of the nonlinear multilevel solver; ii) the ability to construct nonlinear multilevel solver for a more general class of nonlinear PDEs problems, including saddle point problems.

Specifically, the version of AMGe used in this paper, [18, 19], constructs operator-dependent coarse spaces for the whole de Rham complex (i.e. the sequence of $H^1$-conforming, $H$-curl conforming, $H$-div conforming and $L^2$-conforming spaces). This gives the foundation to cover a broad range of applications such as elliptic PDEs, Maxwell equations, Darcy flow equations, etc. Such AMGe technique with guaranteed approximation properties on coarse agglomerated meshes provides the coarse spaces used for the restriction, prolongation and nonlinear operators. These coarse spaces have desirable properties analogous to the original (fine grid) finite element spaces both in term of approximation properties and stability: Nédélec, Raviart-Thomas and piecewise discontinuous polynomials. This is ensured by introducing additional degrees of freedom associated with non-planar interfaces/edges between coarse elements/faces (agglomerates of finer level elements/faces). The necessary number of degrees of freedom on coarse faces or coarse edges is automatically found via singular value decomposition. While these coarse spaces were originally intended for use as an upscaling tool; in this work, we demonstrate that the same coarse spaces can be reused to construct scalable multilevel solvers. We also remark that the construction of stable algebraic multigrid hierarchies for mixed discretization of saddle point systems is indeed a very challenging problem on its own and it is an active area of research also for linear solvers, see e.g. [20].

It is worth mentioning that the construction of robust solvers for nonlinear PDE problems has been an area of active research also in the domain decomposition (DD) community. For example, the additive Schwarz preconditioned inexact Newton algorithm (ASPIN), originally proposed in [21], has been extended to a two-level method in [22, 23] thus improving scalability with respect to both the number of processors and the size of the fine mesh. While, in the DD area, the main goal is to design DD solvers that are robust with respect to coefficient variations and stiff nonlinear behavior for broad classes of PDEs, our goal is to construct coarse spaces with guaranteed approximation properties so that they can be also used as discretization (upscaling) tools for nonlinear PDEs. These goals are motivated and explained in more details in the survey
[24], which deals with the use of appropriate AMG-based coarse spaces as accurate discretization spaces (i.e., as an upscaling tool) in the context of linear problems.

The FAS-AMGe method implemented in this paper is tested on a nonlinear saddle point problem with applications in porous media flow. It is compared to exact and inexact Newton’s method and Picard iterations. The comparison is done in a fair way by letting the FAS, Newton’s method and Picard iterations utilize the same underlying components, namely the multilevel divergence free solver [25], for the solution of the mixed discretization of the Darcy problem.

The rest of the paper is structured as follows. In Section 2 we give a brief outline of the AMGe method we use. Section 3 summarizes the FAS algorithm in general terms. The model problem of our main interest is introduced in Section 4. The main part of this paper, consisting of a large set of numerical tests, is given in Section 5. At the end, in Section 6 we draw some conclusions and suggest future research directions.

2. Element-based Algebraic Multigrid (AMGe)

AMGe is a framework of multilevel methods for the solution of systems stemming from finite element discretizations. In contrast to AMG, where only system coefficients are used, AMGe also employs grid topology information and local finite element matrices. The specific version of AMGe used in this paper was introduced in [26, 18, 19] to build stable discretizations of the whole de Rham complex (i.e., the sequence of $H^1$-conforming, $H$ (curl)-conforming, $H$ (div)-conforming and $L^2$-conforming spaces) on agglomerated coarse meshes. In particular, the method facilitates the construction of operator-dependent coarse spaces that can be shown to guarantee good approximation properties on coarse levels for general unstructured meshes. Thanks to the guaranteed approximation properties, this AMGe technique can be used as a discretization tool (upscaling) on coarse (agglomerated) meshes and allows for the generation of accurate coarse spaces for the FAS hierarchy.

In a setup phase, a hierarchy of agglomerated meshes is formed. Each agglomerate is formed by grouping together finer-grid elements (or agglomerates if already on a coarse level). For unstructured meshes, the agglomeration can be accomplished using graph partitioners, such as METIS [27]. Once the hierarchy of agglomerated meshes is generated, coarse spaces are computed by restricting certain basis functions and by solving local saddle point problems for each agglomerated entity. In addition, this version of AMGe allows assembling coarse grid residuals and Jacobians directly on coarse agglomerated meshes without visiting the fine grid, as we demonstrate for the model problem in Section 4.4. For details on the assembly procedure see [28], where a time-dependent two-phase porous media flow is solved with optimal complexity on coarse (upscaled) levels. The software developed for this paper uses the Element-Agglomeration Algebraic Multigrid and Upscaling Library (ParElag, [29]) developed at Lawrence Livermore National Laboratory. ParElag is based...
Nonlinear MG with Accurate AMGe Coarse Spaces

on the MFEM library, [30], for the finite element discretization and supports several solvers from the hypre library, [31].

In what follows we describe the properties of the de Rham sequence in the continuous spaces and we discuss the stability of the discrete sequences based on the commutativity between finite element canonical interpolators and differential operators. We conclude the section by explaining the requirements for the construction of coarse stable sequences on general unstructured agglomerated meshes. Note that to solve the model problem in Section 4, one only needs to build the hierarchy of $H(\text{div})$ and $L^2$-conforming spaces; in addition, to apply the preconditioner in Section 4.5, one also needs to construct the hierarchy of $H(\text{curl})$-conforming spaces. In the appendix we provide details on the construction of the coarse $L^2 - H(\text{div})$-conforming spaces. The construction of the coarse $H(\text{curl})$-conforming space is similar in spirit, but more involved (c.f. [26, 19]).

2.1. Exact de Rham sequences

We now introduce some notation used throughout the paper. Let $\Omega$ be a bounded connected domain in $\mathbb{R}^d$ with a regular (Lipschitz continuous) boundary $\partial \Omega$, which has a well-defined unit outward normal vector $\mathbf{n} \in \mathbb{R}^d$. For the cases considered in this paper, $d = 3$. For the vectorial functions $\mathbf{u}, \mathbf{v} \in \mathbf{L}^2(\Omega) = [L^2(\Omega)]^d$ and scalar functions $p, w \in L^2(\Omega)$, we define the inner products $(\mathbf{u}, \mathbf{v}) = \int_\Omega \mathbf{u} \cdot \mathbf{v} \, d\Omega$ and $(p, w) = \int_\Omega p \, w \, d\Omega$. Finally, we introduce the functional spaces $S, Q, R, W$ defined as

\begin{align*}
S & := H^1(\Omega) = \{ s \in L^2(\Omega) \mid \nabla s \in \mathbf{L}^2(\Omega) \} \\
Q & := H(\text{curl;} \Omega) = \{ \sigma \in \mathbf{L}^2(\Omega) \mid \text{curl} \, \sigma \in \mathbf{L}^2(\Omega) \} \\
R & := H(\text{div;} \Omega) = \{ \mathbf{u} \in \mathbf{L}^2(\Omega) \mid \text{div} \, \mathbf{u} \in L^2(\Omega) \} \\
W & := L^2(\Omega).
\end{align*}

The above spaces form a de Rham sequence

\begin{align}
S \xrightarrow{\nabla} Q \xrightarrow{\text{curl}} R \xrightarrow{\text{div}} W,
\end{align}

in the sense that $\nabla S \subseteq \ker(\text{curl})$ and $\text{curl} \, Q \subseteq \ker(\text{div})$, where we have set $\ker(\text{curl}) := \{ \sigma \in Q \mid \text{curl} \, \sigma = 0 \}$ and $\ker(\text{div}) := \{ \mathbf{u} \in R \mid \text{div} \, \mathbf{u} = 0 \}$. In addition, a de Rham sequence is said to be exact if we have that $(\nabla S)^{\perp} \cap \ker(\text{curl}) = \{ 0 \}$ and $(\text{curl} \, Q)^{\perp} \cap \ker(\text{div}) = \{ 0 \}$, where $(\cdot)^{\perp}$ denotes the orthogonal complement of the subspace $(\cdot)$. Under certain topological assumptions on the domain $\Omega$, namely that $\Omega$ is simply connected with connected boundary, the spaces $S, Q, R, W$ form an exact sequence, see e.g. [32].

A stable (i.e. inf-sup compatible) discretization of the de Rham sequence can be obtained as follows. Let $T_h$ denote a conforming triangulation of the domain $\Omega$, and define appropriate finite dimensional spaces $S_h \subset S$, $Q_h \subset Q$, $R_h \subset R$, $W_h \subset W$ associated with the triangulation $T_h$. By appropriate, we mean that the discrete spaces should also preserve the de Rham complex structure of the continuous case (see [33, 34] for more details), i.e. they should
form the commuting diagram

\[
\begin{array}{cccccc}
S & \nabla & Q & \text{curl} & R & \text{div} & W \\
S_h & \nabla & Q_h & \text{curl} & R_h & \text{div} & W_h.
\end{array}
\]

A standard choice for numerical discretization of the de Rham sequence is the following. For a given integer \( r \geq 0 \), we let \( S_h \) be the space of continuous piecewise polynomials of degree at most \( r + 1 \), \( Q_h \) the \( (r + 1) \)-th order Nédélec finite elements [35], \( R_h \) the \( r \)-th order Raviart-Thomas finite elements [36, 35], and \( W_h \) the discontinuous piecewise polynomial finite elements of degree at most \( r \). In fact, this choice of finite elements guarantees that the commutativity property

\[
\nabla \Pi^Q_h = \Pi^Q \nabla, \quad \text{curl} \Pi^Q_h = \Pi^R \text{curl}, \quad \text{and div} \Pi^R_h = \Pi^W \text{div}
\]

holds [37, 38]. Here, \( \Pi^V_h : V \mapsto V_h \) denotes the canonical interpolation operator from the continuous space \( V \) into the finite element counterpart \( V_h \), \( V \in \{ S, Q, R, W \} \).

In what follows, we denote with \( G_h, C_h, D_h \) the matrix representation of the discrete gradient, curl, and divergence operators, i.e., the matrix representation of the mappings \( \nabla : S_h \mapsto Q_h \) such that \( \sigma_h = \nabla s_h \in Q_h \) for all \( s_h \in S_h \), \( \text{curl} : Q_h \mapsto R_h \) such that \( u_h = \text{curl} \sigma_h \in R_h \) for all \( \sigma_h \in Q_h \), and similarly, \( \text{div} : R_h \mapsto W_h \) such that \( p_h = \text{div} u_h \in W_h \) for all \( u_h \in R_h \).

### 2.2. Exact de Rham sequences on agglomerated meshes

Given a fine mesh \( T_h \), we define an agglomerated mesh \( T_H \) whose elements \( T \) consist of a non-overlapping partition of fine grid elements \( \tau \in T_h \). The objective is then to generate a coarse sequence \( S_H, Q_H, R_H, W_H \) associated with the agglomerated mesh \( T_H \) such that: (1) the sequence is exact (provided that \( \Omega \) meets the necessary topological requirements); (2) the commutativity property between the two sequences is preserved; (3) the coarse spaces are respectively \( H^1 \), \( H(\text{curl}) \), \( H(\text{div}) \), \( L^2 \)-conforming; (4) the approximation order of the original spaces is preserved. Requirements (1)-(3) guarantee the stability of the coarse sequences. Such stable coarse sequences can be built following the approach presented in [26]. However the approximation properties of the resulting coarse spaces are not preserved for general unstructured agglomerated meshes \( T_H \). This was remedied in [18] for the coarse \( H(\text{div}) \)-conforming space and in [19] for the complete sequence, by adding additional coarse degrees of freedom corresponding to non-planar coarse faces and non-straight coarse edges.

In summary, the AMGe technique described in [26, 18, 19] allows us to
construct the commutative diagram

\[
\begin{array}{ccccccc}
\Pi^S & \xrightarrow{P^S} & S_h & \xrightarrow{G_h} & Q_h & \xrightarrow{C_h} & R_h & \xrightarrow{D_h} & W_h & \Pi^W \\
\Pi^H & \xrightarrow{P^H} & S_H & \xrightarrow{G_H} & Q_H & \xrightarrow{C_H} & R_H & \xrightarrow{D_H} & W_H & \Pi^W
\end{array}
\] (4)

where, for each \( V \in \{ S, Q, R, W \} \), \( P^V : V_H \mapsto V_h \) represents the prolongation operator from the coarse to the fine space, and \( \Pi^V : V_h \mapsto V_H \) represents the commuting projection operator from the fine to the coarse space. Also, \( G_H, C_H, D_H \) stem from the matrix representations of the \( \nabla \), curl, div differential operators associated with the agglomerated mesh \( T_H \). All these operators and matrices are locally (agglomerated element by agglomerated element) constructed by the AMGe technique and satisfy, by construction, the following requirements:

\[
\begin{align*}
\Pi^V P^V &= I_H, \quad \forall V \in \{ S, Q, R, W \}; \\
C_H G_H &= 0, \quad D_H C_H = 0; \\
G_H &= \Pi^Q G_H P^S, \quad C_H = \Pi^R C_H P^Q, \quad D_H = \Pi^W D_H P^R;
\end{align*}
\] (5)

where \( I_H \) represents the identity operator associated to the coarse spaces.

Finally, we can recursively apply this coarsening procedure to obtain the hierarchy of coarse de Rham sequences used, in Section 4.3, to build the FAS solver and, in Section 4.5, to build the multilevel preconditioner.

3. Full Approximation Scheme (FAS)

FAS, [7, 8, 9], can be considered as a generalization of multigrid methods to nonlinear problems. We first introduce the algorithm in a two-grid setting, and then we present its multilevel generalization in Algorithm 1.

Consider the nonlinear discrete problem:

\[
A_h(x_h) = f_h,
\] (6)

where \( A_h : V_h \mapsto V_h^* \) is a nonlinear operator. Here, \( V_h \) represents the finite dimensional space where the discrete solution \( x_h \) is sought, and the adjoint space \( V_h^* \) represents the finite dimensional space where discretizations of linear forms are sought\(^1\). The subscript \( h \) indicates that all quantities are discretized on the fine grid. Introducing the approximate solution \( v_h \in V_h \), the residual equation is given by

\[
A_h(x_h) - A_h(v_h) = d_h,
\] (7)

\(^1\)Note that for a Galerkin method \( V_h \) and \( V_h^* \) are both isomorphic to \( \mathbb{R}^n \), where \( n \) denotes the number of degrees of freedom. However, we prefer not to identify \( V_h \) with its adjoint \( V_h^* \) to better motivate the use of different restriction operators for discrete functions in \( V_h \) and discretized forms in \( V_h^* \).
where \( d_h \in \mathcal{V}_h \) is the residual \( f_h - A_h(v_h) \). Introducing the subscript \( H \) to refer to quantities defined on the coarse mesh, the coarse residual equation can be written as

\[
A_H(x_H) - A_H(v_H) = d_H \iff A_H(v_H + e_H) - A_H(v_H) = d_H, \tag{8}
\]

where \( e_H \in \mathcal{V}_H \) is the error \( x_H - v_H \).

We use the projection operator \( \Pi : \mathcal{V}_h \rightarrow \mathcal{V}_H \) to restrict the approximate solution \( v_h \in \mathcal{V}_h \) to the coarse grid, and we use the adjoint of the prolongation operator \( P : \mathcal{V}_H \rightarrow \mathcal{V}_h \) to restrict the residual \( d_h \in \mathcal{V}_h^* \) to the coarse grid. The operators \( P \) and \( \Pi \) are constructed by our AMGe algorithm such that \( \Pi P = I_H \), c.f. Equation (5a). Then the coarse grid problem reads

\[
A_H \left( \Pi v_h + e_H \right) = A_H(\Pi v_h) + P^T d_h, \tag{9}
\]

where \( P^T \) is the matrix representation of \( P^* : \mathcal{V}_h^* \rightarrow \mathcal{V}_H^* \).

Finally, the coarse grid correction \( e_H = x_H - \Pi v_h \) is prolonged to the fine grid level by using the prolongation operator \( P \) and the solution \( v_h \) is updated accordingly.

Algorithm 1 contains a pseudo code for the multilevel implementation of a FAS V-cycle.

\begin{algorithm}
\caption{Pseudo code for FAS V-cycle implementation.}
\begin{algorithmic}[1]
\Function{FAS\_Vcycle}{l}
\If{\( l = \text{nLevels}-1 \) (coarsest grid)}
\State Approximately solve \( A_l(x_l) = f_l \)
\Else
\State Nonlinear smoothing of \( A_l(x_l) = f_l \)
\State Compute nonlinear residual: \( d_l = f_l - A_l(x_l) \)
\State Restrict nonlinear residual: \( d_{l+1} = P^T d_l \)
\State Restrict solution: \( x_{l+1} = \Pi x_l \)
\State Store approximate solution: \( v_{l+1} = x_{l+1} \)
\State Compute right hand side for residual equation: \( f_{l+1} = d_{l+1} + A_{l+1}(x_{l+1}) \)
\State Apply \text{FAS\_Vcycle}(l+1) to compute updated \( x_{l+1} \)
\State Compute correction: \( e_{l+1} = x_{l+1} - v_{l+1} \)
\State Prolongate correction: \( e_l = P e_{l+1} \)
\State Correct approximation: \( x_l = x_l + e_l \)
\State Nonlinear smoothing of \( A_l(x_l) = f_l \)
\EndIf
\EndFunction
\end{algorithmic}
\end{algorithm}
4. Model problem

The saddle point problem of focus in this paper is chosen for the simplicity of its formulation, while still presenting itself as a numerically challenging nonlinear problem with applications to porous media flow. The problem stems from Darcy’s law and reads

\[
\begin{aligned}
\begin{cases}
    k^{-1}(p)u + \nabla p = f & \text{in } \Omega \\
    \text{div } u = q & \text{in } \Omega,
\end{cases}
\end{aligned}
\]  

(10)

where \( k \) is the conductivity field (or permeability field as it is commonly called in petroleum engineering), \( p \) is the pressure and \( u \) is the velocity. The problem can be used to model the steady-state single-phase primary depletion of an oil reservoir, where the permeability \( k \) depends — in a differentiable manner — on the pressure \( p \).

For simplicity, we assume Dirichlet boundary conditions \( p = 0 \) on \( \partial \Omega \), but non-homogeneous and Neumann boundary conditions (i.e., to prescribe the flux \( u \cdot n \) at the boundary) can be handled in a similar way.

4.1. Weak formulation

To derive the weak formulation of the mixed system in Equation (10), we multiply Equation (10) with test functions \( v \in \mathcal{R} \) and \( w \in \mathcal{W} \) and integrate over the domain \( \Omega \). After integrating by parts of the non-conforming terms and applying the homogeneous Dirichlet boundary condition \( p = 0 \) on \( \partial \Omega \), we obtain the following nonlinear variational problem

**Problem 1** Find \((u, p) \in \mathcal{R} \times \mathcal{W}\) such that

\[
\begin{aligned}
\begin{cases}
    k(p)^{-1}u, v - \left(p, \text{div } v\right) = (f, v), & \forall v \in \mathcal{R} \\
    \left(\text{div } u, w\right) = -(q, w), & \forall w \in \mathcal{W},
\end{cases}
\end{aligned}
\]  

(12)

To solve the non-linear Problem 1 we consider both the Newton’s and Quasi-Newton’s (Picard) methods. In a compact notation, the Newton’s/Picard’s step reads

Solve:\quad a(\delta u, \delta p; v, w) = -r(u_{\text{old}}, p_{\text{old}}; v, w), \quad \forall (v, w) \in \mathcal{R} \times \mathcal{W};

(11)

Update:\quad u_{\text{new}} = u_{\text{old}} + \delta u, \quad p_{\text{new}} = p_{\text{old}} + \delta p,

where the residual variational form is

\[
\begin{aligned}
    r(u_{\text{old}}, p_{\text{old}}; v, w) = & \left(k(p_{\text{old}})^{-1}u_{\text{old}}, v\right) - \left(p_{\text{old}}, \text{div } v\right) - (f, v) \\
    & - \left(\text{div } u_{\text{old}}, w\right) + (q, w), \quad \forall (v, w) \in (\mathcal{R}, \mathcal{W}),
\end{aligned}
\]  

(12)
and the bilinear form for the Jacobian (Approximate Jacobian) evaluated at $(\mathbf{u}_{\text{old}}, p_{\text{old}})$ is

$$a(\delta \mathbf{u}, \delta p; \mathbf{v}, w) = \left( k(p_{\text{old}})^{-1} \delta \mathbf{u}, \mathbf{v} \right) + \beta \cdot \left( \frac{\partial k^{-1}}{\partial p} \mathbf{u}_{\text{old}} \delta p, \mathbf{v} \right)$$

$$- \left( \delta p, \text{div} \mathbf{v} \right) - \left( \text{div} \delta \mathbf{u}, w \right), \quad \forall (\mathbf{v}, w) \in (\mathcal{R}, \mathcal{W}).$$

(13)

Here $\beta = 0$ leads to Picard’s linearization and $\beta = 1$ leads to Newton’s method.

4.2. Mixed Finite Element discretization

The variational nonlinear Problem 1 and its linearization in (11) are discretized with the Mixed Finite Element method. In particular, we let $\mathcal{T}_h$ denote a triangulation of the domain $\Omega$, $\mathcal{R}_h \subset \mathcal{R}$ be the (lowest order) Raviart–Thomas finite element space consisting of vector functions with a continuous normal component across the interfaces between the elements of $\mathcal{T}_h$ and $\mathcal{W}_h \subset \mathcal{W}$ be the space of piecewise discontinuous polynomials (constant) scalar functions. It is well known that this choice of finite element spaces satisfies the Ladyzhenskaya-Babuška-Brezzi condition, and therefore allows for a stable discretization [39].

To obtain the discrete version of the residual in (12) and of the Jacobian operator in (13), we denote with $\{ \phi^j \}_{j=1}^{n^R_h}$ a basis for the space $\mathcal{R}_h$ and with $\{ \psi^j \}_{j=1}^{n^W_h}$ a basis for the space $\mathcal{W}_h$. Here $n^R_h := \dim(\mathcal{R}_h)$ and $n^W_h := \dim(\mathcal{W}_h)$ denote the dimension (number of degrees of freedom) of the spaces $\mathcal{R}_h$ and $\mathcal{W}_h$, respectively. With this notation, the finite element solution $(\mathbf{u}_h, p_h)$ can be written as a linear combination of the basis functions $\left( \{ \phi^j \}_{j=1}^{n^R_h}, \{ \psi^j \}_{j=1}^{n^W_h} \right)$.

More specifically, letting $\mathbf{u} \in \mathbb{R}^{n^R_h}$ and $p \in \mathbb{R}^{n^W_h}$ denote the vectors collecting the finite element degrees of freedom $\mathbf{u}^j_h$, $j = 1, \ldots, n^R_h$ and $p^j_h$, $j = 1, \ldots, n^W_h$, we write

$$\mathbf{u}_h = \sum_{j=1}^{n^R_h} \mathbf{u}^j_h \phi^j, \quad p_h = \sum_{j=1}^{n^W_h} p^j_h \psi^j.$$

(14)

We then introduce the residual vectors $\mathbf{r}_\mathbf{u} \in \mathbb{R}^{n^R_h}$ and $\mathbf{r}_p \in \mathbb{R}^{n^W_h}$ whose entries are given by

$$(\mathbf{r}_\mathbf{u})_i = \left( k(p_{\text{old}}) \phi^i \right) \mathbf{u}_{\text{old}} - \left( \phi^i \right) \mathbf{f}, \quad i = 1, \ldots, n^R_h;$$

$$\mathbf{r}_p = - \left( \text{div} \mathbf{u}_h, \psi^i \right) + \left( \phi^i \right) \mathbf{q}, \quad i = 1, \ldots, n^W_h.$$  

(15)

Similarly, we define the finite element matrices $M$, $B$ and $N$ whose entries are given by

$$M_{ij} = \left( \frac{\partial k^{-1}}{\partial p} \mathbf{u}_{\text{old}} \phi^j, \phi^i \right), \quad i, j = 1, \ldots, n^R_h;$$

$$B_{ij} = - \left( \text{div} \phi^j, \psi^i \right), \quad i = 1, \ldots, n^R_h, j = 1, \ldots, n^W_h;$$

$$N_{ij} = \left( \frac{\partial k^{-1}}{\partial p} \mathbf{u}_{\text{old}} \psi^j, \phi^i \right), \quad i = 1, \ldots, n^R_h, j = 1, \ldots, n^W_h.$$  

(16)
After Galerkin projection of the residuals and Jacobian on the finite element spaces, the Newton’s/Picard’s step in (11) leads to the solution of the sparse linear system

\[ A\chi = -r, \]  

where the block matrix \( A \) and block vectors \( \chi \) and \( r \) read:

\[ A = \begin{bmatrix} M & B^T + \beta N \\ B & 0 \end{bmatrix}, \quad \chi = \begin{bmatrix} \delta u \\ \delta p \end{bmatrix}, \quad r = \begin{bmatrix} r_u \\ r_p \end{bmatrix}. \]

Here \( \beta = 0 \) leads to Picard’s iterations and \( \beta = 1 \) leads to Newton’s method.

### 4.3. Multilevel formulation

Using the element-based algebraic multigrid (AMGe) methodology summarized in Section 2, we are able to construct operator-dependent coarse spaces with guaranteed approximation properties on general, unstructured grids.

Let \( T_0 \) denote a fine grid triangulation of the domain \( \Omega \). This fine grid is recursively agglomerated into a hierarchy of \( L \) coarser algebraic levels, \( \{T_l\}_{l=1}^L \) where \( L \) denotes the coarsest grid. Agglomerates are formed by grouping together fine-grid elements. Based upon this hierarchy, we build the sequence of spaces \( R_l, W_l \) for \( l = 0, \ldots, L \), that are the discrete analogues for \( H(\text{div}, \Omega) \) and \( L^2(\Omega) \) respectively associated with the (agglomerated) mesh \( \{T_l\}_{l=0}^L \). The space \( R_0 \) is discretized by lowest order Raviart-Thomas finite elements and \( W_0 \) is discretized by piecewise constant finite elements, as explained in Section 4.2. The inf-sup compatibility of the coarse spaces is then a direct consequence of the compatibility of the fine grid spaces \( R_0 \) and \( W_0 \) and of the commutativity of Diagram (4).

In addition, for each \( V \in \{R, W\} \), we define the following prolongation operators from the coarser space \( V_{l+1} \) to the finer space \( V_l \) as

\[ P_l^V : V_{l+1} \mapsto V_l, \quad V \in \{R, W\}. \]  

We also define the following projection operators from the finer space \( V_l \) to the coarser space \( V_{l+1} \) as

\[ \Pi_l^V : V_l \mapsto V_{l+1}, \quad V \in \{R, W\}. \]  

Finally, to apply the FAS V-cycle, we let \( x_l \) collect the unknowns \((u_l, p_l)\) and we define the block prolongation operators

\[ P_l := \begin{bmatrix} P_l^{R_l} & 0 \\ 0 & P_l^{W_l} \end{bmatrix}, \quad \text{for } l = 0, \ldots, L-1 \]

and the block projection operators

\[ \Pi_l := \begin{bmatrix} \Pi_l^{R_l} & 0 \\ 0 & \Pi_l^{W_l} \end{bmatrix}, \quad \text{for } l = 0, \ldots, L-1. \]

Accordingly, we define the nonlinear differential operator \( A_l : (R_l, W_l) \mapsto (R_l^*, W_l^*) \) by the duality mapping

\[ \langle A_l(x_l), y_l \rangle := r(u_l, p_l; v_l, w_l), \quad \forall y_l = (v_l, w_l) \in (R_l, W_l), \]

where \( r \) is the residual form in (12).
4.4. Efficient evaluation of the residual and Jacobian forms at coarse levels

A key component to ensure optimal complexity of the FAS algorithm is the ability to evaluate the residual and the Jacobian forms at coarse levels of the hierarchy without visiting the fine grid level. This is achieved in our framework by exploiting the sophisticated data structures of the AMGe hierarchies. In fact, such data structures — which closely mimic the same data structures of geometric multigrid — store topological tables (i.e. element-element, element-face connectivities) and degree-of-freedom to element mapping for all levels of the hierarchy. Specifically, at each level $l$ and for each space $V_l \otimes \mathbb{P}_p \mathcal{S}_l \otimes \mathcal{Q}_l \otimes \mathcal{R}_l \otimes \mathcal{W}_l$, of the de Rham sequence we denote with $L_{el}^{W_l}$ the mapping between local (to the element $e_l \in \mathcal{T}_l$) and global (for the space $V_l$) degrees of freedom. These local to global mappings are used to assemble local (to each agglomerated element $e_l \in \mathcal{T}_l$) mass matrices into the global one for each space in the de Rham sequence. For example, the global mass matrix $W_l$ for the space $W_l$ at level $l$ is computed as

$$W_l = \sum_{e_l \in \mathcal{T}_l} (L_{el}^{W_l})^T W_{el}(T_{el}^{W_l}),$$

where $L_{el}^{W_l}$ is the local-to-global mapping and $\{W_{el}\}_{e_l \in \mathcal{T}_l}$ are the local (i.e. restricted to each element $e_l$) mass matrices for the space $W_l$. These local mass matrices are computed once during the construction of the AMGe hierarchy by local (to each agglomerated element) Galerkin projections of partially assembled mass matrices from the previous (finer) level, and then stored for future use. In addition, the AMGe technique builds the discrete derivatives operators — such as the discrete divergence operator $D_l$ — at each level of the hierarchy such that (5c) is satisfied.

Then given the current iterate $x_l = [u_l; p_l]$ at level $l$ we can directly evaluate the residual $\eta$ stemming from the discretization of (12) on the coarse spaces $(\mathcal{R}_l, \mathcal{W}_l)$ using local matrices and simple linear algebra as follow

$$\eta = \begin{bmatrix} M_{l}^{p} u_l - B_{l}^{T} p_l - f_l, \\ -B_{l} u_l + q_l \end{bmatrix},$$

where

- $f_l$ and $q_l$ are the Galerkin projection at level $l$ of the forcing term weak forms $(f, \mathbf{v})$ and $(q, w)$, which can be precomputed before invoking the FAS algorithm;
- $B_{l} = W_{l} D_{l}$ is the matrix stemming from the discretization of the divergence weak form, which is computed as the product of the fully-assembled mass matrix $W_{l}$ (for the space $W_{l}$) and the discrete divergence operator $D_{l}$;
- finally, $M_{l}^{p}$ is a weighted fully assembled mass matrix for the space $\mathcal{R}_l$ computed as

$$M_{l}^{p} = \sum_{e_l \in \mathcal{T}_l} s_{e_l} (L_{el}^{R_l})^T M_{e_l} L_{el}^{R_l}.$$
Nonlinear MG with Accurate AMGe Coarse Spaces

Above, we denoted with $M_{el}$ and $L_{el}^{Rl}$ the local (i.e. for the coarse element $e_l$) mass matrix and the local-to-global mapping for the space $R_l$, respectively. The scaling factor $s_{el} = k^{-1}(p_l|_{e_l})$ is computed by restricting $p_l$ (recall $p_l$ is piecewise constant on the elements of $T_l$) to the element $e_l$ and evaluating the nonlinear permeability model.

In a similar way, the coarse Jacobian $A_l$ is given by

$$A_l = \begin{bmatrix} M_{el}^p & B_{el}^T + \beta N_l \\ B_{el} & 0 \end{bmatrix},$$

where $\beta = 0$ leads to Picard’s approximation of the Jacobian and $\beta = 1$ leads to the true Jacobian. The matrix $N_l$ can also be computed directly at level $l$ without visiting the fine grid at the additional cost of storing a small dense 3-dimensional tensor $N_{el}$ for each coarse element $e_l$. Specifically, $N_{el}$ corresponds to the restriction to element $e_l$ of the trilinear form $\left< \psi^j_l, \phi^i_l - \phi^k_l, \phi^l_l \right>$, where we denote with $\psi^j_l$ the local (to element $e_l$) basis functions of $W_l$, and with $\phi^i_l$ and $\phi^k_l$ the local basis functions of $R_l$. Then we compute $N_l$ as

$$N_l = \sum_{e_l \in T_l} s_{el} (L_{el}^{Rl})^T N_{el}^W L_{el}^W,$$

where $(N_{el}^W)_{ij} = \sum_k (N_{el})_{ijk} (L_{el}^{Rl} u_l)^k$ is obtained by saturating the $k$-axis of $N_{el}$ with the restriction of $u_l$ to element $e_l$, and $s_{el} = \frac{\partial k^{-1}}{\partial p} |_{p_l}$ is computed by restricting $p_l$ (recall $p_l$ is piecewise constant on the elements of $T_l$) to the element $e_l$ and evaluating the derivative of the nonlinear permeability model.

In our FAS algorithm we will make use only of the approximated Jacobian $A_l$ for the Picard iteration (i.e. with $\beta = 0$) to avoid the need to store $N_{el}$ and because of the intrinsic difficulty in deriving multilevel solvers for nonsymmetric saddle point systems.

4.5. Linear solvers and preconditioner for the upscaled Jacobian operators

As smoother in the FAS method we apply a global Picard linearization of the nonlinear operator $A_l$, although local linearization is also possible. Specifically, in the smoothing step at each level $l$ of the FAS hierarchy, we compute an approximate solution to the linear system,

$$A_l x_l = r_l,$$

where the matrix $A_l$ in (21) is indefinite (saddle point system). To this aim we employ a single iteration of a specialized indefinite AMGe preconditioner (the Multilevel Divergence Free preconditioner - MLDivFree [25]) developed at LLNL for the solution of the mixed formulation of the Darcy equations. The name of the preconditioner, MLDivFree, is motivated by the fact – shown in [40, 41] – that a single iteration of the preconditioner produces a solution that satisfies the divergence constraint exactly. More specifically, the MLDivFree preconditioner
uses the hierarchy of AMGe coarse spaces — constructed in Section 2 — to form a multilevel preconditioner for symmetric indefinite saddle point problems of the form in (21). It consists of a symmetric V-cycle with a sophisticated multiplicative Hiptmair-type smoother\(^2\). The pre-smoothing involves two stages: first we solve a local saddle point problem for each agglomerated element and then we perform one smoothing iteration (e.g. symmetric Gauss-Seidel) in the divergence free subspace \((\text{curl } \mathcal{Q})\). The post-smoother consists of the same two components but in the reverse order. We refer to [25] for a detailed description of the preconditioner and for numerical results that demonstrate the robustness and scalability properties of the MLDivFree preconditioner for the solution of fine grid and upscaled saddle point systems.

Similarly, each Newton/Picard step requires the solution of a linear system of the form (17) on the fine grid. In the numerical results section, we solve such fine grid system using the GMRES method preconditioned by MLDivFree, thus allowing for a fair comparison among Krylov-(Quasi-)Newton and FAS nonlinear solvers.

5. Numerical results

Five different solver schemes are compared, namely FAS Picard, Exact-GMRES Picard, Inexact-GMRES Picard, Exact-GMRES Newton, and Inexact-GMRES Newton. FAS Picard is a FAS solver with a global Picard linearization of the nonlinear operator at each level. Each smoothing step consists of one V-cycle of the MLDivFree preconditioner. Exact-GMRES Picard and Inexact-GMRES Picard are Krylov-Quasi-Newton solvers with a Picard-like linearization. Each Jacobian system is solved using GMRES preconditioned by MLDivFree with high accuracy for the exact version and less accuracy for the inexact version. Finally, Exact-GMRES Newton and Inexact-GMRES Newton are Krylov-Newton solvers where each Jacobian system is solved with high accuracy (for the first version) and inexacty (for the second version) using GMRES. We stress upon the fact that, even though the Newton linearization leads to a non-symmetric saddle point system, the MLDivFree preconditioner is constructed using the symmetric saddle point system arising from the Picard linearization.

The stopping criterion of the nonlinear solvers used for all experiments is based on the norm of the nonlinear residual \(r^{(k)}\) at iteration \(k\). More specifically, we require \(||r^{(k)}||_2 \leq \max(\text{rtol}||r^{(0)}||_2, \text{atol})\), where \(\text{rtol} = 10^{-6}\) and \(\text{atol} = 10^{-8}\) are the relative and absolute tolerances. For the exact versions of the GMRES Picard and GMRES Newton solvers, we solve the linear system using preconditioned GMRES up to a relative tolerance of \(10^{-9}\) and absolute tolerance of \(10^{-10}\). In the inexact versions we use an Eisenstat-Walker type condition, [43].

\(^2\)Multiplicative or additive smoothers that act both on the original operator and on its restriction onto an auxiliary space representing the near null space are commonly referred as Hiptmair smoothers. This is in recognition of Hiptmair’s seminal work on multigrid methods for the Maxwell’s equations [42].
to determine the relative tolerance for the linear solver GMRES in each nonlinear iteration and to prevent oversolving. More specifically, we set the relative tolerance equal to \( \min(0.5, \sqrt{||r^{(0)}||_2/||r^{(0)}||_2}) \), so that linear systems are solved with less accuracy when we are far from the solution and then the accuracy is increased as we approach the solution to allow for superlinear convergence in the Newton method. Globalization of the Newton/Picard method is achieved by backtracking to guarantee a reduction of the residual norm at each nonlinear iteration, although a more sophisticated line search algorithm (see e.g. [6, 44]) could have also been implemented.

Finally, let us introduce two measures of multigrid performance, namely the arithmetic complexity and the operator complexity that will be used in the following numerical results. The arithmetic complexity \( C_a \) is defined as the ratio of the total number of degrees of freedom on all levels (fine and coarse grids) to the fine grid number of degrees of freedom. In a similar way, the operator complexity \( C_o \) is the ratio of the total number of non-zeros (in the mixed system) on all levels to the number of non-zeros on the fine grid. More specifically, we have

\[
C_a = \frac{\sum_{l=0}^{\text{levels}-1} \text{dim}(\mathcal{R}_l \times \mathcal{W}_l)}{\text{dim}(\mathcal{R}_0 \times \mathcal{W}_0)}, \quad C_o = \frac{\sum_{l=0}^{\text{levels}-1} \text{nnz}(\mathcal{A}_l)}{\text{nnz}(\mathcal{A}_0)}. \tag{22}
\]

We stress upon the fact that many methods can achieve \( C_a \) close to unity and have acceptable approximation properties. However, it is also of vital importance to ensure that \( C_o \) is small (at least sufficiently less than two) to guarantee that the coarse systems take up much less memory than the fine grid problem.

5.1. Unit cube domain and isotropic permeability field

We solve the nonlinear Darcy equations (10) in the domain \( \Omega = (0,1)^3 \) with Dirichlet boundary conditions \( p = p_0 \) on \( \partial \Omega \). We let the forcing terms be \( f := 0 \) and \( q := -1 \). The pressure dependency of the permeability is modeled as

\[
k(p) = k_0 e^{-\alpha(p-p_0)}, \tag{23}
\]

where \( \alpha > 0 \) is the compressibility coefficient and \( k_0 \) is the permeability at a reference pressure \( p_0 = 0 \). We set \( \alpha = 10 \) and let \( k_0 \) be a realization of a lognormal spatially correlated random field with covariance function

\[
C(x, y) := \sigma^2 \exp \left( -\frac{||x - y||_1}{\lambda} \right), \quad x, y \in \Omega. \tag{24}
\]

Here \( || \cdot ||_1 \) denotes the \( l_1 \)-norm in \( \mathbb{R}^3 \), \( \sigma^2 \) denotes the pointwise marginal variance, and \( \lambda \) the correlation length. The covariance function in (24) was originally proposed in [45] and leads to realizations of the random field that have some similarities with actual permeability fields. The realization of \( k_0 \) used in the following tests is shown in Figure 1 and it is obtained with the choice \( \sigma^2 = 9 \) and \( \lambda = 0.1 \). It is generated by means of a truncated Karhunen-Loève (KL)
expansion with 6 eigenmodes in each direction (see [46] for an analytic expression of the eigenvalues and eigenmodes). The KL expansion is chosen for its ability to generate a grid-independent permeability field such that scaling experiments can be easily performed.

In what follows, we present scaling experiments for both a structured hexahedral mesh and an unstructured tetrahedral mesh of the computational domain $\Omega$.

5.1.1. Structured grid scaling

The first study is a comparison between FAS, Newton’s method and Picard iterations for the model problem discretized on a structured cartesian hexahedral mesh. The coarsest mesh consists of 64 elements (4 elements in each spatial direction), and the finest mesh of 262144 elements (64 elements in each spatial direction). In Figure 2 we show the numerical solution computed on the finest mesh. The hierarchy of agglomerated meshes is structured (cartesian) and with a coarsening factor of 2 in each direction. The coarsest level consists of a single agglomerate and the number of multigrid levels ranges from 3 (on the coarsest initial mesh) to 7 (on the finest initial mesh) for these experiments. Table 1 reports the number of fine grid degrees of freedom and the operator and arithmetic complexities of the AMGe hierarchies as a function of the number of elements of the mesh. Both operator and arithmetic complexities are reasonably small independently of the size of the problem; indeed we observe some reduction in the operator complexity as the size of the problem increases. Figure 3 shows the computational time as a function of the number of degrees of freedom for all solver schemes. For the given problem, all solver schemes show optimal scaling with respect to the number of unknowns: the solver times increase linearly with the problem sizes. Moreover, the inexact solvers are faster than the exact counterparts and FAS is the fastest overall. Table 2 reports additional details
on the performances of the five methods we tested. All solver schemes show mesh-independent convergence with respect to the number of nonlinear iterations (#nonlinears) necessary to meet the convergence criterion. The number of nonlinear FAS iterations is roughly one half of those of the other methods, resulting in a speedup (last column of the table) of the time to solution between 1.4x and 4x. It is also worth noticing that, for this particular problem, the Picard’s method converges in nearly the same number of iterations as Newton’s method. One possible explanation of the suboptimal convergence of Newton is that the basin of attraction for the Newton’s method is rather small and that, therefore, backtracking is needed to ensure global converge of the Newton method. Finally, the table also reports the total number of linear (GMRES) iterations (#linears) for the (Quasi)Newton-Krylov solver schemes. All methods show mesh independent convergence also with respect to the number of linear iterations. In addition, early termination of the GMRES inner solvers based on the Eisenstat-Walker condition allows for significant reduction in the total number of linear iterations.

### 5.1.2. Unstructured grid scaling

In this section, we carry out a scaling experiment for an unstructured tetrahedral mesh. A unit cube is meshed with NETGEN, [47, 48], to produce 8
Nonlinear MG with Accurate AMGe Coarse Spaces

<table>
<thead>
<tr>
<th>Scheme</th>
<th>#elements</th>
<th>#linears</th>
<th>#nonlinears</th>
<th>time/time(FAS)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FAS Picard</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>-</td>
<td>5</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>512</td>
<td>-</td>
<td>4</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>4096</td>
<td>-</td>
<td>5</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>32768</td>
<td>-</td>
<td>4</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>262144</td>
<td>-</td>
<td>4</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td><strong>Inexact-GMRES Picard</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>12</td>
<td>9</td>
<td>1.41</td>
<td></td>
</tr>
<tr>
<td>512</td>
<td>11</td>
<td>7</td>
<td>1.45</td>
<td></td>
</tr>
<tr>
<td>4096</td>
<td>13</td>
<td>9</td>
<td>1.41</td>
<td></td>
</tr>
<tr>
<td>32768</td>
<td>12</td>
<td>9</td>
<td>1.71</td>
<td></td>
</tr>
<tr>
<td>262144</td>
<td>12</td>
<td>10</td>
<td>2.16</td>
<td></td>
</tr>
<tr>
<td><strong>Inexact-GMRES Newton</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>16</td>
<td>9</td>
<td>1.65</td>
<td></td>
</tr>
<tr>
<td>512</td>
<td>12</td>
<td>7</td>
<td>1.54</td>
<td></td>
</tr>
<tr>
<td>4096</td>
<td>13</td>
<td>9</td>
<td>1.43</td>
<td></td>
</tr>
<tr>
<td>32768</td>
<td>12</td>
<td>9</td>
<td>1.73</td>
<td></td>
</tr>
<tr>
<td>262144</td>
<td>12</td>
<td>10</td>
<td>2.52</td>
<td></td>
</tr>
<tr>
<td><strong>Exact-GMRES Picard</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>37</td>
<td>9</td>
<td>2.79</td>
<td></td>
</tr>
<tr>
<td>512</td>
<td>33</td>
<td>6</td>
<td>2.85</td>
<td></td>
</tr>
<tr>
<td>4096</td>
<td>34</td>
<td>7</td>
<td>2.30</td>
<td></td>
</tr>
<tr>
<td>32768</td>
<td>37</td>
<td>7</td>
<td>3.17</td>
<td></td>
</tr>
<tr>
<td>262144</td>
<td>35</td>
<td>7</td>
<td>3.55</td>
<td></td>
</tr>
<tr>
<td><strong>Exact-GMRES Newton</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>59</td>
<td>9</td>
<td>4.07</td>
<td></td>
</tr>
<tr>
<td>512</td>
<td>38</td>
<td>6</td>
<td>3.22</td>
<td></td>
</tr>
<tr>
<td>4096</td>
<td>41</td>
<td>7</td>
<td>2.78</td>
<td></td>
</tr>
<tr>
<td>32768</td>
<td>42</td>
<td>7</td>
<td>3.48</td>
<td></td>
</tr>
<tr>
<td>262144</td>
<td>44</td>
<td>7</td>
<td>3.96</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Information on the scaling experiments (structured grid). #linears is the total number of linear iterations for all nonlinear iterations.
unstructured meshes with increasing resolution. For all 8 meshes, the performance of FAS is compared to the performance of inexact Picard. We restrict ourselves to these two nonlinear methods, since they have proven to be the fastest in the previous example.

Agglomerated meshes are generated using graph-partitioning algorithms (METIS) leading to agglomerated elements with arbitrary shapes, non-planar faces, and arbitrary number of neighboring elements. In addition, a post-processing of agglomerated elements is required to ensure certain topological requirements are met. Agglomerated elements that do not meet these requirements are split into agglomerated elements from the previous (finer) level and left unagglomerated until the next (coarser) level, where they are subject to agglomeration again. These irregularities in the coarse topology usually lead to an increase of both the arithmetic and operator complexity (for the same target accuracy), adding additional challenges to the upscaling procedure. We use aggressive coarsening and one initial level of geometric coarsening to ensure small operator and arithmetic complexities. Specifically, the first level of the mesh hierarchy is geometric (i.e. mesh de-refinement) and the following levels are algebraic. The unstructured coarsening factor (METIS) used for the algebraic levels is 100 finer elements per agglomerated element. Figures 4 - 7 provide an example of the topology produced by the procedure. Table 3 reports the number of fine grid degrees of freedom, the operator and arithmetic complexities, and the number of levels of the AMGe hierarchies as a function of the number of elements in

\[3\] Specifically, agglomerated elements need to be simply connected with simply connected boundary.
Nonlinear MG with Accurate AMGe Coarse Spaces

Figure 4: Level 0
Uniformly refined once

Figure 5: Level 1
Agglomerates are based on uniform refinement

Figure 6: Level 2
Agglomerates are formed by METIS

Figure 7: Level 3
Coarsest level is a single agglomerate

the mesh. Note that the aggressive coarsening allows us to obtain arithmetic and operator complexities comparable to those obtained using the structured hierarchy in the previous experiment, however we notice a slight increase in the operator complexity for the larger problems.

From Table 4, it can be seen that both algorithms perform optimally in terms of linear and nonlinear iterations. If we look at Figure 8, the computational time is good, but slightly suboptimal due to the operator complexity not remaining constant. This can be remedied by increasing the coarsening factor for the larger scale problems.

5.2. Egg model

The Egg model [49] has 101 synthetic realizations of channelized reservoir permeability fields. In particular, we are using the default one, which can be downloaded from the model’s website [50]. The name of the model comes from its shape being like an egg. The egg shape is obtained from a structured grid by deleting some elements. The resulting boundary of the model is irregular, so we treat the resulting grid as unstructured. As in the previous example, we impose Dirichlet boundary conditions $p = p_0$, and we let the forcing terms be $f := 0$ and $q := -1$. The pressure dependency of the permeability is modeled as in (23),
Table 3: Number of levels, operator and arithmetic complexities of the AMGe hierarchies as a function of problem size (unstructured grid).

<table>
<thead>
<tr>
<th>#elements</th>
<th>#DoFs</th>
<th>operator complexity</th>
<th>arithmetic complexity</th>
<th>#levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>32280</td>
<td>98644</td>
<td>1.19</td>
<td>1.135</td>
<td>4</td>
</tr>
<tr>
<td>53832</td>
<td>164268</td>
<td>1.21</td>
<td>1.135</td>
<td>4</td>
</tr>
<tr>
<td>67624</td>
<td>207056</td>
<td>1.21</td>
<td>1.136</td>
<td>4</td>
</tr>
<tr>
<td>215512</td>
<td>654332</td>
<td>1.23</td>
<td>1.136</td>
<td>4</td>
</tr>
<tr>
<td>405632</td>
<td>1227916</td>
<td>1.29</td>
<td>1.140</td>
<td>5</td>
</tr>
<tr>
<td>496800</td>
<td>1507968</td>
<td>1.34</td>
<td>1.144</td>
<td>5</td>
</tr>
<tr>
<td>679808</td>
<td>2066320</td>
<td>1.34</td>
<td>1.140</td>
<td>5</td>
</tr>
<tr>
<td>827144</td>
<td>2500552</td>
<td>1.34</td>
<td>1.140</td>
<td>5</td>
</tr>
</tbody>
</table>

Figure 8: Computational time for FAS and inexact Picard as a function of problem size (unstructured grid).
where we set $\alpha = 0.5$ and let $k_0$ be the default permeability field for the egg model. As shown in Figure 9, the reference permeability field $k_0$ features high-permeability channels in a low-permeable background representing a typical pattern encountered in fluvial environments. $k_0$ is an anisotropic tensor whose $z$-component is one order of magnitude smaller than the $x/y$-components.

The model consists of 7 layers in the vertical direction and contains 18553 elements, each of them of size $8m \times 8m \times 4m$. To perform the scaling experiments we uniformly refine the mesh twice. The number of fine-scale degrees of freedom ranges from 77738 (59205 for the velocity and 18533 for the pressure) on the original mesh to 4806304 (3618912 and 1187392, respectively) on the finest mesh. Figure 10 shows the numerical solution computed on the intermediate mesh. For all 3 meshes, we compare the performance of FAS Picard to that of Inexact-GMRES Picard.

To generate the agglomerated meshes we use first geometric de-refinement (if applicable) and then two levels of unstructured agglomeration. The agglomerated meshes for the two unstructured levels are shown in Figure 11 and feature a coarsening factor of 32 elements per agglomerated element. Table 5 shows the number of levels, operator and arithmetic complexities of the AMGe hierarchy as a function of the number of fine grid elements and degrees of freedom. Operator and arithmetic complexities are very close to one and do not show any significant increase as we refine the mesh. Figure 12 and Table 6 show that, also for this more realistic example, both FAS Picard and Inexact-GMRES Picard perform optimally: the computational time scales linearly with the number of fine grid degrees of freedom, and the number of (nonlinear and linear) itera-
Nonlinear MG with Accurate AMGe Coarse Spaces

Figure 9: x/y component of the reference permeability field (egg model). The z component is one order of magnitude smaller.

Figure 10: Numerical solution computed on the intermediate mesh: pressure (horizontal clip) on the left and velocity magnitude on the right (egg model).

6. Conclusion & perspectives

In the present work, AMGe coarse spaces with guaranteed approximation properties have been exploited to define a hierarchy of discretizations for nonlinear Darcy PDEs arising from subsurface modeling, which lead to challenging saddle-point problems. The resulting discretization operators and their linearizations were utilized within classical FAS with a specialized divergence-free smoother, which resulted in a scalable nonlinear solver.

To the best of our knowledge, this work is the first to show optimal scaling results for a nonlinear algebraic multilevel solver on general unstructured

<table>
<thead>
<tr>
<th>#elements</th>
<th>#DoFs</th>
<th>operator complexity</th>
<th>arithmetic complexity</th>
<th>#levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>18553</td>
<td>77758</td>
<td>1.112</td>
<td>1.042</td>
<td>3</td>
</tr>
<tr>
<td>148424</td>
<td>607880</td>
<td>1.139</td>
<td>1.132</td>
<td>4</td>
</tr>
<tr>
<td>1187392</td>
<td>4806304</td>
<td>1.143</td>
<td>1.143</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 5: Number of levels, operator and arithmetic complexities of the AMGe hierarchies as a function of problem size (egg model).
Nonlinear MG with Accurate AMGe Coarse Spaces

Figure 11: Hierarchy of agglomerated meshes (egg model).

Figure 12: Computational time for FAS Picard and Inexact-GMRES Picard as a function of the problem size (egg model).

<table>
<thead>
<tr>
<th>Scheme</th>
<th>#elements</th>
<th>#linears</th>
<th>#nonlinears</th>
<th>time/time(FAS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FAS Picard</td>
<td>18553</td>
<td>-</td>
<td>3</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>148424</td>
<td>-</td>
<td>3</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>1187392</td>
<td>-</td>
<td>3</td>
<td>-</td>
</tr>
<tr>
<td>Inexact-GMRES Picard</td>
<td>18553</td>
<td>21</td>
<td>8</td>
<td>3.60</td>
</tr>
<tr>
<td></td>
<td>148424</td>
<td>22</td>
<td>8</td>
<td>3.11</td>
</tr>
<tr>
<td></td>
<td>1187392</td>
<td>24</td>
<td>9</td>
<td>3.35</td>
</tr>
</tbody>
</table>

Table 6: Information on scaling experiments (egg model). #linears is the total number of linear iterations for all nonlinear iterations.
Previous attempts to apply FAS to unstructured mesh problems (see e.g. [13, 14, 15]) suffered from a degradation of performances due to the lack of approximation properties of the coarse spaces. Having the right kind of coarse spaces is the key to obtain mesh independent convergence: nonlinear multilevel solvers without approximation properties of coarse spaces should not (and in fact do not) work.

Specifically, we performed numerical tests, both for structured and unstructured meshes, demonstrating the mesh independent convergence of our FAS algorithm in terms of number of V-cycles. We compared the FAS-AMGe solver to exact and inexact Newton’s method and respective Picard iterations; FAS outperformed the exact methods (4x faster) and also proved faster than the inexact versions. These are, indeed, very promising results: to beat by a constant factor well established optimal complexity algorithms is the best one could possibly hope for.

In this paper, only global linearization has been considered, in part because of the intrinsic difficulty — even in the case of linear solvers (see e.g. [20]) — of developing multilevel algebraic solvers for saddle point problems. A study comparing with local linearization techniques is worth exploring in the future. The results are encouraging and offer the potential for exploring the method for other classes of nonlinear PDEs that can utilize components from the de Rham complex of AMGe coarse spaces with guaranteed approximation properties, which is an active (ongoing) work.

References

Nonlinear MG with Accurate AMGe Coarse Spaces


Nonlinear MG with Accurate AMGe Coarse Spaces


Nonlinear MG with Accurate AMGe Coarse Spaces


[38] D. N. Arnold, R. S. Falk, R. Winther, Multigrid in H(div) and H(curl), Numerische Mathematik 85 (2000) 197 – 217.


Appendix A. AMGe construction of the coarse $H(\text{div})-L^2$-conforming spaces

In this appendix we summarize the construction of the coarse $H(\text{div})-L^2$-conforming spaces with guaranteed approximation properties. The construction of the coarse $H(\text{curl})$-conforming space, used by the MLDivFree preconditioner, is similar in spirit but more involved. For additional details we refer to [26, 18, 19].

In what follows we denote with $T_H$ the set of coarse (agglomerated) elements $T$, and with $F_H$ the set of coarse (agglomerated) faces $F$. Following [26], for each $F \in F_H$, we denote with $\mathbf{n}_F$ the unit normal vector to $F$ defined. The orientation of $\mathbf{n}_F$ is arbitrarily chosen and fixed to point outside one of the two agglomerated elements that share the coarse face $F$. We also define the following local finite element spaces for each coarse element $T \in T_H$,

\[
\begin{align*}
\mathcal{R}_T &= \{ \mathbf{v}_h \in \mathcal{R}_h | \text{supp(} \mathbf{v}_h \text{) } \subset T \text{ and } \mathbf{v}_h \cdot \mathbf{n} = 0 \text{ on } \partial T \} , \\
\mathcal{W}_T &= \{ w_h \in \mathcal{W}_h | \text{supp(} w_h \text{) } \subset T \text{ and } (w_h, 1)_T = 0 \} .
\end{align*}
\]

The coarse $L^2$-conforming space $W_H \subset W_h$ is constructed as in [26] and consists of piecewise constant functions on agglomerated elements. A basis of the space $W_H$ is then used to form the columns of the prolongation matrix $P : W_H \rightarrow W_h$.

We then construct a coarse $H(\text{div})$-conforming space $\mathcal{R}_H \subset \mathcal{R}_h$ such that $\text{div } \mathcal{R}_H = W_H$. This property is necessary to preserve the stability of the upscaled discretization and to guarantee that the spaces $(\mathcal{R}_H, W_H)$ are inf-sup compatible. The construction closely follows that of [18], where coarse $H(\text{div})$-conforming spaces with guaranteed approximation properties are introduced for the first time. The method is a two-step process, where we first find the coarse basis functions on coarse faces and then we extend such basis functions into the interior of the neighboring agglomerated elements.

The coarse basis functions are defined in terms of their fine degrees of freedom (DoFs). Given a sufficiently smooth vector function $\mathbf{r}$ and a fine face $f$, the value of the DoF associated with $f$ is defined as

\[
\text{DoF}_f(\mathbf{r}) = \int_f \mathbf{r} \cdot \mathbf{n}_f \, dA ,
\]

where $\mathbf{n}_f$ is the unit normal to the fine face and $A$ is the surface area of the face.

For each coarse face $F \in F_H$, a matrix $\mathbf{W}_F$ is formed. $\mathbf{W}_F$ consists of the values of the DoFs of the fine faces $f_1, \ldots, f_{|F|}$ constituting $F$, i.e.

\[
\mathbf{W}_F = \begin{bmatrix}
\text{DoF}_{f_1}(\mathbf{n}_F) & \text{DoF}_{f_2}(\mathbf{e}_1) & \text{DoF}_{f_3}(\mathbf{e}_2) & \text{DoF}_{f_4}(\mathbf{e}_3) \\
\vdots & \vdots & \vdots & \vdots \\
\text{DoF}_{f_{|F|}}(\mathbf{n}_F) & \text{DoF}_{f_{|F|}}(\mathbf{e}_1) & \text{DoF}_{f_{|F|}}(\mathbf{e}_2) & \text{DoF}_{f_{|F|}}(\mathbf{e}_3)
\end{bmatrix} ,
\]
where $|F|$ is the number of fine faces in the coarse face $F$, $n_F$ is the unit normal vector to $F$, and $e_i$ ($i = 1, 2, 3$) stand for the three coordinate constant vector-functions. The goal is to ensure that the coarse Raviart-Thomas space contains locally (on each agglomerated element) these constant vectors, hence has first order of approximation in $L^2$ as the fine-grid Raviart-Thomas space. The above construction is fairly general; we can include any given functions in $H^{p}(\text{div})$ of our interest in the coarse $H^{p}(\text{div})$ space and maintain the compatibility. Note that if $W_F$ only contains $n_F$ the method from [26] is recovered. Using an SVD decomposition $W_F = U \Sigma V^T$, the linearly dependent columns of $W_F$ are eliminated. The left singular vectors (columns of $U$) $u_j$ are chosen based on the corresponding singular values $\sigma_j$. If $\sigma_j \geq \epsilon \sigma_{\text{max}}$, where $\epsilon \in (0,1]$ is a user-given input, then $u_j$ defines a coarse basis function for the coarse face $F$ denoted $r_j^F$. More precisely, $r_j^F$ is only equal to $u_j$ on the coarse face $F$ and zero everywhere else.

The above procedure describes the first step to construct the coarse $H^{p}(\text{div})$ space. The second step involves taking the partially defined functions $r_j^F$ and extending these coarse basis functions into the interior of the coarse elements. The extension is performed using the approach in [26], which guarantees that the divergence of the coarse Raviart-Thomas space belongs to the coarse $L^2$ space.

Given a partially defined function $r_j^F$ on the coarse face $F$ belonging to the boundary of coarse element $T$, the local (element-based) mixed system reads

\begin{align*}
\text{Problem 2 Element extension: } \text{Find } (\tilde{r}_T, p_h) & \in \mathcal{R}_T \times \tilde{W}_T \text{ such that } \\
\begin{cases} 
(\alpha (\tilde{r}_T + r_j^F), v_h)_T + (p_h, \text{div } v_h)_T = 0, & \forall v_h \in \mathcal{R}_T \\
(\text{div } (\tilde{r}_T + r_j^F), w_h)_T = 0, & \forall w_h \in \tilde{W}_T.
\end{cases}
\end{align*}

The coefficient matrix $\alpha$ (a $3 \times 3$ SPD matrix) can be set equal to the coefficients from the original problem (such as permeabilities), but this is not strictly necessary. Problem 2 is guaranteed to have a unique solution, [18, 26], provided that $T$ is connected. By solving these local problems on each pair of agglomerates $(T^+, T^-)$ adjacent to a coarse face $F$, we obtain the coarse basis functions $r_h = r_j^F + r_{T^+} + r_{T^-}$ of the space $\mathcal{R}_H$. We finally let the columns of the prolongation matrix $P : \mathcal{R}_H \rightarrow \mathcal{R}_h$ be the collection of the coarse basis functions $r_h$. 