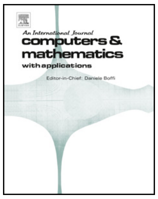
Computers and Mathematics with Applications 157 (2024) 15–26

Contents lists available at ScienceDirect

Computers and Mathematics with Applications

journal homepage: www.elsevier.com/locate/camwa



Automatic variationally stable analysis for finite element computations: Transient convection-diffusion problems

Eirik Valseth [a,b,c,*], Pouria Behnoudfar [d], Clint Dawson [a], Albert Romkes [e]

[a] Oden Institute for Computational Engineering and Sciences, University of Texas at Austin, Austin, TX 78712, USA
[b] The Department of Data Science, The Norwegian University of Life Science, Drøbakveien 31, Ås 1433, Norway
[c] Department of Scientific Computing and Numerical Analysis, Simula Research Laboratory, Kristian Augusts gate 23, Oslo, 0164, Norway
[d] Mineral Resources, Commonwealth Scientific and Industrial Research Organisation (CSIRO), Kensington, Perth, WA 6152, Australia
[e] Department of Mechanical Engineering, South Dakota School of Mines & Technology, Rapid City, SD 57701, USA

ARTICLE INFO

We dedicate this work to Prof. Lsezek F. Demkowicz on the occasion of his 70th birthday

Keywords:
Stability
Discontinuous Petrov-Galerkin
Method of lines
Space-time finite element method
Adaptive mesh refeinement

ABSTRACT

We present an application of stable finite element (FE) approximations of convection-diffusion initial boundary value problems (IBVPs) using a weighted least squares FE method, the automatic variationally stable finite element (AVS-FE) method [1]. The transient convection-diffusion problem leads to issues in classical FE methods as the differential operator can be considered a singular perturbation in both space and time. The stability property of the AVS-FE method, allows us significant flexibility in the construction of FE approximations in both space and time. Thus, in this paper, we take two distinct approaches to the FE discretization of the convection-diffusion problem: i) considering a space-time approach in which the temporal discretization is established using finite elements, and ii) a method of lines approach in which we employ the AVS-FE method in space whereas the temporal domain is discretized using the generalized-α method. We also consider another space-time technique in which the temporal direction is partitioned, thereby leading to finite space-time "slices" in an attempt to reduce the computational cost of the space-time discretizations.

We present numerical verifications for these approaches, including numerical asymptotic convergence studies highlighting optimal convergence properties. Furthermore, in the spirit of the discontinuous Petrov-Galerkin (DPG) method by Demkowicz and Gopalakrishnan [2–6], the AVS-FE method also leads to readily available *a posteriori* error estimates through a Riesz representer of the residual of the AVS-FE approximations. Hence, the norm of the resulting local restrictions of these estimates serves as error indicators in both space and time for which we present multiple numerical verifications in mesh adaptive strategies.

1. Introduction

Transient BVPs are commonplace in engineering applications and to date still pose significant challenges in numerical analysis and numerical modeling. Time dependency in many BVPs, such as the heat equation, involve partial derivatives of the trial variable with respect to time and leads to numerical instabilities unless careful considerations are taken. The reason being that the time derivative is a convective transport term, i.e., transient problems may lead to unstable discretizations, particularly in the FE context. Additionally, the target problem of convection-diffusion also result in numerical instabilities in its spatial discretizations which lead to the development of the AVS-FE method in [1]. To overcome the stability issues in both space and time we propose two distinct approaches employing the AVS-FE method. First, we take a space-time approach in which space and time are discretized directly considering time an additional dimension using the AVS-FE method. Second, we consider a method of lines to decouple the computations in space and time and employ a generalized α method for the temporal discretization [7–9].

The use of space-time FE methods remains attractive as the approximations are standard FE approximations and therefore inherit attractive features of FE methods such as *a priori* and *a posteriori* error estimation and mesh adaptive strategies. Examples of space-time FE methods can be found in, e.g., [10–12]. The AVS-FE method [1] being stable for any differential operator is therefore a prime candidate for space-time FE discretizations. Its stability property is a consequence of the philoso-

[*] Corresponding author at: Oden Institute for Computational Engineering and Sciences, University of Texas at Austin, Austin, TX 78712, USA.
E-mail address: eirik.valseth@nmbu.no (E. Valseth).

https://doi.org/10.1016/j.camwa.2023.12.025

Available online 4 January 2024

phy of the DPG method in which the test space consists of functions that are computed on-the-fly from Riesz representation problems [2–6]. In [13], the AVS-FE method is successfully employed in space and time for the Cahn-Hilliard BVP. The goal here was the extension of the AVS-FE method to a nonlinear BVP as well as an initial verification of AVS-FE space-time solutions. Similarly, in [14], the AVS-FE method is employed for space time solutions of a nonlinear transient wave propagation problem, the Korteweg de-Vries equation. Furthermore, its built-in *a posteriori* error estimate and their corresponding error indicators can be directly applied to drive adaptivity. The DPG method has been successfully applied to several transient problems, e.g., convection-diffusion and the Navier-Stokes equations [15–17]. These space-time formulations are available in the DPG FE code Camellia of Nathan Roberts [18]. Recent efforts in DPG methods for transient problems include the use of optimal testing in time, see e.g., [19,20]

Alternatively, the method of lines can be employed to decouple the discretization of space and time where the spatial dimension is discretized to obtain a discrete system at each time step. Then, using a time integrator, the discretization of the temporal domain subsequently results in a fully discrete system of equations. Here, we employ the AVS-FE method in space and the generalized-α method in time. Chung and Hulbert introduced the generalized-α method in [9] to solve hyperbolic problems and extended it to parabolic differential equations such as Navier-Stokes equations in [21]. The method provides second-order accuracy in the temporal domain as well as unconditional stability. Although the method allows us to control the numerical dissipation in high-frequency regions, it delivers adequately accurate results in low-frequency domains. Introduction of a user-defined parameter provides this control and includes the HHT-α method of Hilber, Hughes, Taylor [22] and the WBZ-α method of Wood, Bossak, and Zienkiewicz [23].

In the following, we introduce the AVS-FE method for transient BVPs by taking the two distinct approaches introduced above. In Section 2 we introduce our model problem and notations in addition to a review of the AVS-FE methodology and present the AVS-FE weak formulation to be used. In this section we also present the discretization of the weak form, an alternative saddle point structure of the AVS-FE method, and its built-in *a posteriori* error estimate. In Section 3 we present the time discretization techniques: the method of lines using AVS-FE method in space and generalized-α method in time is presented in Section 3.1; and the space-time AVS-FE method in Section 3.2. Results from numerical verifications for numerous PDEs and applications are presented in Section 4. Finally, we draw conclusions and discuss potential directions of future work in Section 5.

2. The AVS-FE methodology

The AVS-FE method [1] allows us to compute stable FE approximations to BVPs for any differential operator, provided its kernel is trivial and the computations of optimal test functions are sufficiently accurate. In this section we introduce our model problem and briefly review the AVS-FE method, a thorough introduction can be found in [1].

2.1. Model problem and notation

Let $\Omega \subset \mathbb{R}^N$, $N \leq 2$ be an open bounded domain with Lipschitz boundary $\partial\Omega$ and outward unit normal vector $\mathbf{n}$, and let T be the final time. Then, define $\Omega_T = \Omega \times (0, T)$ to be the space time domain which is open and bounded with a Lipschitz boundary $\partial\Omega_T = \overline{\Gamma_{in} \cup \Gamma_{out} \cup \Gamma_0 \cup \Gamma_T}$. Γ_{in} and Γ_{out} are the in and outflow boundaries, respectively, and Γ_0 and Γ_T are the initial and final time boundaries, respectively. The transient model problem is therefore the following linear convection-diffusion IBVP:

Find u such that:

$$\begin{aligned} \frac{\partial u}{\partial t} - \boldsymbol{\nabla} \cdot (\varepsilon \boldsymbol{\nabla} u) + \mathbf{b} \cdot \boldsymbol{\nabla} u &= f, && \text{in } \Omega_T, \\ u &= u_{in}, && \text{on } \Gamma_{in}, \\ \varepsilon \boldsymbol{\nabla} u \cdot \mathbf{n} &= g, && \text{on } \Gamma_{out}, \\ u &= u_0, && \text{on } \Gamma_0, \end{aligned} \tag{1}$$

where $\varepsilon \in L^\infty(\Omega)$ denotes the isotropic diffusion parameter; $\mathbf{b} \in [L^2(\Omega)]^N$ the convection coefficient; $f \in L^2(\Omega)$ the source function; and $g \in H^{-1/2}(\Gamma_{out})$ the Neumann boundary data. Note that the gradient operator $\boldsymbol{\nabla}$ refers to the spatial gradient operator, e.g., $\boldsymbol{\nabla}(\cdot) = \{\frac{\partial(\cdot)}{\partial x}, \frac{\partial(\cdot)}{\partial y}\}^{\mathrm{T}}$.

2.2. Weak formulation and FE discretization

We omit the full derivation of the weak formulation here and mention key points only. The derivation of a weak formulation for the AVS-FE method is shown in, e.g. [1]. To establish a weak formulation of (1), we need a regular partition $\mathcal{P}_h$ of Ω_T into elements K_m, such that:

$$\Omega_T = \text{int}\Big(\bigcup_{K_m \in \mathcal{P}_h} \overline{K_m} \Big).$$

We introduce a flux variable $\mathbf{q} = \varepsilon \boldsymbol{\nabla} u$, and recast (1) as a system of (distributional) first-order PDEs:

Find $(u, \mathbf{q}) \in L^2([0, T]; H^1(\Omega) \times H(\text{div}, \Omega))$ such that:

$$\begin{aligned} \boldsymbol{\nabla} u - \frac{1}{\varepsilon}\mathbf{q} &= 0, && \text{in } \Omega_T, \\ \frac{\partial u}{\partial t} - \boldsymbol{\nabla} \cdot \mathbf{q} + \mathbf{b} \cdot \boldsymbol{\nabla} u &= f, && \text{in } \Omega_T, \\ u &= u_{in}, && \text{on } \Gamma_{in}, \\ \mathbf{q} \cdot \mathbf{n} &= g, && \text{on } \Gamma_{out}, \\ u &= u_0, && \text{on } \Gamma_0. \end{aligned} \tag{2}$$

Note that the flux variable $\mathbf{q}$ depends on time but only has the same number of components as the dimension of Ω and in the weak enforcement of the PDE, it belongs to $L^2([0, T]; H(\text{div}, \Omega))$.

To derive the AVS-FE weak formulation, we enforce the PDEs (2) weakly on each element $K_m \in \mathcal{P}_h$, apply integration by parts to shift all derivatives to the test functions except the time derivative. After subsequent summation of the local contributions we arrive at the global variational formulation:

Find $(u, \mathbf{q}) \in U(\Omega_T)$ such that:

$$B((u, \mathbf{q}), (v, \mathbf{w})) = F((v, \mathbf{w})), \quad \forall (v, \mathbf{w}) \in V(\mathcal{P}_h), \tag{3}$$

In (3), the bilinear form, $B : U(\Omega_T) \times V(\mathcal{P}_h) \longrightarrow \mathbb{R}$, and linear functional, $F : V(\mathcal{P}_h) \longrightarrow \mathbb{R}$, are defined:

$$\begin{aligned} B((u, \mathbf{q}), (v, \mathbf{w})) \overset{\text{def}}{=} \sum_{K_m \in \mathcal{P}_h} \Bigg\{ \int_{K_m} \Bigg[-u \boldsymbol{\nabla} \cdot \mathbf{w}_m - \frac{1}{\varepsilon}\mathbf{q} \cdot \mathbf{w}_m + \frac{\partial u}{\partial t} v_m \\ + \mathbf{q} \cdot \boldsymbol{\nabla} v_m - (\mathbf{b} \cdot \boldsymbol{\nabla} v_m) u \Bigg] \,\mathrm{d}\mathbf{x} \\ + \oint_{\partial K_m} \Big[(\mathbf{b} \cdot \mathbf{n}) \gamma_0^m(u) \gamma_0^m(v_m) + \gamma_{\mathbf{n}}^m(\mathbf{w}_m) \gamma_0^m(u) - \gamma_{\mathbf{n}}^m(\mathbf{q}) \gamma_0^m(v_m) \Big] \,\mathrm{d}s \Bigg\}, \\ F((v, \mathbf{w})) \overset{\text{def}}{=} \sum_{K_m \in \mathcal{P}_h} \int_{K_m} f\, v_m \,\mathrm{d}\mathbf{x}, \end{aligned} \tag{4}$$

where the *continuous* trial and *broken* test function spaces, $U(\Omega_T)$ and $V(\mathcal{P}_h)$, are defined as follows:

$$\begin{aligned} U(\Omega_T) &\overset{\text{def}}{=} \Big\{ (u,\mathbf{q}) \in L^2([0,T]; H^1(\Omega)\times H(\text{div},\Omega)) : \\ &\qquad u|_{\Gamma_0} = u_0,\, u|_{\Gamma_{in}} = u_{in},\, \mathbf{q}\cdot\mathbf{n}|_{\Gamma_{out}} = g \Big\}, \\ V(\mathcal{P}_h) &\overset{\text{def}}{=} \Big\{ (v,\mathbf{w}) \in H^1(\mathcal{P}_h)\times H(\text{div},\mathcal{P}_h) \Big\}. \end{aligned} \tag{5}$$

The broken Hilbert spaces are defined:

$$\begin{aligned} H^1(\mathcal{P}_h) &\overset{\text{def}}{=} \Big\{ v \in L^2(\Omega_T) :\ v_m \in H^1(K_m),\ \forall K_m \in \mathcal{P}_h \Big\}, \\ H(\text{div},\mathcal{P}_h) &\overset{\text{def}}{=} \Big\{ \mathbf{w} \in [L^2(\Omega_T)]^2 :\ \mathbf{w}_m \in H(\text{div},K_m),\ \forall K_m \in \mathcal{P}_h \Big\}, \end{aligned} \tag{6}$$

and the norms on these spaces $\|\cdot\|_{U(\Omega_T)} : U(\Omega_T) \longrightarrow [0,\infty)$ and $\|\cdot\|_{V(\mathcal{P}_h)} : V(\mathcal{P}_h) \longrightarrow [0,\infty)$ are defined as follows:

$$\begin{aligned} \|(u,\mathbf{q})\|_{U(\Omega_T)} &\overset{\text{def}}{=} \sqrt{\int_\Omega \Big[\nabla u\cdot\nabla u + u^2 + (\nabla\cdot\mathbf{q})^2 + \mathbf{q}\cdot\mathbf{q} \Big]\, \text{d}\mathbf{x}}. \\ \|(v,\mathbf{w})\|_{V(\mathcal{P}_h)} & \\ &\overset{\text{def}}{=} \sqrt{\sum_{K_m\in\mathcal{P}_h} \int_{K_m} \Big[h_m^2 \nabla v_m\cdot\nabla v_m + v_m^2 + h_m^2(\nabla\cdot\mathbf{w}_m)^2 + \mathbf{w}_m\cdot\mathbf{w}_m \Big]\, \text{d}\mathbf{x}}. \end{aligned} \tag{7}$$

The operators $\gamma_0^m : H^1(K_m) :\longrightarrow H^{1/2}(\partial K_m)$ and $\gamma_{\mathbf{n}}^m : H(\text{div},K_m) \longrightarrow H^{-1/2}(\partial K_m)$ denote the trace and normal trace operators on K_m.

The bilinear form and linear functional in (4) differs from the ones presented in [1] due to the term $\frac{\partial u}{\partial t}$ and the application of integration by parts to all terms involving spatial derivatives. This weak formulation (3) represents a DPG formulation as the test space is broken and continuity of the trial space is a result of the definition of its subspaces. In the following we review important points of the AVS-FE method and for the sake of simplicity, consider the case with homogeneous Dirichlet boundary conditions ($u|_{\partial\Omega_T} = 0$) which are enforced strongly in the trial space $U(\Omega_T)$. We focus only on the equivalent saddle point formulation to (3) of the AVS-FE, and other discrete least squares methods, and refer to the extensive literature on the subject, e.g., the classical text of Bochev and Gunzberger [24] and in particular the work of Keith et al. [25] for further details on the solution of the normal equation (3) using optimal test functions.

2.3. Equivalent saddle point problem

The discretization of (3) can be implemented in existing FE software by redefining routines that compute the element stiffness matrices to compute optimal test functions on-the-fly, as in e.g., [1]. However, in several commonly used FE solvers, such as FEniCS [26] or Firedrake [27], manipulations of the element assembly routines may not as easily be performed. Thus, to enable straightforward implementation into these FE solvers, we will introduce an equivalent interpretation of the AVS-FE method as a global saddle point problem. We omit several details here and highlight only key features of this interpretation, interested readers are referred to [28] for a complete presentation.

The AVS-FE method is a weighted least squares, or minimum residual method, in the sense that its solution realizes the minimum of a functional according to the following principle:

$$\boxed{u^h = \underset{v^h\in U^h(\Omega)}{\arg\min}\ \frac{1}{2}\|\mathbb{B}v^h - \mathbb{F}\|^2_{V(\mathcal{P}_h)'},} \tag{8}$$

where $\mathbb{B}$ and $\mathbb{F}$ are operators induced by the bilinear and linear forms, respectively. Furthermore, we can relate the norm on the dual space $V(\mathcal{P}_h)'$ $\|\cdot\|_{V(\mathcal{P}_h)'}$ to the energy norm:

$$\|(u,\mathbf{q})\|_{\text{B}} \overset{\text{def}}{=} \sup_{(v,\mathbf{w})\in V(\mathcal{P}_h)\setminus\{(0,\mathbf{0})\}} \frac{|B((u,\mathbf{q}),(v,\mathbf{w}))|}{\|(v,\mathbf{w})\|_{V(\mathcal{P}_h)}}, \tag{9}$$

using the Riesz representers $(p,\mathbf{r})$ that solves the following problem for $(u,\mathbf{q})$:

$$(\,(p,\mathbf{r}),(v,\mathbf{w})\,)_{V(\mathcal{P}_h)} = B(\,(u,\mathbf{q}),(v,\mathbf{w})), \quad \forall (v,\mathbf{w}) \in V(\mathcal{P}_h). \tag{10}$$

Analogously, we can consider a Riesz representer of the approximation error $(u-u^h, \mathbf{q}-\mathbf{q}^h)$, which we refer to as an *error representation function* [3]. This error representation function $(\hat{e},\hat{\mathbf{E}})$ is then defined as the solution of the following weak problem:

$$\boxed{\begin{aligned} &\text{Find } (\hat{e},\hat{\mathbf{E}}) \in V(\mathcal{P}_h) \quad \text{such that:} \\ &((\hat{e},\hat{\mathbf{E}}),(v,\mathbf{w}))_{V(\mathcal{P}_h)} \\ &\quad = \underbrace{F(v,\mathbf{w}) - B(\,(u^h,\mathbf{q}^h),(v,\mathbf{w})\,)}_{\text{Residual}} \quad \forall (v,\mathbf{w}) \in V(\mathcal{P}_h). \end{aligned}} \tag{11}$$

The energy norm of $(u-u^h, \mathbf{q}-\mathbf{q}^h)$ can be identified by the $V(\mathcal{P}_h)$ norm of the error representation function:

Proposition 2.1. *Let $(u,\mathbf{q}) \in U(\Omega)$ be the solution of the AVS-FE weak form (3) and $(u^h,\mathbf{q}^h) \in U^h(\Omega)$ its corresponding AVS-FE approximation. Then, the energy norm of $(u-u^h,\mathbf{q}-\mathbf{q}^h)$ is identical to the $V(\mathcal{P}_h)$ norm of $(\hat{e},\hat{\mathbf{E}})$:*

$$\|(u-u^h,\mathbf{q}-\mathbf{q}^h)\|_{\text{B}} = \|(\hat{e},\hat{\mathbf{E}})\|_{V(\mathcal{P}_h)}. \tag{12}$$

Proof. This proof is known from existing DPG literature (see Section 1 and equation (1.17) in [6]). The identity is a consequence of the definition of the energy norm (9) and the weak problem governing the error representation function (11). □

The norm of approximate error representation function $(\hat{e}_h,\hat{\mathbf{E}}_h)$ is therefore an *a posteriori* error estimate, i.e.,

$$\|(u-u^h,\mathbf{q}-\mathbf{q}^h)\|_{\text{B}} \approx \|(\hat{e}_h,\hat{\mathbf{E}}_h)\|_{V(\mathcal{P}_h)}. \tag{13}$$

Furthermore, its local restriction can be computed element-wise as the space $V(\mathcal{P}_h)$ is broken to yield the error indicator:

$$\eta = \|(\hat{e}_h,\hat{\mathbf{E}}_h)\|_{V(K_m)}. \tag{14}$$

This type of error indicator has been applied with great success to multiple problems (see, e.g., [3,6,29,30]), and we show several numerical experiments using this indicator for the AVS-FE method in Section 4. It should be noted that this error estimate and the error indicator are known to be robust (i.e., bounded above and below) under the assumption of the existence of DPG Fortin operators and localizable norms [3,31,32].

The minimum residual interpretation allows us to establish the following AVS-FE saddle point formulation to which we seek the approximate solution $(u^h,\mathbf{q}^h)$ under the constraint that the error representation function minimizes the residual of the AVS-FE method, see (11):

$$\boxed{\begin{aligned} &\text{Find } (u^h,\mathbf{q}^h) \in U^h(\Omega), (\hat{e}_h,\hat{\mathbf{E}}_h) \in V^h(\mathcal{P}_h) \quad \text{such that:} \\ &\quad \Big((\hat{e}_h,\hat{\mathbf{E}}_h),(v^h,\mathbf{w}^h) \Big)_{V(\mathcal{P}_h)} + B((u^h,\mathbf{q}^h),(v^h,\mathbf{w}^h)) = F(v^h,\mathbf{w}^h), \\ &\quad \forall (v^h,\mathbf{w}^h) \in V^h(\mathcal{P}_h), \\ &\quad B((p^h,\mathbf{r}^h),(\hat{e}_h,\hat{\mathbf{E}}_h)) = 0, \quad \forall (p^h,\mathbf{r}^h) \in U^h(\Omega). \end{aligned}} \tag{15}$$

Solution of (15) gives both the AVS-FE solution for $(u^h,\mathbf{q}^h)$ and its error representation functions $(\hat{e}_h,\hat{\mathbf{E}}_h)$ in a single global solution step. This is very convenient as we now have a built-in *a posteriori* error estimate and error indicators immediately upon solving (15). However, the computational cost of doing so has been shifted from local computations for optimal test functions to the global cost of a larger system of equations. Fortunately, the global nature of (15) allows for very simple implementation of the AVS-FE method in readily available FE solvers like FEniCS [26] and Firedrake [27]. Note that dropping the weighted derivative terms from the inner product corresponding to the norm $\|\cdot\|_{V(\mathcal{P}_h)}$ reduces (15) to a DPG implementation of the first-order system least squares method. Note that the analysis of (15) can be performed using the famous Brezzi theory [33,34]. In [25], a general framework for saddle point problems arising in discrete least squares methods, such as (15) is presented which is also applicable in the present setting. Since the inner product is a coercive linear operator, and the bilinear form satisfies a discrete *inf-sup* condition, the saddle point system is also well posed. □

3. Time discretization

In the weak formulation (3) we have made no assumptions on the type of discretization of the time domain. Here, we consider two distinctive cases of time discretization techniques. In both cases the spatial discretizations are performed with finite elements and the AVS-FE methodology. First, we consider a discretization of the time domain by employing the method of lines to decouple the spatial and time discretization and subsequently employing the generalized-α method. Second, the discrete stability property of the AVS-FE method allows us to discretize the time domain with finite elements in a space-time approach.

3.1. Method of lines

In this section, we first discuss the method in an abstract setting before proceeding to the particular case of the AVS-FE method and generalized-α methods. To this end, we define two Hilbert spaces $\mathrm{U}(\Omega)$ and $\mathrm{V}(\Omega)$, and introduce a well-posed weak formulation for a transient BVP, e.g., the convection-diffusion problem of Section 2.1:

$$\boxed{\begin{aligned}&\text{Find } \mathrm{u}\in\mathrm{U}(\Omega) \quad \text{such that:}\\ &\qquad\qquad \mathrm{b}(\mathrm{u},\mathrm{v}) = F(\mathrm{v}), \quad \forall\, \mathrm{v}\in\mathrm{V}(\Omega),\end{aligned}} \tag{16}$$

where the bilinear form b contains all spatial and temporal terms. To seek approximations of (16), we introduce the time derivative operator $\mathcal{L}$, and consider FE polynomial subspaces of $\mathrm{U}(\Omega)$ and $\mathrm{V}(\Omega)$, i.e., $\mathrm{U}^h(\Omega)$ and $\mathrm{V}^h(\Omega)$. Thus, we get:

$$\boxed{\begin{aligned}&\text{Find } \mathrm{u}^h\in\mathrm{U}^h(\Omega) \quad \text{such that:}\\ &\qquad \left(\mathcal{L}(\mathrm{u}^h),\mathrm{v}^h\right)_{L^2(\Omega)} + \mathrm{b}_h(\mathrm{u}^h,\mathrm{v}^h) = F(\mathrm{v}^h), \quad \forall\, \mathrm{v}^h\in\mathrm{V}^h(\Omega),\end{aligned}} \tag{17}$$

where $(\cdot,\cdot)_{L^2(\Omega)}$ denotes the $L^2(\Omega)$ inner product, and the bilinear operator $\mathrm{b}_h(\cdot,\cdot)$ contains only spatial derivative terms. We assume that this formulation is well-posed.

To advance the solution in time, we consider a uniform partition of the time domain from $t_0 = 0$ to the final time $t_N = T$, with τ the distance between each step t_i. We compute approximations to u^h at each step using second-order accurate generalized-α methods presented in [9,21]. For parabolic or first-order hyperbolic problems, the generalized-α method for the transient term $\mathcal{L}(\mathrm{u}^h)$ in (17) is to find $\mathrm{u}_h^{n+1}\in\mathrm{U}^h(\Omega)$, such that:

$$(\vartheta_h^{n+\alpha_g},\mathrm{v}_h)_{L^2(\Omega)} + \mathrm{b}_h(\mathrm{u}_h^{n+\alpha_f},\mathrm{v}_h) = F^{n+\alpha_f}(\mathrm{v}_h), \quad \forall\,\mathrm{v}_h\in\mathrm{V}(\Omega)_h, \tag{18}$$

where u_h^n, ϑ_h^n are the approximations to $\mathrm{u}(.,t_n)$ and $\dfrac{\partial \mathrm{u}(.,t_n)}{\partial t}$, respectively. We discuss the initialization of ϑ_h^n in Remark 3.1 and Section 3.1.2. The unknowns at time step $n+1$ are updated using the solutions at $n+\alpha_f$ and $n+\alpha_g$ as:

$$\boxed{\begin{aligned}\mathrm{u}^{n+\alpha_f} &= \mathrm{u}^n + \alpha_f\delta(\mathrm{u}^n), \quad && \delta(\mathrm{u}^n) = \mathrm{u}^{n+1} - \mathrm{u}^n.\\ \vartheta_h^{n+\alpha_g} &= \vartheta_h^n + \alpha_g\delta(\vartheta_h^n), \quad && \delta(\vartheta_h^n) = \vartheta_h^{n+1} - \vartheta_h^n.\end{aligned}} \tag{19}$$

Using a Taylor expansion, we have $\mathrm{u}^{n+1} = \mathrm{u}^n + \tau\vartheta^n + \tau\gamma\delta(\vartheta^n)$ as a linear combination of $\mathrm{u}^n, \vartheta^n$ with γ guaranteeing second-order accuracy. Substitution of the expressions in (19) into (18) gives:

$$(\vartheta_h^{n+1},\mathrm{v}_h)_{L^2(\Omega)} + \mathrm{b}_h(\zeta\,\vartheta_h^{n+1},\mathrm{v}_h) = (\frac{1}{\alpha_g}l^{n+1},\mathrm{v}_h), \quad \forall\,\mathrm{v}_h\in\mathrm{V}(\Omega)_h, \tag{20}$$

where $\zeta = \frac{\tau\gamma\alpha_f}{\alpha_g}$, and:

$$\begin{aligned}l^{n+1} = F^{n+\alpha_f} &+ (\alpha_g - 1)\left(\vartheta_h^n,\mathrm{v}_h\right)_{L^2(\Omega)} + \tau\alpha_f(\gamma-1)\,\mathrm{b}_h\left(\vartheta_h^n,\mathrm{v}_h\right)\\ &- \mathrm{b}_h\left(\mathrm{u}_h^n,\mathrm{v}_h\right).\end{aligned} \tag{21}$$

It can be shown that this scheme is formally second order accurate (see [21]) if we select:

$$\gamma = \frac{1}{2} + \alpha_g - \alpha_f. \tag{22}$$

To avoid having to select both α_g and α_f, these parameters are defined in terms of the spectral radius ρ_∞. This parameter essentially provides a means to control numerical dissipation. Hence, α_g and α_f are defined:

$$\alpha_g = \frac{1}{2}\left(\frac{3-\rho_\infty}{1+\rho_\infty}\right), \qquad \alpha_f = \frac{1}{1+\rho_\infty}. \tag{23}$$

Remark 3.1. *The generalized-α method requires additional initial data for ϑ_h^0. This value is obtained by setting $\alpha_f = \alpha_g = n = 0$ and solving (18).*

Remark 3.2. *The spectral radius ρ_∞ is a user-defined parameter that provides control on the numerical dissipation such that for $\rho_\infty = 1$ there is no dissipation control, and the maximum control is delivered by setting $\rho_\infty = 0$. Numerical dissipation can occur for example in the case of poor spatial resolution (for more details, see, [35,36]).*

3.1.1. Generalized-α and the AVS-FE method

Having introduced the generalized-α method for a well defined weak formulation, we now extend it to the AVS-FE method for our model IBVP of convection-diffusion. Hence, let us consider the AVS-FE weak formulation (3), and the trial and test spaces $U(\Omega)$ and $V(\mathcal{P}_h)$ analogous to (5). The generalized-α method for the AVS-FE method is:

$$\boxed{\begin{aligned}&\text{Find } (\vartheta_h^{n+1},\mathbf{q}_h^{n+1})\in U^h(\Omega) \text{ such that:}\\ &\quad (\vartheta_h^{n+1},v^*)_{L^2(\Omega)} + B_h((\zeta\,\vartheta_h^{n+1},\mathbf{q}_h^{n+1}),(v^*,\mathbf{w}^*)) = \tfrac{1}{\alpha_g}\ell^{n+1}(v^*,\mathbf{w}^*),\\ &\qquad\qquad \forall\,(v^*,\mathbf{w}^*)\in V^*(\mathcal{P}_h),\end{aligned}} \tag{24}$$

where the operators are defined:

$$\begin{aligned}&B_h((u,\mathbf{q}),(v,\mathbf{w}))\\ &\overset{\text{def}}{=} \sum_{K_m\in\mathcal{P}_h}\Bigg\{\int_{K_m}\Big[-u\,\varepsilon\boldsymbol{\nabla}\cdot\mathbf{w}_m - \mathbf{q}\cdot\mathbf{w}_m + \mathbf{q}\cdot\boldsymbol{\nabla}v_m - (\mathbf{b}\cdot\boldsymbol{\nabla}v_m)\,u\Big]\,\mathrm{d}\mathbf{x}\\ &+ \oint_{\partial K_m}\Big[(\mathbf{b}\cdot\mathbf{n})\,\gamma_0^m(u)\gamma_0^m(v_m) + \gamma_{\mathbf{n}}^m(\mathbf{w}_m)\,\gamma_0^m(u) - \gamma_{\mathbf{n}}^m(\mathbf{q})\,\gamma_0^m(v_m)\Big]\,\mathrm{d}s\Bigg\},\end{aligned} \tag{25}$$

$$\begin{aligned}\ell^{n+1}((v,\mathbf{w})) &\overset{\text{def}}{=} \sum_{K_m\in\mathcal{P}_h}\int_{K_m}(f^{n+\alpha_f}\,v)\,\mathrm{d}\mathbf{x} + (\alpha_g-1)\left(\vartheta_h^n,v\right)_{L^2(\Omega)}\\ &+\tau\alpha_f(\gamma-1)\cdot B_h\left((\vartheta_h^n,\mathbf{0}),(v,\mathbf{w})\right) - B_h\left((u_h^n,\mathbf{q}_h^n),(v,\mathbf{w})\right).\end{aligned}$$

To establish the solutions to (24) we take the same approach introduced in Section 2.3 and define a saddle point system similar to (15). The major difference between the "original" weak form (3) and the one corresponding to the generalized-α method, i.e., (24) other than the adjusted bilinear and linear forms, is the term $(\vartheta_h^{n+1}, \mathrm{v})_{L^2(\Omega)}$. Analogous to the case in Section 2.3, the approximation to (24) is governed by the following minimization problem:

$$\boxed{\vartheta_h^{n+1} = \underset{z_h \in U^h(\Omega)}{\arg\min}\ \frac{1}{2}\|\mathbb{l}^{n+1} - \left(\mathbb{M} + \zeta \mathbb{B}_h\right)\, z_h\|^2_{V_h'},} \tag{26}$$

where the operators $\mathbb{B}_h$ and $\mathbb{l}^{n+1}$ correspond to the actions of the adjusted forms B_h and ℓ^{n+1}, respectively, and $\mathbb{M}$ to the new term $(\vartheta_h^{n+1}, \mathrm{v})_{L^2(\Omega)}$. Thankfully, the Riesz map (induced by the equivalent of the Riesz representation problem (10) for (24)) allows us to relate the norm on the dual space $\|\cdot\|_{V_h'}$ to the energy norm on $U(\Omega)$. Hence, we define the following error representation function:

$$\boxed{\begin{aligned}
&\text{Find } (\hat{e}^{n+1}, \hat{\mathbf{E}}^{n+1}) \in V(\mathcal{P}_h) \quad \text{such that:}\\
&((\hat{e}^{n+1}, \hat{\mathbf{E}}^{n+1}), (\upsilon, \mathbf{w}))_{V(\mathcal{P}_h)}\\
&\quad = \underbrace{\ell^{n+1}(\upsilon, \mathbf{w}) - (\vartheta_h^{n+1}, \upsilon)_{L^2(\Omega)} + B_h((\zeta\ \vartheta_h^{n+1}, \mathbf{q}_h^{n+1}), (\upsilon, \mathbf{w}))}_{\text{Residual}},\\
&\qquad\qquad\qquad\qquad\qquad\qquad \forall\, (\upsilon, \mathbf{w}) \in V(\mathcal{P}_h).
\end{aligned}} \tag{27}$$

which now measures how far we are from the best approximation of $(\vartheta_h^{n+1}, \mathbf{q}_h^{n+1})$ at the current time step. In the same fashion as in Section 2.3, the norm of this function is an *a posteriori* error estimate and its restriction to each $K_m \in \mathcal{P}_h$ an error indicator. We finally can introduce the saddle point problem for each time step:

$$\boxed{\begin{aligned}
&\text{Find } (\vartheta_h^{n+1}, \mathbf{q}_h^{n+1}) \in U^h(\Omega), (\hat{e}_h^{n+1}, \hat{\mathbf{E}}_h^{n+1}) \in V_h(\mathcal{P}_h) \text{ such that:}\\
&((\hat{e}_h^{n+1}, \hat{\mathbf{E}}_h^{n+1}), (\upsilon_h, \mathbf{w}_h))_{V_h} + ((\vartheta_h^{n+1}, \mathbf{0}), (\upsilon_h, \mathbf{w}_h))_{L^2(\Omega)}\\
&\quad + \zeta \cdot B_h((\vartheta_h^{n+1}, \mathbf{q}_h^{n+1}), (\upsilon_h, \mathbf{w}_h)) = \frac{1}{\alpha_g}\ell^{n+1}((\upsilon_h, \mathbf{w}_h)),\\
&\qquad\qquad\qquad\qquad\qquad\qquad \forall\, (\upsilon_h, \mathbf{w}_h) \in V_h(\mathcal{P}_h),\\
&((z_h, \mathbf{r}_h), (\hat{e}_h^{n+1}, \mathbf{0}))_{L^2(\Omega)} + \zeta \cdot B_h((z_h, \mathbf{r}_h), (\hat{e}_h^{n+1}, \hat{\mathbf{E}}_h^{n+1})) = 0,\\
&\qquad\qquad\qquad\qquad\qquad\qquad \forall\, (z_h, \mathbf{r}_h) \in U^h(\Omega),
\end{aligned}} \tag{28}$$

where the inner product $(\cdot,\cdot)_{V_h}$ is defined:

$$\begin{aligned}
&((\hat{e}_h^{n+1}, \hat{\mathbf{E}}_h^{n+1}), (\upsilon_h, \mathbf{w}_h))_{V_h}\\
&\quad = ((\zeta \cdot \hat{e}_h^{n+1}, \hat{\mathbf{E}}_h^{n+1}), (\upsilon_h, \mathbf{w}_h))_{V(\mathcal{P}_h)} + ((\hat{e}_h^{n+1}, \mathbf{0}), (\upsilon_h, \mathbf{w}_h))_{L^2(\Omega)}.
\end{aligned} \tag{29}$$

Computing ϑ_h^{n+1} from (28), we obtain u_h^{n+1} from a Taylor expansion at each time step. The overall procedure requires a matrix solve at each time step as well as two explicit updates. Note that as both (25) and the inner product (29) both depend on ζ in the same fashion. Hence, the solutions u_h computed from (28) do not change with ζ, whereas the error representation function $(\hat{e}_h^{n+1}, \hat{\mathbf{E}}_h^{n+1})$ does.

Next, we show that our proposed saddle-point problem (28) is unconditionally stable in the temporal domain. To achieve this, we must show that our AVS-FE spatial discretization scheme does not alter the unconditional stability of generalized-α method.

Theorem 3.1. *The saddle-point problems in* (28) *provide unconditionally stable solutions in temporal domain.*

Proof. Our proof relies on established bounds from literature for the generalized-α method [35], and reasoning based on the properties of the AVS-FE saddle point problem (28). By applying the generalized-α method on the continuous parabolic problem (2) to discrete the temporal domain, we obtain a fully discrete problem that can be written:

$$\begin{bmatrix} u^{n+1} \\ \tau\vartheta^{n+1} \end{bmatrix} = \Xi \begin{bmatrix} u^{n} \\ \tau\vartheta^{n} \end{bmatrix} + \Pi\, l^{n+\alpha_f}, \tag{30}$$

with Ξ and Π being a 2×2 amplification matrix and a 2×1 matrix, respectively. The amplification matrix allows us to write the solution at time step $n+1$ using an initial condition and a forcing term. The derivation and development of the amplification matrix can be found in, e.g., [7–9]. If the eigenvalues of this amplification matrix are bounded by one, the method is stable see Theorem 2 in [8]. Hence:

$$\tau\|\vartheta^{n+1}\|^2_{L^2(\Omega)} \leq \tau\|\vartheta^{n}\|^2_{L^2(\Omega)} + \frac{1}{\tau}\|\pi_2\, l^{n+\alpha_f}\|^2_{L^2(\Omega)}, \tag{31}$$

where π_2 is the second component of Π. Considering the saddle-point problem (28) with unknown ϑ_h^{n+1}, we add the term $\|\vartheta_h^{n+1} - \vartheta^{n+1}\|_{L^2(\Omega)}$ to the right-side of the inequality (31) and the inequality still holds. Next, using $\|\vartheta^{n+1} - \vartheta_h^{n+1}\|_{L^2(\Omega)} \leq \|\vartheta^{n+1} - \vartheta_h^{n+1}\|_{V(\mathcal{P}_h)}$, the Cauchy-Schwartz inequality, and the error representation provided by the AVS-FE method, we get:

$$\begin{aligned}
\tau\|\vartheta_h^{n+1}\|^2_{L^2(\Omega)} &\leq \tau\|\vartheta^{n}\|^2_{L^2(\Omega)} + \tau\|\vartheta_h^{n+1} - \vartheta^{n+1}\|^2_{L^2(\Omega)} + \frac{1}{\tau}\|\pi_2\, l^{n+\alpha_f}\|^2_{L^2(\Omega)}\\
&\leq \tau\|\vartheta^{n}\|^2_{L^2(\Omega)} + C\sqrt{\tau}\|(\hat{e}_h^{n+1}\|^2_{V(\mathcal{P}_h)} + \frac{1}{\tau}\|\pi_2\, l^{n+\alpha_f}\|^2_{L^2(\Omega)},\\
&\leq \tau\|\vartheta^{0}\|^2_{L^2(\Omega)}\\
&\quad + \sum_{j=0}^{j=N-1}\left(C\sqrt{\tau}\|(\hat{e}_h^{j}\|^2_{V(\mathcal{P}_h)} + \frac{1}{\tau}\|\pi_2\, l^{j+\alpha_f}\|^2_{L^2(\Omega)}\right),
\end{aligned}$$

where $C > 0$ is a constant. Hence, the solution is bounded by the initial solution, forcing, and error representation terms. □

3.1.2. Retrieving initial data

As pointed out in Remark 3.1, we need to retrieve the additional initial data ϑ_h^0 to solve (28). Hence, we set $\alpha_f = \alpha_g = 0$ and get:

$$\boxed{\begin{aligned}
&\text{Find } (\vartheta_h^{0}, \mathbf{q}_h^{0}) \in U^h(\Omega), (\hat{e}_h^{0}, \hat{\mathbf{E}}_h^{0}) \in V_h(\mathcal{P}_h) \text{ such that:}\\
&((\hat{e}_h^{0}, \hat{\mathbf{E}}_h^{0}, (\upsilon_h, \mathbf{w}_h))_{V_h} + ((\vartheta_h^{0}, \mathbf{0}), (\upsilon_h, \mathbf{w}_h))_{L^2(\Omega)}\\
&\quad = \ell^{0}((\upsilon_h, \mathbf{w}_h)) - \zeta \cdot b_h((u^0, \mathbf{q}_h^{0}), (\upsilon_h, \mathbf{w}_h)),\\
&\qquad\qquad\qquad\qquad\qquad \forall\, (\upsilon_h, \mathbf{w}_h) \in V_h(\mathcal{P}_h),\\
&((z_h, \mathbf{r}_h), (\hat{e}_h^{0}, \mathbf{0}))_{L^2(\Omega)} = 0, \qquad \forall\, (z_h, \mathbf{r}_h) \in U^h(\Omega),
\end{aligned}} \tag{32}$$

where u^0, $\mathbf{q}^0$, and $\ell^0((\upsilon_h, \mathbf{w}_h))$ correspond to the initial data. To ascertain that the problem for the initial data is well posed (32), we have the following proposition.

Proposition 3.1. *Let $(\upsilon_h, \mathbf{w}_h) \in V_h$ be arbitrary test functions. Then, $\vartheta_h^0 \in U^h(\Omega)$ exists and is unique.*

We omit the proof here as it is trivial to show that $((\vartheta_h^0, \mathbf{0}), (\upsilon_h, \mathbf{w}_h))_{L^2(\Omega)}$, i.e. the $L^2(\Omega)$ inner product, satisfies the following three properties:

- Stability: There exists a constant $C_{sta} > 0$ independent of the mesh size, such that:

$$\inf_{0 \neq z_h \in U^h(\Omega)}\ \sup_{0 \neq \upsilon_h \in V_h} \frac{|((z_h, \mathbf{0}), (\upsilon_h, \mathbf{w}_h))_{L^2(\Omega)}|}{\|z_h\|_{L^2(\Omega)}\|\upsilon_h\|_{L^2(\Omega)}} \geq C_{sta}. \tag{33}$$

- Consistency: Employing a similar argument as [30] to study the consistency of the saddle-point problem, we can state the consistency as:

$$\begin{aligned}
&((\vartheta_h^0, \mathbf{0}), (\upsilon_h, \mathbf{w}_h))_{L^2(\Omega)} = (f^0, (\upsilon_h, \mathbf{w}_h))\\
&- \zeta \cdot b_h((u_h^0, \mathbf{q}_h^0), (\upsilon_h, \mathbf{w}_h)), \quad \forall\, (\upsilon_h, \mathbf{w}_h) \in V_h
\end{aligned} \tag{34}$$

- Boundedness: There exists a constant $C_{bnd} < \infty$, uniformly with respect to the mesh size, such that:

$$((z,\mathbf{0}),(v_h,\mathbf{w}_h))_{L^2(\Omega)} \le C_{bnd}\,\|z\|_{L^2(\Omega)}\|v_h\|_{L^2(\Omega)}, \quad \forall (z,v_h)\in U\times V_h. \tag{35}$$

See [37] for details on these conditions. □

Thus, using (32), we have a stable and adaptive method to find the initial data which is critical for the generalized-α method to ensure second-order accuracy in time.

3.2. *Space-time FE approach*

The use of FE discretizations for transient problems is commonly avoided due to the inherently unstable nature of transient problems. The discretizations must be very carefully constructed to achieve discrete stability using the classical FE method. However, the stability of the AVS-FE method allows us to discretize the entire space-time domain with finite elements in a straightforward manner. Furthermore, *a posteriori* error estimates and error indicators are immediately available to us as error indicators are obtained directly in the saddle point approach of the AVS-FE method (15).

To establish AVS-FE space-time approximations of weak formulation (3) or (15), we pick appropriate discretizations of the space $L^2([0,T];H^1(\Omega)\times H(\text{div},\Omega))$. For $L^2([0,T];H^1(\Omega))$, the choice is classical FE basis functions that are C^0 continuous functions in the space-time domain Ω_T such as Lagrange or Legendre polynomials. For $L^2([0,T];H(\text{div},\Omega))$, a conforming choice of basis is, e.g., a tensor product basis of Raviart-Thomas and $L^2[0,T]$ functions. However, as in [1], we employ approximations for $\mathbf{q}^h$ by C^0 polynomials for each of its components as this has shown to yield superior results for convex domains and sufficiently regular sources. In particular, we generate 3D FE meshes on which the bases are defined, which for the scalar valued function is trivial, and in the case of $\mathbf{q}^h$, we employ restrictions of the 3D C^0 basis functions to 2D. The discretized saddle point problem is therefore:

$$\boxed{\begin{aligned}&\text{Find } (u^h,\mathbf{q}^h)\in U^h(\Omega_T),(\hat{e}_h,\hat{\mathbf{E}}_h)\in V^h(\mathcal{P}_h) \quad \text{such that:}\\ &\Big((\hat{e}_h,\hat{\mathbf{E}}_h),(v^h,\mathbf{w}^h)\Big)_{V(\mathcal{P}_h)} + B((u^h,\mathbf{q}^h),(v^h,\mathbf{w}^h))\\ &\quad = F(v^h,\mathbf{w}^h), \quad \forall (v^h,\mathbf{w}^h)\in V^h(\mathcal{P}_h),\\ &B((p^h,\mathbf{r}^h),(\hat{e}_h,\hat{\mathbf{E}}_h)) = 0, \quad \forall (p^h,\mathbf{r}^h)\in U^h(\Omega_T).\end{aligned}} \tag{36}$$

where the components of $U^h(\Omega_T)$ are spanned by continuous FE basis functions and $V^h(\mathcal{P}_h)$ by discontinuous FE basis functions.

3.2.1. *Time slice approach*

As an alternative to the space-time discretization of the full space-time domain Ω_T, in this section we introduce a time slice approach for the AVS-FE method. While the space-time approach introduced in the preceding section allows straightforward implementation of the AVS-FE method and its "built-in" error indicator can drive mesh adaptive refinements, the large number of degrees of freedom quickly makes the method intractable. In an effort to localize the computational cost of the space-time approach, we propose to partition the space-time domain into "space-time slices". The slices can be constructed in a number of ways, from uniformly to a graded mesh structure as considered in [15, 38] for the DPG method.

To advance in time, a solution can be obtained on a slice which can be transferred to the neighboring slice as an initial condition. Hence, we can perform mesh refinements on each slice to ensure the complete resolution of any interior or boundary layer (i.e., physical features) before proceeding to the next. This is of particular interest in applications in which physical parameters are time dependent leading to widely different solution features as time progresses. In Fig. 1, an arbitrary domain Ω_T is shown and is partitioned into two space-time slices. Note that the approach of time slices is not fully equivalent to the full space-time domain as only information of u^h is transferred between slices.

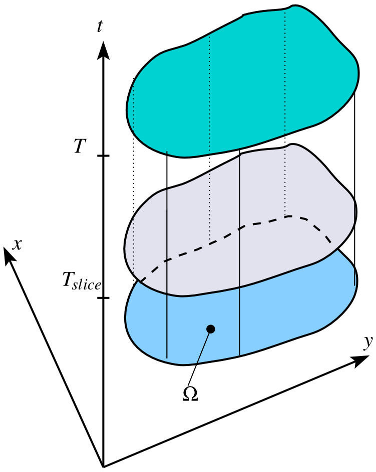


Fig. 1. Partition of space-time domain into slices.

4. Numerical verifications

To conduct numerical verifications, we consider the following form of our model scalar-valued convection diffusion problem (1):

$$\begin{aligned}\frac{\partial u}{\partial t} - \varepsilon\Delta u + \mathbf{b}\cdot\nabla u &= f, && \text{in } \Omega_T,\\ u &= u_0, && \text{on } \partial\Omega_T,\\ u &= u_{intial}, && \text{on } \partial\Omega_T\cap\{t=0\},\end{aligned} \tag{37}$$

where the coefficient ε is a constant diffusion coefficient. We first study the effect of approximation degree of the optimal test functions in Section 4.1. Next, we verify the convergence properties of the AVS-FE method for both time discretization schemes in Section 4.2. In this section, we also investigate the use of time slices as well as compare the space-time method to the method of lines with generalized-α time stepping. Last, in Section 4.4 we present verifications of a problem with both a hyperbolic and a parabolic part, i.e., a transient convection-diffusion problem. The particular case we investigate corresponds to a challenging physical application, a shock wave problem.

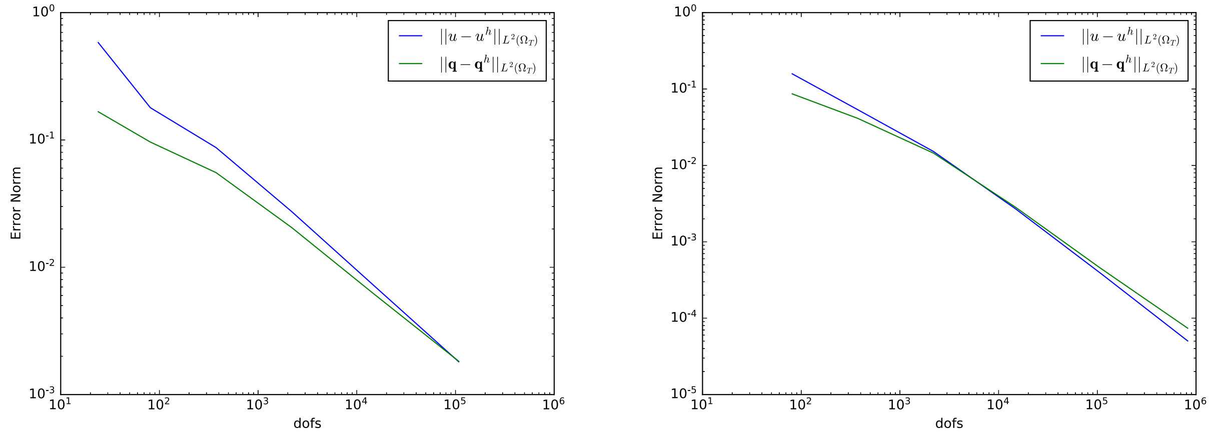
In all the presented numerical experiments we use the saddle point description in (15) implemented in legacy FEniCS [26] with the latest stable release from Anaconda. The verifications in which we employ adaptive refinements all use the same criterion as in [30], i.e., the built-in error indicator (14) as well as a Dörfler marking strategy [39] using the approximate energy error computed using (12). To solve the system of linear algebraic equations, we use the direct solver MUMPS [40,41]. Also note that in all cases where we report the number of degrees of freedom, we do not include the degrees of freedom for the error representation function in the saddle point systems (15) and (28). The polynomial degree of approximation used for this error representation function is identical to the degree of the trial space with the results in Section 4.1 being the sole exception.

4.1. *Optimal test function resolution*

As an initial verification, we perform a study to ensure proper resolution of the optimal test space. To this end, we consider the following exact solution:

$$u(x,y,t) = e^{-t}\left[x + \frac{e^{\frac{b_x}{\varepsilon}x}-1}{1-e^{\frac{b_x}{\varepsilon}}}\right]\left[y + \frac{e^{\frac{b_y}{\varepsilon}y}-1}{1-e^{\frac{b_y}{\varepsilon}}}\right], \tag{38}$$

from which we establish initial and exact solution boundary conditions and a corresponding source term f. For these studies we consider the



(a) Linear polynomial approximations.

(b) Quadratic polynomial approximations.

Fig. 2. Convergence histories for the space-time convergence study.

Table 1
Increasing degree of approximation for the error representation function.

p_{trial}	p_{test}	$\|u-u^h\|_{L^2(\Omega_T)}$ (coarsest mesh)	$\|u-u^h\|_{L^2(\Omega_T)}$ (finest mesh)
1	1	1.1439e-01	2.6374e-03
1	2	1.1439e-01	2.6559e-03
1	3	1.1439e-01	2.6576e-03
1	4	1.1439e-01	2.6582e-03
2	1	6.8837e-02	1.4028e-04
2	2	6.7822e-02	1.3770e-04
2	3	6.8246e-02	1.3716e-04
2	4	6.8277e-02	1.3698e-04
2	5	6.8288e-02	1.3695e-04

moderately convection dominated case with $\varepsilon = 0.1$, $\mathbf{b} = \{1, 1\}$, and select the final time of computation to be $T = 0.5s$. We consider only the space-time case here and assume that the conclusions apply to the generalized-α case as well. Due to the smoothness of the exact solution, we consider continuous polynomial approximations for both variables of equal order - p. The error representation functions are then discretized with discontinuous polynomials of order $p + 0, 1, 2, 3$, as well as $p - 1$ for $p \geq 2$. In Table 1, these results are presented for linear and quadratic trial functions for two uniform meshes: 6 and 24,576 space-time tetrahedrons, respectively. The results in this table indicate that for linear and quadratic bases for the trial space, the impact of increasing test space degree is vanishing small. We observe the same trend for $p > 2$. Note that for $p = 2$, we observe satisfactory results for a test space degree $p = 1$.

4.2. Convergence studies

To numerically investigate the convergence properties of our methods, we consider a well-known example of transient convection-diffusion, the Eriksson-Johnson problem [42]. This problem has a known exact solution that satisfies the following form of (37):

$$\frac{\partial u}{\partial t} - \varepsilon\, \Delta u + \frac{\partial u}{\partial x} = f, \qquad \text{in } \Omega_T. \tag{39}$$

Additionally, Dirichlet boundary conditions on u, the initial condition on u, and the source f are ascertained from the exact solution:

$$u(x, y, t) = e^{-l\,t}\left(e^{\lambda_1\, x} - e^{\lambda_2\, x}\right) + \cos(\pi\, y)\,\frac{e^{\delta_2\, x} - e^{\delta_1\, x}}{e^{-\delta_2} - e^{-\delta_1}}, \tag{40}$$

where $l = 2$, and:

$$\begin{aligned} \lambda_{1,2} &= \frac{-1 \pm \sqrt{1 - 4\,\varepsilon\, l}}{-2\,\varepsilon}, \\ \delta_{1,2} &= \frac{1 \pm \sqrt{1 + 4\,\pi^2\,\varepsilon^2}}{2\,\varepsilon}, \end{aligned} \tag{41}$$

The problem domain $\Omega_T = (-1, 0) \times (-0.5, 0.5) \times (0, 0.5)$. For these studies we consider the moderately convection dominated case of (39) with $\varepsilon = 0.075$.

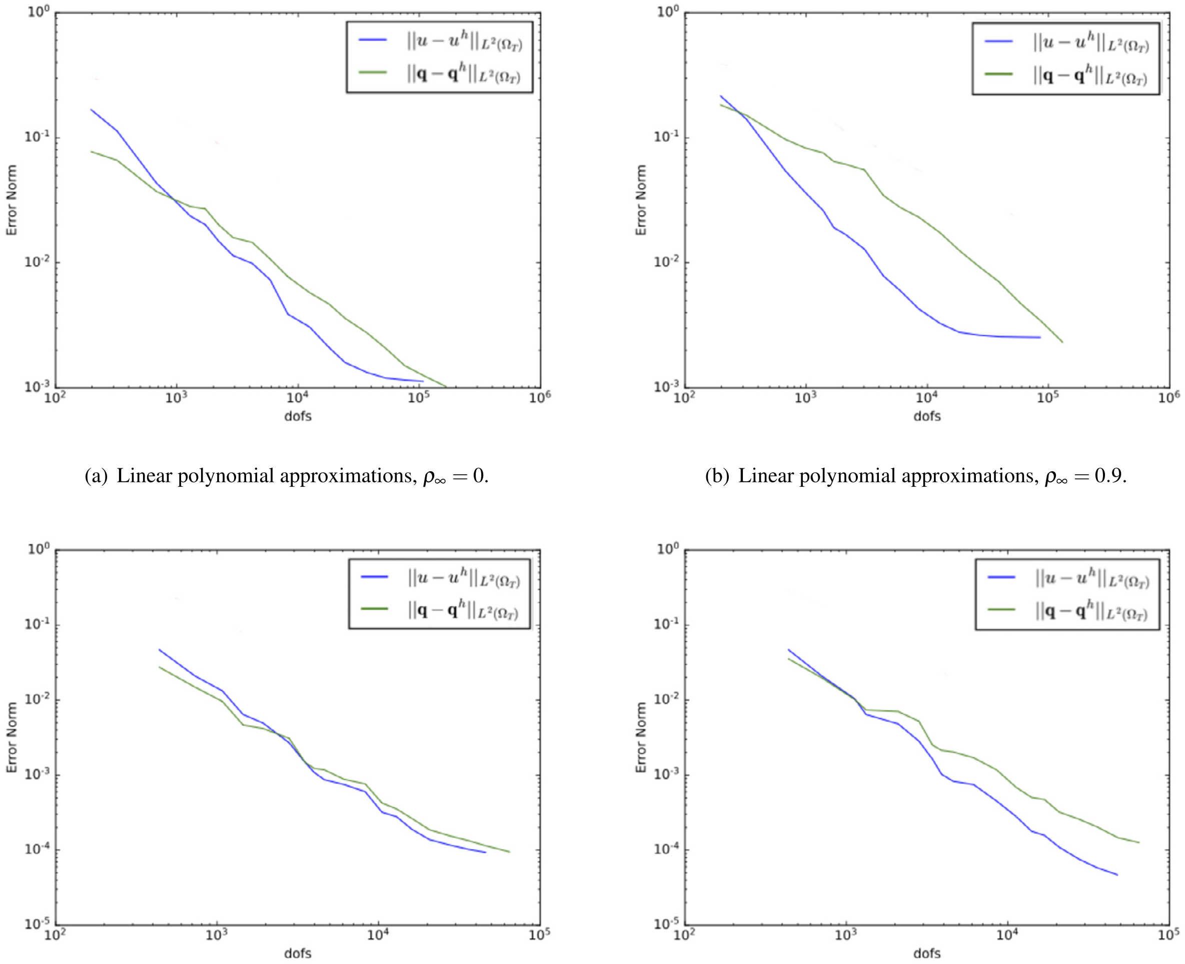
In Fig. 2 the convergence plots for linear and quadratic polynomial degrees for the space-time approach are shown. In Fig. 2, we plot error norms versus the number of degrees of freedom N, which increases at $\mathcal{O}(h^{-2})$, i.e., the h-convergence rates of the FE approximations can be extracted from these by a simple adjustment. For the case of $\|u - u^h\|_{L^2(\Omega_T)}$, we get $\mathcal{O}(N^{-1}) = \mathcal{O}(h^2) = \mathcal{O}(h^{p+1})$ order of convergence. The observed rates for $\|\mathbf{q} - \mathbf{q}^h\|_{L^2(\Omega_T)}$ are slightly lower, whereas the energy error converges at the expected rates of $\mathcal{O}(h^p)$. In error bounds for the AVS-FE method applied to a second order PDE, see, e.g., [43], it is only guaranteed that the energy norm (9) and the error in the norm on $U(\Omega_T)$ converges at $\mathcal{O}(h^p)$.

Analogously, in Fig. 3, the convergence plots for generalized-α are presented to study the convergence of the method at the final time $T = 0.5s$ with time step of $\tau = 10^{-3}$. The observed rates of convergence in Fig. 3 are the optimal rates expected from the polynomial approximations employed. Note that the L^2 errors in the base variable u become flat near the end of the refinement process as the temporal discretization error becomes dominant. Comparison of the results in Figs. 2 and 3 for the two methods reveal that the number of degrees of freedom is significantly larger for the space-time approach.

4.2.1. $H - (div)$ conforming basis functions

To complete our numerical verifications we consider the generalized-α system (28) and use Raviart-Thomas basis functions for the flux q^h. Following the known results from e.g., [33], the Raviart-Thomas functions are of order $p - 1$, where p is the order of the approximations for u^h. We also use discontinuous Raviart-Thomas bases for the vector valued error representation function of order p and the scalar valued function of the same order p.

We again consider the same Eriksson-Johnson problem with $T_{final} = 1.0s$, set $\varepsilon = 1 \times 10^{-3}$, $p = 2$, $\rho_\infty = 0.9$ and perform both uniform and adaptive mesh refinements. In the adaptive refinements and the Dörfler marking strategy, we pick the parameter $\theta = 0.5$. In Fig. 4, we present the corresponding convergence histories. Clearly, for the strongly convection-dominated case considered, the uniform refinements are not an optimal choice. However, the adaptive refinement scheme performs significantly better and is able to reduce the considered errors approximately two orders of magnitude.



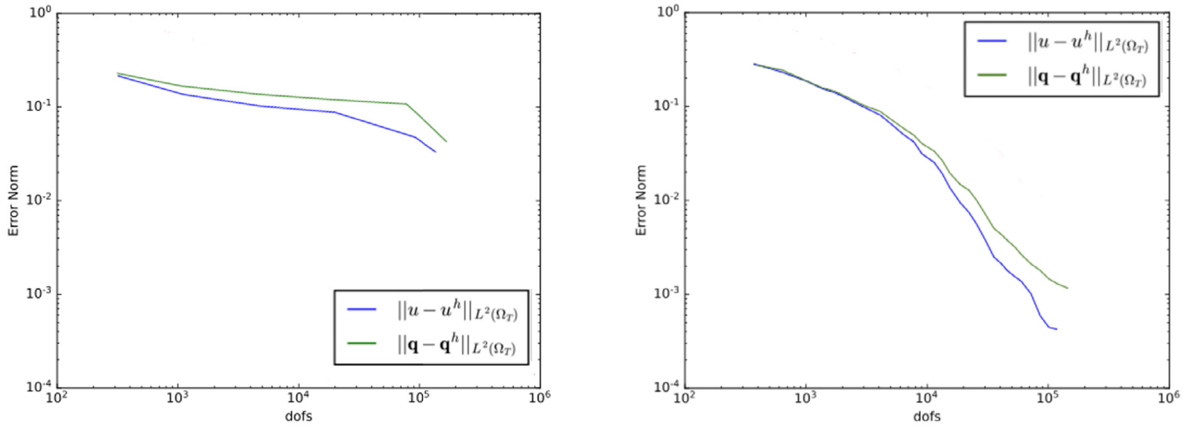
(a) Linear polynomial approximations, $\rho_\infty = 0$.

(b) Linear polynomial approximations, $\rho_\infty = 0.9$.

(c) Quadratic polynomial approximations, $\rho_\infty = 0$.

(d) Quadratic polynomial approximations, $\rho_\infty = 0.9$.

Fig. 3. Convergence study of the solution obtained using the generalized-α method for time discretization at final time $T = 0.5\,s$.



(a) Uniform mesh refinements.

(b) Adaptive mesh refinements.

Fig. 4. Convergence study of the solution obtained using the generalized-α approach $T = 1.0$ using fully conforming FE basis functions.

4.3. Comparison between space-time and time stepping

As the space-time and time-stepping methods are fundamentally different, a comparison between the two methods is not trivial. Comparison of accuracy of the two methods is not straightforward to compare, as the errors reported in Fig. 2 are global for the full space-time domain and the errors in Fig. 3 are at the final time step. Furthermore, the computational cost is distributed differently in the two methods. In this subsection, we perform heuristic comparisons between the two methods by considering the results at the final time of simulation.

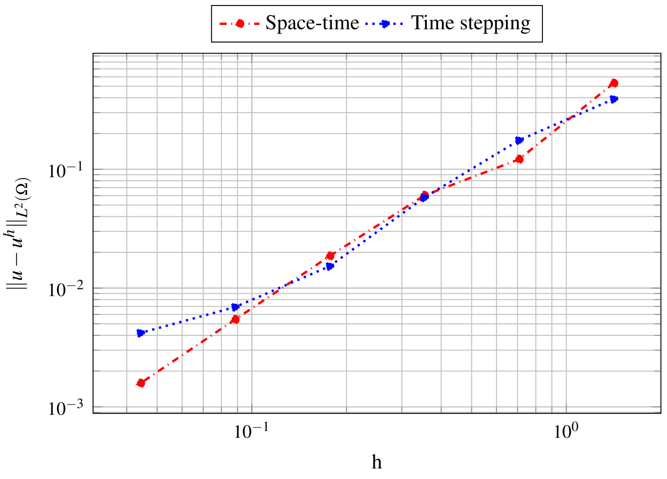


Fig. 5. Convergence at the final time $T = 0.5$ for increasingly fine uniform meshes.

4.3.1. *Uniform refinement*

First, we compare the error at the final time $T = 0.5$ for the case of $p = 1$ with the problem setup from Section 4.2. In the space-time approach the initial mesh consists of six uniform space-time tetrahedrons whereas in the generalized -α method it consists of two triangular elements. In the generalized -α method we set $\rho_\infty = 0.9$ and perform 5 time steps. We perform uniform refinements to the initial mesh and compute the errors in the space-time approach at the final time step and plot them alongside the final time error from generalized-α against the (2 dimensional) element size h at the final time in Fig. 5. It is interesting to observe that the errors in both methods shown in this figure are nearly identical. In terms of computational time, the space-time approach required 75 seconds whereas the generalized -α method took 49 seconds. In both cases the experiments were performed on a 2022 MacBook pro with the Apple M2 chip.

4.3.2. *Adaptive refinement*

As a second comparison, we compare the final time meshes obtained using adaptive mesh refinements. We consider the following exact solution:

$$u(x,y,t) = (t+0.1)(1+\tanh(0.5(0.1-|0.5-\sqrt{x^2+y^2}|))), \tag{42}$$

from which we ascertain boundary and initial conditions, as well as the source f. The problem domain $\Omega_T = (0,1)\times(0,1)\times(0,0.25)$ and we consider the moderately convection dominated case of (37) with $\varepsilon = 0.1$ and $\mathbf{b} = \{-y, x\}$. In the space-time approach the initial mesh again consists of six uniform space-time tetrahedrons whereas in the generalized -α method it consists of two triangular elements, in both cases we select $p = 1$. In the generalized -α method we set $\rho_\infty = 0.9$ and perform 250 time steps. We perform adaptive mesh refinements to the initial meshes using the Dörfler marking strategy, and select the parameter $\theta = 0.3$. In the space-time case we perform 24 adaptation cycles, and in the generalized-α case 30 adaptation cycles. In Fig. 6 we present the traces of the meshes produced in the refinement process at $T = 0.25s$. The final meshes in both cases are similar in that the refinements are focused around the internal layer. Both experiments were performed on the same desktop computer with an Intel Xeon Processor type W-2245 from 2020, where the space-time case required 10833 seconds, and in the case of the generalized-α 2430 seconds.

4.4. *Shock problem*

As a final numerical verification, we present a consideration of (37) in which the solution behaves as two shocks traveling through the space-time domain while rotating about the origin. Furthermore, the choices we make for the problem parameters are such that the interface of the shock is skewed and rotates in the space-time domain as $t \to T_{final}$. Thus, we have the following choices:

$$\begin{aligned} \mathbf{b} &= \{-x+2y, 0\}^T, \\ u_0 &= 0, \\ \varepsilon &= 10^{-3}, \\ u_0 &= 0, \\ f &= -2x\varepsilon + x(1-y^2). \\ T_{final} &= 2.50s \\ \Omega &= (-1,1)\times(-1,1) \end{aligned} \tag{43}$$

For this particular problem, we present the time slice approach in which we perform mesh adaptations between each slice and we apply linear polynomial approximations for the trial functions. Experience has shown that the slice containing the initial condition is critical to the proper resolution of the space-time process. Thus, we consider the case of three space-time slices, the first from $0s$ to $0.2s$ and the final two of equal size from $0.2s$ to $2.5s$. In Figs. 7, 8, and 9 we present the AVS-FE solution for the base variable at different time steps. As expected, two shock-waves originate at the boundaries of $x = \pm 1$, and as time progress, the two waves approach the center of the domain while rotating. The adaptively refined meshes shown in Figs. 7(b), 8(b), and 9(b) (the final times of each slice) show that the mesh refinements are focused at the interfaces of the shocks, further indicating the applicability of the built-in error indicators.

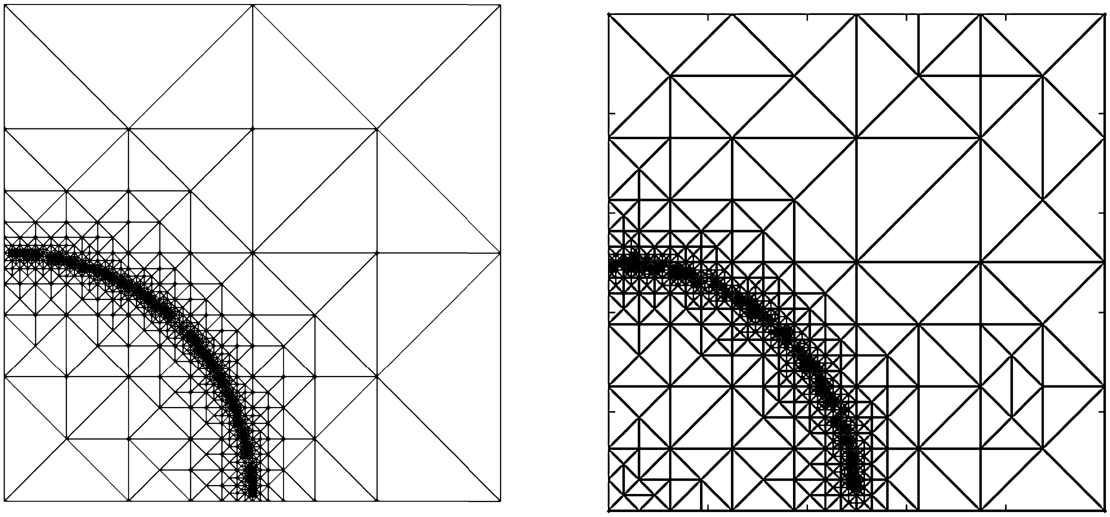
5. Conclusions

The AVS-FE method is a Petrov-Galerkin method which uses classical continuous FE trial basis functions, while the test space consists of functions that are discontinuous across element edges. This broken topology in the test space allows us to employ the DPG philosophy and introduce an equivalent saddle point problem which we implement using high level FE solvers. We have introduced two distinct approaches to transient problems using the AVS-FE method. First, we take a space-time approach in which the entire space-time domain is discretized using finite elements, and second, using the method of lines to discretize the spatial domain independently. Then, using a time-marching method, we obtain a fully discrete system.

The space-time method allows us to exploit the unconditional stability of the AVS-FE method and perform a single global solve governing the FE approximation. As the AVS-FE approximations computed from the saddle point system (26) come with built-in error indicators, we are capable of utilizing mesh adaptive strategies in space and time. In an effort to control the computational cost of the space-time approach in solving the global system of equations, we consider a time slice approach. Here, the space-time domain is partitioned into finite sized space-time slices on which we employ the AVS-FE method. The advantage here is that the size of the global system is reduced and we are able to employ mesh adaptive strategies on each slice.

The method of lines, in which we use the AVS-FE method for the spatial discretization and a generalized-α method to derive a fully-discretized system. In this case, the discrete stability in the temporal domain is ensured by the generalized-α method leading to highly efficient stable FE computations. We show that the AVS-FE method uses a corresponding norm as a function of the time-step. Another distinguishing feature of this method is that due to the influence of the initial data on the accuracy of the solution, we find a stable approximation for $\frac{\partial u}{\partial t}$ at the initial time. Accordingly, at each time step, one is required to solve a system with a smaller number of degrees of freedom in comparison with the space-time approach.

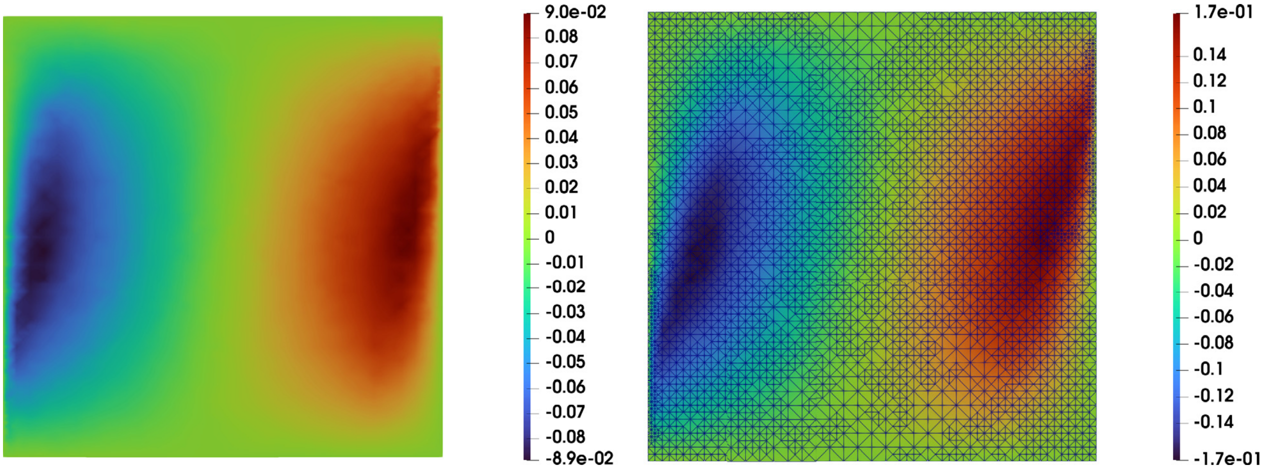
Numerical verifications for several cases of the transient convection-diffusion IBVP show that both methods exhibit optimal asymptotic convergence behavior as well as similar norms of the numerical approximation error. For degrees of approximation above 2, the space-time approach becomes more accurate as it is not limited to the second-order



(a) Space-time.

(b) Generalized -α.

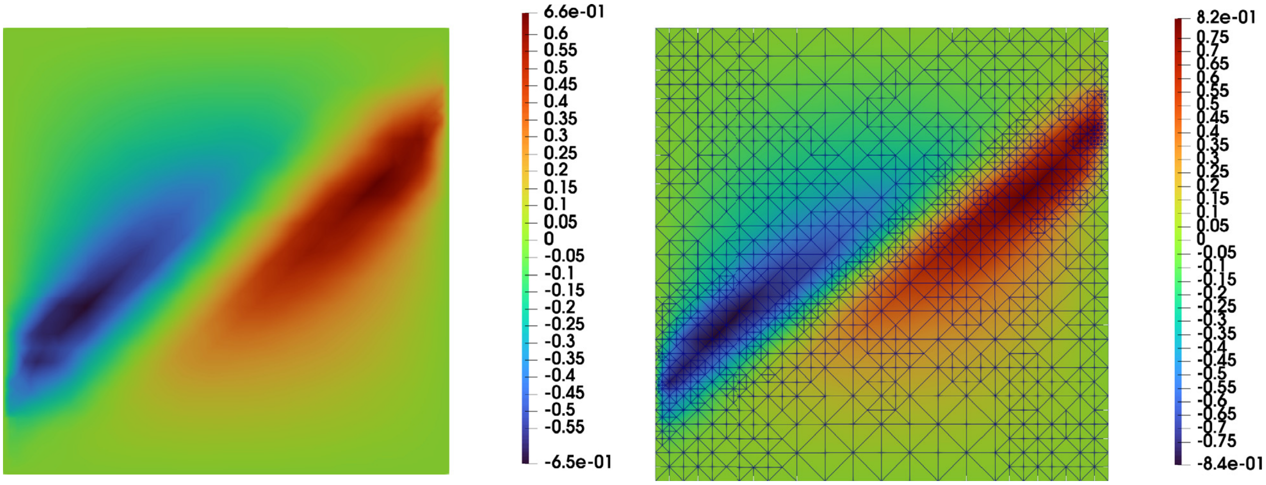
Fig. 6. Comparison of the adaptive refined meshes at the final time.



(a) Solution u^h at $t = 0.1s$.

(b) Solution u^h at $t = 0.2s$ with final adapted mesh.

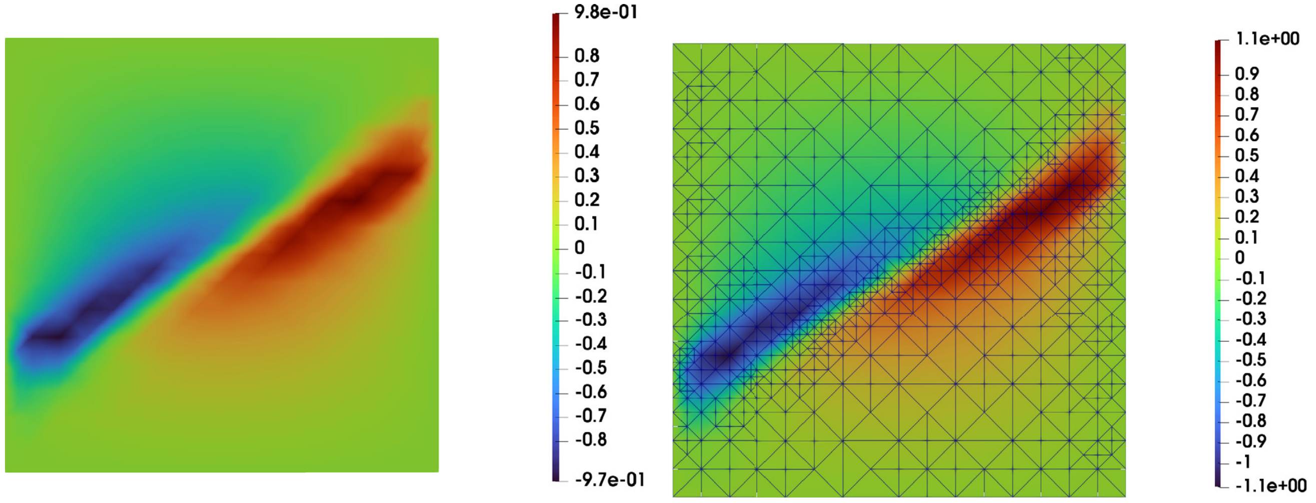
Fig. 7. AVS-FE approximations of the shock problem, i.e., (37) with parameters from (43).



(a) Solution u^h at $t = 1.0s$.

(b) Solution u^h at $t = 1.35s$ with final adapted mesh.

Fig. 8. AVS-FE approximations of the shock problem, i.e., (37) with parameters from (43).



(a) Solution u^h at $t = 2.0s$.

(b) Solution u^h at $t = 2.5s$ with final adapted mesh.

Fig. 9. AVS-FE approximations of the shock problem, i.e., (37) with parameters from (43).

accuracy of the generalized-α method. However, we do not advocate one method over the other but we point out these differences for potential users as their available computational resources will likely dictate which approach to use. For both cases, we present additional numerical verifiactions highlighting the adaptive mesh refinement capabilities. In future efforts, we expect to pursue alternative error estimators and indicators as well as the AVS-FE approximation of challenging transient physical phenomena. The use of basis functions that are of higher order regularity, e.g., as in [44] is another potential direction of future research efforts.

Data availability

Data will be made available on request.

Acknowledgements

Authors Valseth and Dawson have been supported by the United States National Science Foundation - NSF PREEVENTS Track 2 Program, under NSF Grant Number 1855047. Authors Valseth and Romkes have been supported by the United States National Science Foundation - NSF CBET Program, under NSF Grant Number 1805550. The authors gratefully acknowledge the assistance of Austin Kaul of the Department of Mechanical Engineering at South Dakota School of Mines and Technology in performing the numerical verifications of Section 4.2 and 4.4. Finally, the authors would also like to gratefully acknowledge the use of the "ADCIRC" and "DMS21031" allocations at the Texas Advanced Computing Center at the University of Texas at Austin.

References

[1] V.M. Calo, A. Romkes, E. Valseth, Automatic variationally stable analysis for FE computations: an introduction, in: G. Barrenechea, J. Mackenzie (Eds.), Boundary and Interior Layers, Computational and Asymptotic Methods BAIL 2018, Springer, 2020, pp. 19–43.

[2] L. Demkowicz, J. Gopalakrishnan, A class of discontinuous Petrov-Galerkin methods. Part I: the transport equation, Comput. Methods Appl. Mech. Eng. 199 (23) (2010) 1558–1572.

[3] C. Carstensen, L. Demkowicz, J. Gopalakrishnan, A posteriori error control for DPG methods, SIAM J. Numer. Anal. 52 (3) (2014) 1335–1353.

[4] L. Demkowicz, J. Gopalakrishnan, Analysis of the DPG method for the Poisson equation, SIAM J. Numer. Anal. 49 (5) (2011) 1788–1809.

[5] L. Demkowicz, J. Gopalakrishnan, A class of discontinuous Petrov-Galerkin methods. II. Optimal test functions, Numer. Methods Partial Differ. Equ. 27 (1) (2011) 70–105.

[6] L. Demkowicz, J. Gopalakrishnan, A class of discontinuous Petrov-Galerkin methods. Part III: adaptivity, Appl. Numer. Math. 62 (4) (2012) 396–427.

[7] P. Behnoudfar, Q. Deng, V.M. Calo, High-order generalized-alpha method, Appl. Eng. Sci. 4 (2020) 100021.

[8] P. Behnoudfar, Q. Deng, V.M. Calo, Higher-order generalized-α methods for hyperbolic problems, Comput. Methods Appl. Mech. Eng. 378 (2021) 113725.

[9] J. Chung, G. Hulbert, A time integration algorithm for structural dynamics with improved numerical dissipation: the generalized-α method, J. Appl. Mech. 60 (2) (1993) 371–375.

[10] T.J.R. Hughes, J.R. Stewart, A space-time formulation for multiscale phenomena, J. Comput. Appl. Math. 74 (1996) 217–229.

[11] T.J. Hughes, G.M. Hulbert, Space-time finite element methods for elastodynamics: formulations and error estimates, Comput. Methods Appl. Mech. Eng. 66 (3) (1988) 339–363.

[12] A.K. Aziz, P. Monk, Continuous finite elements in space and time for the heat equation, Math. Comput. 52 (186) (1989) 255–274.

[13] E. Valseth, A. Romkes, A.R. Kaul, A stable FE method for the space-time solution of the Cahn-Hilliard equation, J. Comput. Phys. 441 (2021) 110426.

[14] E. Valseth, C. Dawson, An unconditionally stable space–time FE method for the Korteweg–de Vries equation, Comput. Methods Appl. Mech. Eng. 371 (2020) 113297, https://doi.org/10.1016/j.cma.2020.113297.

[15] T.E. Ellis, L. Demkowicz, J. Chan, R.D. Moser, Space-time DPG: Designing a method for massively parallel CFD, ICES report, The Institute for Computational Engineering and Sciences, The University of Texas at Austin, 2014, pp. 14–32.

[16] T. Ellis, J. Chan, L. Demkowicz, Robust DPG methods for transient convection-diffusion, in: Building Bridges: Connections and Challenges in Modern Approaches to Numerical Partial Differential Equations, Springer, 2016, pp. 179–203.

[17] N.V. Roberts, L. Demkowicz, R. Moser, A discontinuous Petrov–Galerkin methodology for adaptive solutions to the incompressible Navier–Stokes equations, J. Comput. Phys. 301 (2015) 456–483.

[18] N.V. Roberts, Camellia: a software framework for discontinuous Petrov–Galerkin methods, Comput. Math. Appl. 68 (11) (2014) 1581–1604.

[19] J. Muñoz-Matute, D. Pardo, L. Demkowicz, A DPG-based time-marching scheme for linear hyperbolic problems, Comput. Methods Appl. Mech. Eng. 373 (2021) 113539.

[20] J. Muñoz-Matute, L. Demkowicz, D. Pardo, Error representation of the time-marching DPG scheme, Comput. Methods Appl. Mech. Eng. 391 (2022) 114480.

[21] K.E. Jansen, C.H. Whiting, G.M. Hulbert, A generalized-α method for integrating the filtered Navier–Stokes equations with a stabilized finite element method, Comput. Methods Appl. Mech. Eng. 190 (3–4) (2000) 305–319.

[22] H.M. Hilber, T.J. Hughes, R.L. Taylor, Improved numerical dissipation for time integration algorithms in structural dynamics, Earthq. Eng. Struct. Dyn. 5 (3) (1977) 283–292.

[23] W. Wood, M. Bossak, O. Zienkiewicz, An alpha modification of Newmark's method, Int. J. Numer. Methods Eng. 15 (10) (1980) 1562–1566.

[24] P.B. Bochev, M.D. Gunzburger, Least-Squares Finite Element Methods, vol. 166, Springer Science & Business Media, 2009.

[25] B. Keith, S. Petrides, F. Fuentes, L. Demkowicz, Discrete least-squares finite element methods, Comput. Methods Appl. Mech. Eng. 327 (2017) 226–255.

[26] M.S. Alnæs, J. Blechta, J. Hake, A. Johansson, B. Kehlet, A. Logg, C. Richardson, J. Ring, M.E. Rognes, G.N. Wells, The FEniCS project version 1.5, Arch. Numer. Softw. 3 (100) (2015) 9–23.

[27] F. Rathgeber, D.A. Ham, L. Mitchell, M. Lange, F. Luporini, A.T. McRae, G.-T. Bercea, G.R. Markall, P.H. Kelly, Firedrake: automating the finite element method by composing abstractions, ACM Trans. Math. Softw. 43 (3) (2017) 24.

[28] L.F. Demkowicz, J. Gopalakrishnan, An overview of the discontinuous Petrov-Galerkin method, in: Recent Developments in Discontinuous Galerkin Finite Element Methods for Partial Differential Equations, Springer, 2014, pp. 149–180.

[29] F. Fuentes, B. Keith, L. Demkowicz, P. Le Tallec, Coupled variational formulations of linear elasticity and the DPG methodology, J. Comput. Phys. 348 (2017) 715–731.

[30] V.M. Calo, A. Ern, I. Muga, S. Rojas, An adaptive stabilized conforming finite element method via residual minimization on dual discontinuous Galerkin norms, Comput. Methods Appl. Mech. Eng. 363 (2020) 112891.

[31] S. Nagaraj, S. Petrides, L.F. Demkowicz, Construction of DPG fortin operators for second order problems, Comput. Math. Appl. 74 (8) (2017) 1964–1980.

[32] L. Demkowicz, P. Zanotti, Construction of DPG Fortin operators revisited, Comput. Math. Appl. 80 (11) (2020) 2261–2271.

[33] F. Brezzi, M. Fortin, Mixed and Hybrid Finite Element Methods, vol. 15, Springer-Verlag, 1991.

[34] F. Brezzi, On the existence, uniqueness and approximation of saddle-point problems arising from Lagrangian multipliers, Publ. Math. Inform. Rennes (1974) 1–26.

[35] P. Behnoudfar, V.M. Calo, Q. Deng, P.D. Minev, A variationally separable splitting for the generalized-α method for parabolic equations, Int. J. Numer. Methods Eng. 121 (5) (2020) 828–841.

[36] P. Behnoudfar, G. Loli, A. Reali, G. Sangalli, V.M. Calo, Explicit high-order generalized-α methods for isogeometric analysis of structural dynamics, Comput. Methods Appl. Mech. Eng. 389 (2022) 114344.

[37] D.A. Di Pietro, A. Ern, Mathematical Aspects of Discontinuous Galerkin Methods, vol. 69, Springer Science & Business Media, 2011.

[38] T.E. Ellis, Space-time discontinuous Petrov-Galerkin finite elements for transient fluid mechanics, Ph.D. thesis, The University of Texas at Austin, 2016.

[39] W. Dörfler, A convergent adaptive algorithm for Poisson's equation, SIAM J. Numer. Anal. 33 (3) (1996) 1106–1124.

[40] P.R. Amestoy, A. Guermouche, J.-Y. L'Excellent, S. Pralet, Hybrid scheduling for the parallel solution of linear systems, Parallel Comput. 32 (2) (2006) 136–156.

[41] P.R. Amestoy, I.S. Duff, J.-Y. L'Excellent, J. Koster, A fully asynchronous multifrontal solver using distributed dynamic scheduling, SIAM J. Matrix Anal. Appl. 23 (1) (2001) 15–41.

[42] K. Eriksson, C. Johnson, Adaptive streamline diffusion finite element methods for stationary convection-diffusion problems, Math. Comput. 60 (201) (1993) 167–188.

[43] E. Valseth, A. Romkes, A.R. Kaul, C. Dawson, A stable mixed finite element method for nearly incompressible linear elastostatics, Int. J. Numer. Methods Eng. 122 (17) (2021) 4709–4729.

[44] M. Łoś, J. Munoz-Matute, I. Muga, M. Paszyński, Isogeometric residual minimization method (iGRM) with direction splitting for non-stationary advection–diffusion problems, Comput. Math. Appl. 79 (2) (2020) 213–229.