Journal of Scientific Computing (2024) 101:11
https://doi.org/10.1007/s10915-024-02650-x

Error Estimates and Adaptivity of the Space-Time Discontinuous Galerkin Method for Solving the Richards Equation

Vít Dolejší[1] · **Hyun-Geun Shin**[1] · **Miloslav Vlasák**[2]

Received: 15 January 2024 / Revised: 28 May 2024 / Accepted: 28 July 2024 /
Published online: 20 August 2024

Abstract
We present a higher-order space-time adaptive method for the numerical solution of the Richards equation that describes a flow motion through variably saturated media. The discretization is based on the space-time discontinuous Galerkin method, which provides high stability and accuracy and can naturally handle varying meshes. We derive reliable and efficient a posteriori error estimates in the residual-based norm. The estimates use well-balanced spatial and temporal flux reconstructions which are constructed locally over space-time elements or space-time patches. The accuracy of the estimates is verified by numerical experiments. Moreover, we develop the hp-adaptive method and demonstrate its efficiency and usefulness on a practically relevant example.

Keywords Space-time discontinuous Galerkin method · Richards equation · A posteriori error estimate · hp-mesh adaptation

Mathematics Subject Classification 65M60 · 65M15 · 65M50

1 Introduction

Fluid flows in variably saturated porous media are usually described by the Richards equation [33], which is expressed in the form

$$\partial_t \vartheta(\psi) - \nabla \cdot \big(\mathbf{K}(\theta(\psi))(\nabla\psi + \nabla z)\big) = g, \tag{1}$$

✉ Vít Dolejší
dolejsi@karlin.mff.cuni.cz

Hyun-Geun Shin
shin@karlin.mff.cuni.cz

Miloslav Vlasák
vlasamil@cvut.cz

[1] Faculty of Mathematics and Physics, Charles University, Sokolovská 83, Prague, Czech Republic

[2] Faculty of Civil Engineering, Czech Technical University, Thakurova 7, Prague 6 166 29, Czech Republic

where ∂_t denotes the derivative with respect to time, $\nabla\cdot$ and ∇ are the divergence and gradient operators, respectively, ψ is the sought pressure head (= normalized pressure), z is the vertical coordinate, θ is the water content function, $\mathbf{K}$ is the hydraulic conductivity tensor and g is the source term. In addition, the active pore volume ϑ is related to θ by the following relation

$$\vartheta(\psi) := \theta(\psi) + \frac{S_s}{\theta_s}\int_{-\infty}^{\psi} \theta(s)\,\mathrm{d}s, \tag{2}$$

where $S_s, \theta_s \geq 0$ are material parameters. The hydraulic conductivity satisfies $\mathbf{K}(\psi) = \mathbf{K}_s\mathcal{K}_r(\psi)$, where $\mathbf{K}_s$ is the saturated conductivity tensor, and $\mathcal{K}_r \in [0, 1]$ is the relative saturation. The functions θ and $\mathcal{K}_r$ are given by constitutive relations, e.g., by van Genuchten's law [27] and by Mualem's law [31], respectively.

The Richards equation belongs to the nonlinear parabolic problems, and it can degenerate, in particular $\mathbf{K} \to 0$ or $\frac{\mathrm{d}\vartheta}{\mathrm{d}\psi} \to 0$. Due to the degeneracy, the numerical solution is challenging, and various techniques have been developed for its solution in the last decades, see [25] for a survey.

In [14], we presented the adaptive space-time discontinuous Galerkin (STDG) method for the numerical solution of (1). This technique is based on a piecewise polynomial discontinuous approximation with respect to both the spatial and temporal coordinates. The resulting scheme is sufficiently stable, provides high accuracy, and is suitable for the hp-mesh adaptation. This is an important property, since the weak solution of the Richards equation is (only) piecewise regular and exhibits singularities along the material interfaces and the unsaturated/saturated zone (when $\psi \approx 0$). Therefore, an adaptive method that allows different meshes at different time levels, can achieve an accurate approximation with a relatively small number of degrees of freedom.

The numerical experiments presented in [14] showed the potential of the adaptive STDG method. However, the mesh adaptation used is based on interpolation error estimates that do not guarantee an upper error bound. The aim of this work is to overcome this bottleneck, derive a posteriori error estimates, and use them in the hp-mesh adaptation framework.

A posteriori error estimates for the numerical solution of the Richards equation have been treated in many papers for different numerical methods. We mention the finite volume framework with multistep time discretization in [5], the mixed finite element method in [6], the two-point finite volume discretization in [8], the lowest-order discretization on polytopal meshes in [38], finite element techniques in [30] and the references cited therein.

Guaranteed error estimates without unknown constants are usually obtained by measuring the error in a dual norm of the residual. Introducing reconstructed fluxes from the space $H^1(\mathrm{div}, \Omega)$, the upper bound can then be obtained directly. In [18], we developed this approach to the higher-order STDG method for nonlinear parabolic problems, where the temporal discontinuities were treated by temporal flux reconstructions considering the time jumps.

In this paper, we extend the approach [18] to the Richards equation (1). Although the definition of the temporal and spatial flux reconstructions as well as the derivation of the upper bounds is straightforward, the proof of the lower bound (efficiency) is rather tricky since the term $\theta(\psi)$ in the time derivative is not a polynomial function for a polynomial ψ. In contrary to [18], the proof of efficiency requires the additional oscillatory data terms. We construct spatial fluxes by solving local Neumann problems defined on space-time patches that generalize the approach from [22]. Moreover, we provide numerical experiments verifying derived error estimates. Compared to [18], the resulting effectivity indices are much closer to one. This is the first novelty of this paper.

Secondly, we deal with the errors arising due to iterative solution of nonlinear algebraic systems. We introduce a cheap stopping criterion for iterative solvers and justify it by numerical experiments. Thirdly, we introduce a space-time adaptive algorithm that employs the anisotropic hp-mesh adaptation technique [15]. The algorithm admits local adaptation of size and shape of mesh elements and the local adaptation of degrees of polynomial approximation with respect to space. However, the size of the time step can vary globally, and the degree of polynomial approximation with respect to time is fixed. Using the equidistribution principle, the algorithm gives an approximate solution with the error estimate under the given tolerance. The performance of the adaptive algorithm is demonstrated numerically, including a practically relevant example.

The rest of the paper is organized as follows. In Sect. 2, we introduce the problem considered, its STDG discretization is briefly described in Sect. 3. The main theoretical results are derived in Sect. 4, where the upper and lower bounds are proved. Two possible spatial reconstructions are discussed in Sect. 5 together with the stopping criteria of iterative solvers. The numerical verification of the error estimates is given in Sect. 6. Furthermore, we present the resulting hp-mesh adaptation algorithm in Sect. 7 and demonstrate its performance by numerical examples. Finally, we conclude with some remarks in Sect. 8.

2 Problem Formulation

Let $\Omega \subset \mathbb{R}^d$ $(d = 2, 3)$ be the domain occupied by a porous medium and $T > 0$ the physical time to be reached. For simplicity, we assume that Ω is polygonal. By $\Gamma := \partial\Omega$, we denote the boundary of Ω which consists of two disjoint parts: the Dirichlet boundary Γ_{D} and the Neumann boundary Γ_{N}. We write the Richards equation (1) in a different form, which is more suitable for the analysis. We seek a function $u = u(x, t) : \Omega \times (0, T) \to \mathbb{R}$, which represents a hydraulic head (with the physical unit L). The quantity u is related to the pressure head ψ by $u = \psi + z$. The Richards equation (1) reads

$$\begin{aligned} &\partial_t \vartheta(u) - \nabla \cdot (\mathbf{K}(u)\nabla u) = g \quad \text{in } \Omega \times (0, T) \\ &u = u_D \text{ on } \Gamma_{\mathrm{D}} \times (0, T) \\ &\mathbf{K}(u)\nabla u \cdot n = g_N \text{ on } \Gamma_{\mathrm{N}} \times (0, T), \\ &u(x, 0) = u_0 \text{ in } \Omega, \end{aligned} \tag{3}$$

where $g : \Omega \times (0, T) \to \mathbb{R}$ represents a source term if g is positive or a sink term if g is negative, $\vartheta : \mathbb{R} \to \mathbb{R}$ denotes the dimensionless active pore volume, and $\mathbf{K} : \mathbb{R} \to \mathbb{R}^{d\times d}$ is the hydraulic conductivity with the physical unit $\mathrm{L} \cdot \mathrm{T}^{-1}$ (L = length, T = time). Moreover, u_D is a trace of a function $u^* \in L^2(0, T; H^1(\Omega))$ on $\Gamma_{\mathrm{D}} \times (0, T)$, $g_N \in L^2(0, T; L^2(\Gamma_{\mathrm{N}}))$ and $u_0 \in L^2(\Omega)$. We note that with respect to (1), we should write $\vartheta = \vartheta(u - z)$ and $\mathbf{K} = \mathbf{K}(\theta(u - z))$, however, we skip this notation for simplicity. We assume that the function $\vartheta(u)$ is non-negative, non-decreasing and Lipschitz continuous. Moreover, the tensor $\mathbf{K}(u)$ is symmetric, positively semi-definite, and Lipschitz continuous.

In order to introduce the weak solution, we set $H(\mathrm{div}, \Omega) = \{v \in L^2(\Omega)^d : \nabla \cdot v \in L^2(\Omega)\}$ and define the spaces

$$\begin{aligned} &X = L^2(0, T, H^1(\Omega)), && V = \{v \in X : v|_{\Gamma_{\mathrm{D}}} = 0\}, \\ &Y = \{v \in X : \vartheta'(v) \in L^2(0, T, L^2(\Omega))\}, && Y^0 = \{v \in Y : v(0) = u_0\}, \end{aligned} \tag{4}$$

where $\vartheta'(u) = \partial_t \vartheta(u) = \frac{\mathrm{d}\vartheta}{\mathrm{d}u} \partial_t u$ denotes the time derivative (in the weak sense). Obviously, if $v \in Y$ then $\vartheta(v) \in C([0,T], L^2(\Omega))$. In order to shorten the notation, we set the physical flux

$$\sigma(u, \nabla u) := \mathbf{K}(u)\nabla u, \qquad u \in X. \tag{5}$$

Definition 1 We say that $u \in Y$ is the *weak solution* of (3) if $u - u^* \in V$ and

$$\int_0^T \left(\left(\vartheta'(u), v\right)_\Omega + (\sigma(u, \nabla u), \nabla v)_\Omega - (g, v)_\Omega - \left(g_N, v\right)_{\Gamma_\mathrm{N}} \right) \mathrm{d}t = 0 \quad \forall v \in V, \tag{6}$$

where $\left(u, v\right)_\Omega := \int_\Omega uv \,\mathrm{d}x$ and $\left(u, v\right)_{\Gamma_\mathrm{N}} := \int_{\Gamma_\mathrm{N}} uv \,\mathrm{d}S$.

The existence and uniqueness of the Richards equation is studied in [2], see also the later works [3, 28].

3 Space-time discretization

We briefly describe the discretization of (6) by the *space-time discontinuous Galerkin* (STDG) method, for more details, see [13, 14]. Let $0 = t_0 < t_1 < \ldots < t_r = T$ be a partition of the time interval $(0, T)$ and set $I_m = (t_{m-1}, t_m)$ and $\tau_m = t_m - t_{m-1}$. For each $m = 0, \ldots, r$, we consider a simplicial mesh $\mathcal{T}_h^m$ covering $\overline{\Omega}$. For simplicity, we assume that $\mathcal{T}_h^m$, $m = 0, \ldots, r$ are conforming, i.e., neighbouring elements share an entire edge or face. However, this assumption can be relaxed by the technique from [12].

For each element $K \in \mathcal{T}_h^m$, we denote by ∂K its boundary, n_K its unit outer normal and $h_K = \mathrm{diam}(K)$ its diameter. In order to shorten the notation, we write $\partial K_N := \partial K \cap \Gamma_\mathrm{N}$. By the generic symbol γ, we denote an edge ($d = 2$) or a face ($d = 3$) of $K \in \mathcal{T}_h^m$ and h_γ denotes its diameter. In the following, we speak only about edges, but we mean faces for $d = 3$. We assume that

- $\mathcal{T}_h^m$, $m = 0, \ldots, r$ are *shape regular*, i.e., $h_K/\rho_K \le C$ for all $K \in \mathcal{T}_h$, where ρ_K is the radius of the largest d-dimensional ball inscribed in K and constant C does not depend on $\mathcal{T}_h^m$ for $h \in (0, h_0)$, $m = 0, \ldots, r$.
- $\mathcal{T}_h^m$, $m = 0, \ldots, r$ are *locally quasi-uniform*, i.e., $h_K \le C h_{K'}$ for any pair of two neighbouring elements K and K', where the constant C does not depend on $h \in (0, h_0)$, $m = 0, \ldots, r$.

Let $p_K \ge 1$ be an integer denoting the degree of polynomial approximation on $K \in \mathcal{T}_h^m$, $m = 0, \ldots, r$ and $P_{p_K}(K)$ be the corresponding space of polynomial functions on K. Let

$$S_{hp,m} = \{v \in L^2(\Omega) : v|_K \in P_{p_K}(K),\ K \in \mathcal{T}_h^m\}, \qquad m = 0, \ldots, r \tag{7}$$

denote the spaces of discontinuous piecewise polynomial functions on $\mathcal{T}_h^m$ with possibly varying polynomial approximation degrees. Furthermore, we consider the space of *space-time* discontinuous piecewise polynomial functions

$$S_{hp}^{\tau q} = \{v \in L^2(\Omega \times (0,T)) :\ v|_{I_m} \in P_q(I_m, S_{hp,m}),\ m = 1, \ldots, r\}, \tag{8}$$

where $q \ge 0$ denotes the time polynomial approximation degree and $P_q(I_m, S_{hp,m})$ is the Bochner space, i.e., $v \in P_q(I_m, S_{hp,m})$ can be written as $v(x,t) = \sum_{j=0}^q t^j v_j(x)$, $v_j \in S_{hp,m}$, $j = 0, \ldots, q$.

For $v \in S_{hp}^{\tau q}$, we define the one-sided limits and time jumps by

$$v_+^m = \lim_{t \to t_m^+} v(t), \quad m = 0, \dots, r-1, \qquad v_-^m = \lim_{t \to t_m^-} v(t), \quad m = 1, \dots, r, \tag{9}$$
$$\{v\}_m = v_+^m - v_-^m, \quad m = 1, \dots, r-1, \qquad v_-^0 = \vartheta(u_0), \qquad \{v\}_0 = v_+^0 - \vartheta(u_0),$$

where u_0 is the initial condition. In the following, we use the notation

$$(u, v)_M = \int_M u\, v \,\mathrm{d}x, \qquad (u, v)_{M,m} = \int_{M \times I_m} u\, v \,\mathrm{d}x\,\mathrm{d}t, \quad m = 1, \dots, r, \tag{10}$$

where M is either element $K \in \mathcal{T}_h^m$ or its (part of) boundary ∂K. The corresponding norms are denoted by $\|\cdot\|_M$ and $\|\cdot\|_{M,m}$, respectively. By $\sum_{K,m} = \sum_{m=1}^r \sum_{K \in \mathcal{T}_h^m}$, we denote the sum over all space-time elements $K \times I_m$, where $K \in \mathcal{T}_h^m$ and $m = 1, \dots, r$.

Moreover, we define the jumps and mean values of $v \in S_{hp,m}$ on edges $\gamma \subset \partial K$, $K \in \mathcal{T}_h^m$ by

$$[v] = \begin{cases} (v^{(+)} - v^{(-)}) n_K & \text{for } \gamma \in \Omega, \\ (v^{(+)} - u_D) n_K & \text{for } \gamma \subset \Gamma_{\mathrm{D}}, \\ 0 & \text{for } \gamma \subset \Gamma_{\mathrm{N}}, \end{cases} \qquad \langle v \rangle = \begin{cases} (v^{(+)} + v^{(-)})/2 & \text{for } \gamma \in \Omega, \\ v^{(+)} & \text{for } \gamma \subset \Gamma_{\mathrm{D}}, \\ 0 & \text{for } \gamma \subset \Gamma_{\mathrm{N}}, \end{cases} \tag{11}$$

where $v^{(+)}$ and $v^{(-)}$ denote the traces of v on ∂K from interior and exterior of K, respectively, and u_D comes from the Dirichlet boundary condition. For vector-valued $v \in [S_{hp,m}]^d$, we set $[v] = (v^{(+)} - v^{(-)}) \cdot n_K$ for $\gamma \in \Omega$ and similarly for boundary edges.

For each space-time element $K \times I_m$, $K \in \mathcal{T}_h^m$, $m = 1, \dots, r$, we define the forms

$$\begin{aligned} a_{K,m}(u, v) &:= \big(\mathbf{K}(u)\nabla u, \nabla v\big)_{K,m} - \big(g, v\big)_{K,m} - \big(g_N, v\big)_{\partial K_N, m}, \\ A_{K,m}(u, v) &:= \big(\mathbf{K}(u)\nabla u, \nabla v\big)_{K,m} - \big(\langle \mathbf{K}(u)\nabla u \rangle \cdot n_K - \alpha [u] \cdot n_K, v\big)_{\partial K \setminus \Gamma_{\mathrm{N}}, m} \\ &\quad + (\beta - \tfrac{1}{2})\big(\mathbf{K}(u)[u], \nabla v\big)_{\partial K \setminus \Gamma, m} + (2\beta - 1)\big(\mathbf{K}(u)[u], \nabla v\big)_{\partial K \cap \Gamma_{\mathrm{D}}, m} \\ &\quad - \big(g, v\big)_{K,m} - \big(g_N, v\big)_{\partial K_N, m}, \end{aligned} \tag{12}$$

where $\alpha > 0$ is a sufficiently large penalization parameter ($\alpha \sim p_K^2 / h_K$) and $\beta \in \{0, \frac{1}{2}, 1\}$ corresponds to the choice of the variants of the interior penalty discretization (SIPG with $\beta = 0$, IIPG with $\beta = 1/2$ and NIPG with $\beta = 1$), see, e.g., [13, Chapter 2].

We introduce the space-time discontinuous Galerkin discretization of (3).

Definition 2 The function $u_h^\tau \in S_{hp}^{\tau q}$ is called the approximate solution of (6) obtained by the space-time discontinuous Galerkin method (STDGM), if

$$\sum_{K,m} B_{K,m}(u_h^\tau, v) = 0 \qquad \forall v \in S_{hp}^{\tau q}, \tag{13}$$

where

$$B_{K,m}(u, v) := \big(\vartheta'(u), v\big)_{K,m} + A_{K,m}(u, v) + \big(\{\vartheta(u)\}_{m-1}, v_+^{m-1}\big)_K \tag{14}$$

with form $A_{K,m}$ given by (12) and $\{\cdot\}$ defined by (9).

Remark 1 We note that u_h^τ is discontinuous with respect to time at t_m, $m = 1, \dots, r-1$. The solution between I_{m-1} and I_m is stuck together by the "time-penalty" term $\big(\{\vartheta(u)\}_{m-1}, v_+^{m-1}\big)_K$ which also makes sense for u and v belonging to different finite element spaces.

Finally, we derive some identities that will be used later. Let $\mathcal{F}_h^m$ denote the set of all interior edges $\gamma \not\subset \Gamma$ of mesh $\mathcal{T}_h^m$ and $\mathcal{F}_D^m$ the set of boundary edges of $\mathcal{T}_h^m$ lying on Γ_{D}. Then, the identity

$$\sum_{K\in\mathcal{T}_h^m} \big(w, z\, n_K\big)_{\partial K\setminus\Gamma_{\mathrm{N}},m} = \sum_{\gamma\in\mathcal{F}_h^m} \Big(\big(\langle w\rangle, [z]\big)_{\gamma,m} + \big([w], \langle z\rangle\big)_{\gamma,m}\Big) + \sum_{\gamma\in\mathcal{F}_D^m} \big(w\cdot n_K, z\big)_{\gamma,m} \tag{15}$$

holds for a piecewise smooth vector-valued function w and a piecewise smooth scalar function z.

Using identity (15) and the following obvious formulas valid for interior edges $\langle\langle \mathbf{K}(u)\nabla u\rangle\rangle = \langle \mathbf{K}(u)\nabla u\rangle$, $\langle \alpha[u]\rangle = \alpha[u]$, $[\langle \mathbf{K}(u)\nabla u\rangle] = 0$, $[\alpha[u]] = 0$, we gain

$$\begin{aligned}
\sum_{K\in\mathcal{T}_h^m} \big(\langle \mathbf{K}(u)\nabla u\rangle\cdot n_K, v\big)_{\partial K\setminus\Gamma_{\mathrm{N}},m} &= \sum_{\gamma\in\mathcal{F}_h^m} \big(\langle \mathbf{K}(u)\nabla u\rangle, [v]\big)_{\gamma,m} + \sum_{\gamma\in\mathcal{F}_D^m} \big(\mathbf{K}(u)\nabla u\cdot n_K, v\big)_{\gamma,m},\\
\sum_{K\in\mathcal{T}_h^m} \big(\alpha[u]\cdot n_K, v\big)_{\partial K\setminus\Gamma_{\mathrm{N}},m} &= \sum_{\gamma\in\mathcal{F}_h^m} \big(\alpha[u], [v]\big)_{\gamma,m} + \sum_{\gamma\in\mathcal{F}_D^m} \big(\alpha[u]\cdot n_K, v\big)_{\gamma,m},\\
\sum_{K\in\mathcal{T}_h^m} \big(\mathbf{K}(u)[u], \nabla v\big)_{\partial K\setminus\Gamma,m} &= \sum_{K\in\mathcal{T}_h^m} \big([u], \mathbf{K}(u)\nabla v\big)_{\partial K\setminus\Gamma,m} = 2\sum_{\gamma\in\mathcal{F}_h^m} \big([u], \langle \mathbf{K}(u)\nabla v\rangle\big)_{\gamma,m},\\
\sum_{K\in\mathcal{T}_h^m} \big(\mathbf{K}(u)[u], \nabla v\big)_{\partial K\cap\Gamma_{\mathrm{D}},m} &= \sum_{\gamma\in\mathcal{F}_D^m} \big([u], \mathbf{K}(u)\nabla v\big)_{\gamma,m}.
\end{aligned} \tag{16}$$

Consequently, from (12) and (16), we obtain the identity

$$\begin{aligned}
\sum_{K\in\mathcal{T}_h^m} A_{K,m}(u,v) = &\sum_{K\in\mathcal{T}_h^m} \big(\mathbf{K}(u)\nabla u, \nabla v\big)_{K,m} - \sum_{\gamma\in\mathcal{F}_h^m} \big(\langle \mathbf{K}(u)\nabla u\rangle, [v]\big)_{\gamma,m}\\
&+ (2\beta-1)\sum_{\gamma\in\mathcal{F}_h^m} \big([u], \langle \mathbf{K}(u)\nabla v\rangle\big)_{\gamma,m} - \sum_{\gamma\in\mathcal{F}_D^m} \big(\mathbf{K}(u)\nabla u\cdot n_K, v\big)_{\gamma,m}\\
&+ (2\beta-1)\sum_{\gamma\in\mathcal{F}_D^m} \big([u], \mathbf{K}(u)\nabla v\big)_{\gamma,m} + \sum_{\gamma\in\mathcal{F}_h^m} \big(\alpha[u], [v]\big)_{\gamma,m}\\
&+ \sum_{\gamma\in\mathcal{F}_D^m} \big(\alpha[u]\cdot n_K, v\big)_{\gamma,m} - \big(g, v\big)_{\Omega,m} - \big(g_N, v\big)_{\Gamma_{\mathrm{N}},m}.
\end{aligned} \tag{17}$$

4 A Posteriori Error Analysis

4.1 Error Measures

In order to proceed to the derivation of error estimators, we define the spaces of piecewise continuous functions with respect to time by

$$Y^\tau = \{v\in X : \vartheta'(v)|_{I_m} \in L^2(I_m, L^2(\Omega))\}, \qquad V^\tau = \{v\in Y^\tau : v|_{\Gamma_{\mathrm{D}}\times(0,T)} = 0\}. \tag{18}$$

Obviously, $Y^0 \subset Y \subset Y^\tau \subset X$ and $S_{hp}^{\tau q} \subset Y^\tau$. Moreover, we have the following result.

Lemma 1 *Let $u \in Y^0$ be the weak solution of (6). Then it satisfies*

$$\sum_{K,m} b_{K,m}(u,v) = 0 \quad \forall v\in V^\tau, \tag{19}$$

where

$$b_{K,m}(u,v) := \big(\vartheta'(u),v\big)_{K,m} + a_{K,m}(u,v) + \big(\{\vartheta(u)\}_{m-1}, v_+^{m-1}\big)_K \tag{20}$$

with $a_{K,m}$ given by (12) and the time jump $\{\cdot\}_{m-1}$ defined by (9). Moreover, there exists a unique solution $u \in Y^\tau$ such that $u - u^* \in V^\tau$ and satisfies (19).

Proof The proof follows directly by comparing formulas (19)–(20) with (6) and the fact that $\big(\{\vartheta(u)\}_{m-1}, v_+^{m-1}\big)_K = 0$ for $u \in Y^0$. For the proof of uniqueness, we employ the fact that $C_0^\infty(\Omega)$ is dense in $L^2(\Omega)$, i.e., there exists a sequence $\{v_\varepsilon\} \subset C_0^\infty(\Omega)$ for any $v \in L^2(\Omega)$ such that $\|v_\varepsilon - v\| \to 0$ as $\varepsilon \to 0$, cf. [34, Theorem 3.14]. We apply $v = v_{s,\varepsilon_1}(x)v_{t,\varepsilon_2}(t)$ in (19), where the spatial component $v_{s,\varepsilon_1} \in \{v \in H^1(\Omega) : v|_{\Gamma_\mathrm{D}} = 0\}$ tends to $\{\vartheta(u)\}_{m-1}$ as $\varepsilon_1 \to 0$ and the time component v_{t,ε_2} is given as 0 outside the interval $(t_{m-1}, t_{m-1}+\varepsilon_2)$ and $v_{t,\varepsilon_2} = 1 - (t - t_{m-1})/\varepsilon_2$ on $(t_{m-1}, t_{m-1}+\varepsilon_2)$, i.e., $v_{t,\varepsilon_2}(t)$ tends to 0 as $\varepsilon_2 \to 0$. Therefore, all the terms containing time integrals in (19) tend to 0 when ε_2 tends to 0. Since $v_+^{m-1} = v_{s,\varepsilon_1}$, the remaining jump term tends to $\|\{\vartheta(u)\}_{m-1}\|^2$ as ε_1 tends to 0. From this it follows that $\{\vartheta(u)\}_{m-1} = 0$. Then it is possible to see that any solution of (19) satisfies the original weak formulation (6). Since the weak problem (6) has a unique solution, cf. [2], the extended problem (19) has a unique solution as well. □

In virtue of [11, § 2.3.1], we define a parameter $d_{K,m}$ associated with the space-time element $K \times I_m$, $K \in \mathcal{T}_h^m$, $m = 1, \dots, r$. The parameter $d_{K,m}$ represents a user-dependent weight, typically with physical units $(\mathrm{T\,L})^{1/2}$ so that the error measure has the same physical unit as the energy norm. In this paper, we use two choices

$$d_{K,m} := \Big(h_K^{-2}\|\mathbf{K}(u_h)\|_{m,\infty} + \tau_m^{-2}T\,\big\|\tfrac{\mathrm{d}\vartheta}{\mathrm{d}u}(u_h)\big\|_{m,\infty}\Big)^{-1/2}, \tag{21a}$$

$$d_{K,m} := \Big(h_K^{2}\|\mathbf{K}(u_h)\|_{m,\infty}^{-1} + \tau_m^{2}/T\,\big\|\tfrac{\mathrm{d}\vartheta}{\mathrm{d}u}(u_h)\big\|_{m,\infty}^{-1}\Big)^{1/2}. \tag{21b}$$

where $\|\cdot\|_{m,\infty} := \|\cdot\|_{L^\infty(\Omega\times I_m)}$. We note that the following error analysis is independent of the choice of $d_{K,m}$. Moreover, we define the norm in the space V^τ (cf. (18)) by

$$\|v\|_{V^\tau}^2 = \sum_{K,m}\|v\|_{V_{K,m}}^2, \quad \|v\|_{V_{K,m}}^2 = d_{K,m}^{-2}\Big(h_K^2\|\nabla v\|_{K,m}^2 + \tau_m^2\big\|v'\big\|_{K,m}^2\Big). \tag{22}$$

In virtue of (19), we introduce the error measure as a dual norm of the residual

$$\mathcal{R}(u_h^\tau) = \sup_{0\neq v\in V^\tau}\frac{\sum_{K,m} b_{K,m}(u_h^\tau, v)}{\|v\|_{V^\tau}}, \tag{23}$$

where $b_{K,m}$ is given by (20). The residual $\mathcal{R}(v)$ represents a natural error measure for $u - v \in V^\tau$, cf. [11, Remark 2.3]. In Sect. 4, we estimate $\mathcal{R}(u_h^\tau)$ for u_h^τ being the solution of (13).

Since the approximate solution u_h^τ belongs to the space of discontinuous function $S_{hp}^{\tau q} \not\subset V^\tau$, we introduce the second building block measuring the nonconformity of the solution in spatial variables. Therefore, similarly to [18], we define the form

$$\mathcal{J}(v) = \sum_{K,m} J_{K,m}(v), \quad J_{K,m}(v) = d_{K,m}^2\,\tau_m^{-1}\,h_K^{-2}\,C_{K,m,\mathbf{K},\alpha}\|[v]\|_{\partial K,m}^2, \tag{24}$$

where $C_{K,m,\mathbf{K},\alpha} = \alpha^2 + \big\|\mathbf{K}(u_h^\tau)\big\|_{L^\infty(K\times I_m)}^2$. The scaling factors are chosen such that $\mathcal{J}(v)^{1/2}$ has the same physical unit as $\mathcal{R}(u_h^\tau)$.

We note that $\mathcal{J}(v)$ measures also the violation of the Dirichlet boundary condition since $\mathcal{J}(v)$ contains the term $\|v - u_D\|_{\partial K \cap \Gamma_D, m}$, cf. (11).

The final error measure is then defined by

$$\mathcal{E}(u_h^\tau) := \left(\mathcal{R}(u_h^\tau)^2 + \mathcal{J}(u_h^\tau)\right)^{1/2}, \tag{25}$$

where $\mathcal{R}(u_h^\tau)$ is given by (23) and $\mathcal{J}(u_h^\tau)$ by (24).

Lemma 2 *The error measure $\mathcal{E}(u_h^\tau) = 0$ if and only if $u_h^\tau = u$ is the weak solution given by (6).*

Proof Obviously, if $u_h^\tau = u$, then $\mathcal{J}(u_h^\tau) = 0$ and $\mathcal{R}(u_h^\tau) = 0$ due to (19). On the other hand, if $\mathcal{J}(u_h^\tau) = 0$, then $u_h^\tau \in Y^\tau$ and $u_h^\tau - u^* \in V^\tau$. Moreover, $\mathcal{R}(u_h^\tau) = 0$ and the uniqueness of (19) imply that u_h^τ is the weak solution (6). □

4.2 Temporal and Spatial Flux Reconstructions

Similarly as in [18], we define a temporal reconstruction $R_h^\tau = R_h^\tau(x,t)$ as a continuous function with respect to time that mimics $\vartheta(u_h^\tau)$, $u_h^\tau \in S_{hp}^{\tau q}$. Let $r_m \in P_{q+1}(I_m)$ be the right Radau polynomial on I_m, i.e., $r_m(t_{m-1}) = 1$ and $r_m(t_m) = 0$, and r_m is orthogonal to $P_{q-1}(I_m)$ with respect to the $L^2(I_m)$ inner product. Then we set

$$R_h^\tau(x,t) := \vartheta(u_h^\tau(x,t)) - \left\{\vartheta(u_h^\tau)\right\}_{m-1}(x)\, r_m(t), \quad x \in \Omega,\ t \in I_m, \tag{26}$$

where $\{\cdot\}$ is given by (9). The *temporal flux reconstruction* $R_h^\tau(x,t)$ is continuous in time, namely $R_h^\tau \in H^1(0,T,L^2(\Omega))$ and it satisfies the initial condition due to

$$\begin{aligned} R_h^\tau(\cdot,0) &= \vartheta(u_h^\tau(\cdot,0)) - \{\vartheta(u_h^\tau)\}_0(\cdot) r_1(0) \\ &= \vartheta(u_h^\tau(\cdot,0)) - (\vartheta(u_h^\tau(\cdot,0)) - \vartheta(u_0(\cdot)) = \vartheta(u_0(\cdot)). \end{aligned} \tag{27}$$

Moreover, by the integration by parts and the properties $r_m(t_{m-1}) = 1$, $r_m(t_m) = 0$, we obtain

$$\begin{aligned} \left((R_h^\tau - \vartheta(u_h^\tau))', v\right)_{K,m} &= -\left(r_m'\left\{\vartheta(u_h^\tau)\right\}_{m-1}, v\right)_{K,m} \\ &= \left(r_m\left\{\vartheta(u_h^\tau)\right\}_{m-1}, v'\right)_{K,m} - r_m(t_m)\left(\left\{\vartheta(u_h^\tau)\right\}_{m-1}, v_-^m\right)_K \\ &\quad + r_m(t_{m-1})\left(\left\{\vartheta(u_h^\tau)\right\}_{m-1}, v_+^{m-1}\right)_K \\ &= \left(r_m\left\{\vartheta(u_h^\tau)\right\}_{m-1}, v'\right)_{K,m} + \left(\left\{\vartheta(u_h^\tau)\right\}_{m-1}, v_+^{m-1}\right)_K, \qquad v \in V^\tau, \end{aligned} \tag{28}$$

which together with definition (26) implies

$$\left((R_h^\tau - \vartheta(u_h^\tau))', v\right)_{K,m} - \left(\left\{\vartheta(u_h^\tau)\right\}_{m-1}, v_+^{m-1}\right)_K = -\left(R_h^\tau - \vartheta(u_h^\tau), v'\right)_{K,m}, \quad v \in V^\tau. \tag{29}$$

Finally, using the orthogonality of r_m to $P_{q-1}(I_m)$, we obtain from (28), the formula

$$\left((R_h^\tau - \vartheta(u_h^\tau))', v\right)_{m,K} = \left(\{\vartheta(u_h^\tau)\}_{m-1}, v_+^{m-1}\right)_K \quad \forall v \in P_q(I_m, L^2(K)). \tag{30}$$

Consequently, if u_h^τ is the approximate solution given by (13), then it satisfies

$$\left((R_h^\tau)', v\right)_{K,m} = \left(\vartheta'(u_h^\tau), v\right)_{K,m} + \left(\left\{\vartheta(u_h^\tau)\right\}_{m-1}, v_+^{m-1}\right)_K = -A_{K,m}(u_h^\tau, v) \qquad (31)$$
$$\forall v \in P_q(I_m, P_{p_K}(K)).$$

Obviously, the reconstruction R_h^τ is local and explicit, so its computation is fast and easy to implement.

The *spatial flux reconstruction* needs to define a function $\sigma_h^\tau \in L^2(0, T, H(\mathrm{div}, \Omega))$ which mimics the flux $\sigma(u_h^\tau, \nabla u_h^\tau) = \mathbf{K}(u_h^\tau)\nabla u_h^\tau$, cf. (5). In particular, $\sigma_h^\tau|_{K\times I_m} \in P_q(I_m, \mathrm{RTN}_p(K))$ where

$$\mathrm{RTN}_p(K) = P_p(K)^d + x\, P_p(K), \qquad K \in \mathcal{T}_h,\ m = 1, \dots, r \qquad (32)$$

is the Raviart-Thomas-Nedelec finite elements, cf. [7] for more details. We assume that the reconstructed flux σ_h^τ has to be equilibrated with the temporal flux R_h^τ

$$\left(\nabla\cdot\sigma_h^\tau, v\right)_{K,m} = \left((R_h^\tau)' - g, v\right)_{K,m} \quad \forall v \in P_q(I_m, P_{p_K}(K)),\ K \in \mathcal{T}_h^m, \qquad (33)$$

and with the Neumann boundary condition

$$\left(\sigma_h^\tau \cdot n, v\right)_{\gamma,m} = \left(g_N, v\right)_{\gamma,m} \quad \forall v \in P_q(I_m, P_{p_K}(\gamma))\ \forall \gamma \subset \partial K_N,\ K \in \mathcal{T}_h^m. \qquad (34)$$

In Sect. 5 we present two possible constructions of σ_h^τ including the choice of the spatial polynomial degree p in (32).

4.3 Auxiliary Results

In the forthcoming numerical analysis, we need several technical tools. We will employ the *scaled space-time Poincarè inequality*, cf. [11, Lemma 2.2]: Let $\varphi_{K,m} \in P_0(K \times I_m)$ be the L^2-orthogonal projection of $\varphi \in H^1(K \times I_m)$ onto a constant in each space-time element $K \times I_m$, $K \in \mathcal{T}_h^m$, $m = 0, \dots, r$. Then,

$$\left\|\varphi - \varphi_{K,m}\right\|_{K,m} \le C_\mathrm{P}\left(h_K^2\|\nabla\varphi\|_{K,m}^2 + \tau_m^2\left\|\varphi'\right\|_{K,m}^2\right)^{1/2} = C_\mathrm{P} d_{K,m}\|\varphi\|_{V_{K,m}}, \qquad (35)$$

where C_P is the Poincarè constant equal to $1/\pi$ for simplicial elements and the last equality follows from (22).

Moreover, we introduce the *space-time trace inequality*

Lemma 3 *Let $\varphi_{\gamma,m} \in P_0(\gamma \times I_m)$ be the L^2-orthogonal projection of $\varphi \in H^1(K \times I_m)$ onto a constant on each $\gamma \times I_m$, where $\gamma \subset \partial K$ is an edge of $K \in \mathcal{T}_h^m$. Then there exists a constant $C_\mathrm{T} > 0$ such that*

$$\left\|\varphi - \varphi_{\gamma,m}\right\|_{\gamma\times I_m} \le C_\mathrm{T}\max(1, h_\gamma^{-1/2}) d_{K,m}\|\varphi\|_{V_{K,m}}, \qquad (36)$$

where $C_\mathrm{T} = \max(c_T, C_\mathrm{P})$, C_P is from (35) and $c_T > 0$ is the constant from the (space) trace inequality.

Proof The proof is straightforward, we present it for completeness. Let $\varphi \in H^1(K \times I_m)$ and, for all $t \in I_m$, set $\tilde\varphi(t) := |\gamma|^{-1}\int_\gamma \varphi(x,t)\,\mathrm{d}S$. Observing that $(\varphi - \tilde\varphi)$ and $(\tilde\varphi - \varphi_{\gamma,m})$ are $L^2(\gamma \times I_m)$-orthogonal, we have

$$\left\|\varphi - \varphi_{\gamma,m}\right\|_{\gamma\times I_m}^2 = \|\varphi - \tilde\varphi\|_{\gamma\times I_m}^2 + \left\|\tilde\varphi - \varphi_{\gamma,m}\right\|_{\gamma\times I_m}^2. \qquad (37)$$

Using the standard trace inequality (e.g., [21, Lemma 3.32]), we have

$$\|\varphi(\cdot,t) - \tilde\varphi(t)\|_\gamma \le c_T h_\gamma^{1/2}\|\nabla\varphi\|_K \qquad \forall t \in I_m, \qquad (38)$$

where $c_T > 0$ is a constant whose values can be set relatively precisely, see the discussion in [37, Section 4.6]. Hence, integrating the square of (38) over I_m and using the fact that $h_\gamma \le h_K$, $\gamma \subset h_K$, we estimate the first term on the right-hand side of (37) as

$$\|\varphi - \tilde{\varphi}\|^2_{\gamma\times I_m} \le c_T^2 h_\gamma \|\nabla\varphi\|^2_{K\times I_m} \le c_T^2 h_\gamma^{-1} h_K^2 \|\nabla\varphi\|^2_{K\times I_m}. \tag{39}$$

Using the fact that $\varphi_{\gamma,m} = \tau_m^{-1}\int_{I_m}\tilde{\varphi}(t)\,\mathrm{d}t$, the one-dimensional Poincarè inequality on I_n and the Cauchy–Schwarz inequality yield

$$\begin{aligned}\left\|\tilde{\varphi} - \varphi_{\gamma,m}\right\|^2_{\gamma\times I_m} &= |\gamma|\int_{I_m}|\tilde{\varphi} - \varphi_{\gamma,m}|^2(t)\,\mathrm{d}t \le |\gamma| C_{\mathrm{P}}^2\tau_m^2\int_{I_m}|\tfrac{\mathrm{d}}{\mathrm{d}t}\tilde{\varphi}(t)|^2\,\mathrm{d}t \\ &= \frac{C_{\mathrm{P}}^2\tau_m^2}{|\gamma|}\int_{I_m}\left(\int_\gamma \partial_t\varphi(x,t)\,\mathrm{d}x\right)^2\mathrm{d}t \le C_{\mathrm{P}}^2\tau_m^2\int_{I_m}\left(\int_\gamma|\partial_t\varphi|^2\,\mathrm{d}x\right)\mathrm{d}t \\ &= C_{\mathrm{P}}^2\tau_m^2\|\partial_t\varphi\|^2_{\gamma\times I_m}.\end{aligned} \tag{40}$$

Collecting bounds (37), (39), (40) and the definition of the norm (22) yields (36). □

4.4 Reliability

We presented the upper bound of $\mathcal{R}(u_h^\tau)$, cf. (23).

Theorem 1 *Let $u \in Y$ be the weak solution of (6) and $u_h^\tau \in S_{hp}^{\tau q}$ be the approximate solution given by (13). Let $R_h^\tau \in H^1(0,T,L^2(\Omega))$ be the temporal reconstruction given by (26) and $\sigma_h^\tau \in L^2(0,T,H(\mathrm{div},\Omega))$ be the spatial reconstruction satisfying (33). Then*

$$\mathcal{R}(u_h^\tau)^2 \le \eta^2 := \sum_{K,m}\eta_{K,m}^2, \quad \eta_{K,m} := C_{\mathrm{P}}\eta_{R,K,m} + (\eta_{S,K,m}^2 + \eta_{T,K,m}^2)^{1/2} + C_{\mathrm{T}}\eta_{N,K,m}, \tag{41}$$

where C_{P} is the constant from Poincarè inequality (35), C_{T} is the constant from the trace inequality (36) and the estimators $\eta_{R,K,m}$, $\eta_{S,K,m}$, $\eta_{T,K,m}$, and $\eta_{N,K,m}$ are given by

$$\eta_{R,K,m} := d_{K,m}\left\|(R_h^\tau)' - \nabla\cdot\sigma_h^\tau - g\right\|_{K,m}, \tag{42a}$$

$$\eta_{S,K,m} := \frac{d_{K,m}}{h_K}\left\|\sigma_h^\tau - \sigma(u_h^\tau,\nabla u_h^\tau)\right\|_{K,m}, \tag{42b}$$

$$\eta_{T,K,m} := \frac{d_{K,m}}{\tau_m}\left\|R_h^\tau - \vartheta(u_h^\tau)\right\|_{K,m}, \tag{42c}$$

$$\eta_{N,K,m} := \sum_{\gamma\subset\partial K_N}\max(1,h_\gamma^{-1/2})d_{K,m}\left\|\sigma_h^\tau\cdot n - g_N\right\|_{\partial K_N,m}. \tag{42d}$$

The proof of Theorem 1 can be found in [19] for the case of the homogeneous Dirichlet boundary condition. For completeness, we present its modification including mixed Dirichlet-Neumann boundary conditions.

Proof Starting from (20), adding the terms $\pm\big(R_h^\tau, v\big)_{K,m}$ and $\pm\big(\nabla\cdot\sigma_h^\tau, v\big)_{K,m}$, and using the integration by parts, we obtain

$$\begin{aligned}&\sum_{K,m} b_{K,m}(u_h^\tau, v) \\ &\quad= \sum_{K,m}\left\{\big(\vartheta'(u_h^\tau) - g, v\big)_{K,m} - \big(g_N, v\big)_{\partial K_N,m} + \big(\sigma(u_h^\tau,\nabla u_h^\tau),\nabla v\big)_{K,m} + \big(\{\vartheta(u_h^\tau)\}_{m-1}, v_+^{m-1}\big)_K\right\}\end{aligned} \tag{43}$$

$$= \sum_{K,m} \left((R_h^\tau)' - \nabla\cdot\sigma_h^\tau - g, v\right)_{K,m} - \sum_{K,m} \left(\sigma_h^\tau - \sigma(u_h^\tau, \nabla u_h^\tau), \nabla v\right)_{K,m}$$
$$- \sum_{K,m} \left\{ \left((R_h^\tau - \vartheta(u_h^\tau))', v\right)_{K,m} - \left(\{\vartheta(u_h^\tau)\}_{m-1}, v_+^{m-1}\right)_K \right\} + \sum_{K,m} \left(\sigma_h^\tau\cdot n - g_N, v\right)_{\partial K_N,m}$$
$$=: \xi_1 + \xi_2 + \xi_3 + \xi_4.$$

The terms ξ_i, $i = 1, \ldots, 4$ are estimated separately.

Let $v_{K,m} \in P_0(K \times I_m)$ be the piecewise constant projection of $v \in V^\tau$ given by the identity $\left(v_{K,m}, 1\right)_{K,m} = \left(v, 1\right)_{K,m}$. Using the Cauchy–Schwarz inequality, assumption (33), the Poincarè inequality (35), and (22), we have

$$|\xi_1| \le \sum_{K,m} \left|\left((R_h^\tau)' - \nabla\cdot\sigma_h^\tau - g, v\right)_{K,m}\right| = \sum_{K,m} \left|\left((R_h^\tau)' - \nabla\cdot\sigma_h^\tau - g, v - v_{K,m}\right)_{K,m}\right| \tag{44}$$
$$\le \sum_{K,m} C_{\mathrm{P}} \left\|(R_h^\tau)' - \nabla\cdot\sigma_h^\tau - g\right\|_{K,m} \left(h_K^2 \|\nabla v\|_{K,m}^2 + \tau_m^2 \|v'\|_{K,m}^2\right)^{1/2}$$
$$= \sum_{K,m} C_{\mathrm{P}}\, d_{K,m} \left\|(R_h^\tau)' - \nabla\cdot\sigma_h^\tau - g\right\|_{K,m} \|v\|_{V_{K,m}} = \sum_{K,m} C_{\mathrm{P}} \eta_{R,K,m} \|v\|_{V_{K,m}}.$$

Furthermore, by the Cauchy–Schwarz inequality and (22), we obtain

$$|\xi_2| \le \sum_{K,m} \left|\left(\sigma_h^\tau - \sigma(u_h^\tau, \nabla u_h^\tau), \nabla v\right)_{K,m}\right| \tag{45}$$
$$\le \sum_{K,m} \frac{d_{K,m}}{h_K} \left\|\sigma_h^\tau - \sigma(u_h^\tau, \nabla u_h^\tau)\right\|_{K,m} \frac{h_K}{d_{K,m}} \|\nabla v\|_{K,m} = \sum_{K,m} \eta_{S,K,m} \frac{h_K}{d_{K,m}} \|\nabla v\|_{K,m}.$$

The use of (29), and a similar manipulations as in (45), give

$$|\xi_3| \le \sum_{K,m} \left|\left((R_h^\tau - \vartheta(u_h^\tau))', v\right)_{K,m} - \left(\{\vartheta(u_h^\tau)\}_{m-1}, v_+^{m-1}\right)_K\right| = \sum_{K,m} \left|\left(R_h^\tau - \vartheta(u_h^\tau), v'\right)_{K,m}\right|$$
$$\le \sum_{K,m} \frac{d_{K,m}}{\tau_m} \left\|R_h^\tau - \vartheta(u_h^\tau)\right\|_{K,m} \frac{\tau_m}{d_{K,m}} \left\|v'\right\|_{K,m} = \sum_{K,m} \eta_{T,K,m} \frac{\tau_m}{d_{K,m}} \left\|v'\right\|_{K,m}. \tag{46}$$

Hence, estimates (45)–(46), the Cauchy inequality and (22) imply

$$|\xi_2| + |\xi_3| \le \sum_{K,m} \left(\eta_{S,K,m} \frac{h_K}{d_{K,m}} \|\nabla v\|_{K,m} + \eta_{T,K,m} \frac{\tau_m}{d_{K,m}} \left\|v'\right\|_{K,m}\right) \tag{47}$$
$$\le \sum_{K,m} \left(\eta_{S,K,m}^2 + \eta_{T,K,m}^2\right)^{1/2} \|v\|_{V_{K,m}}.$$

Furthermore, let $v_{\gamma,m} \in P_0(\gamma \times I_m)$, $\gamma \subset \partial K_N$ be the L^2-orthogonal projection from Lemma 3. Then using assumption (34), the Cauchy inequality and the space-time trace inequality (36), we have

$$|\xi_4| = \sum_{K,m} \sum_{\gamma\subset\partial K_N} \left(\sigma_h^\tau\cdot n - g_N, v - v_{\gamma,m}\right)_{\gamma,m} \le \sum_{K,m} \sum_{\gamma\subset\partial K_N} \left\|\sigma_h^\tau\cdot n - g_N\right\|_{\gamma,m} \left\|v - v_{\gamma,m}\right\|_{\gamma,m}$$
$$\le C_{\mathrm{T}} \sum_{K,m} \sum_{\gamma\subset\partial K_N} \max(1, h_\gamma^{-1/2}) d_{K,m} \left\|\sigma_h^\tau\cdot n - g_N\right\|_{\gamma,m} \|v\|_{V_{K,m}}. \tag{48}$$

The particular estimates (44), (47), and (48), together with the discrete Cauchy–Schwarz inequality, imply (41). □

Remark 2 Obviously, if $\partial K \cap \Gamma_{\mathrm{N}} \neq \emptyset$, then $\eta_{N,K,m} = 0$.

4.5 Efficiency

The aim is to show that the local individual error estimators $\eta_{R,K,m}$, $\eta_{S,K,m}$ and $\eta_{T,K,m}$ from (41)–(42) are locally efficient, i.e., they provide local lower bounds to the error measure up to a generic constant $C > 0$ which is independent of u, u_h^τ, h and τ, but may depend on data problems and the degrees of polynomial approximation p and q. A dependence of the estimate up to this generic constant we will denote by $\lesssim$.

In order to derive the local variants of the error measure, we denote by ω_K the set of elements sharing at least a vertex with $K \in \mathcal{T}_h^m$, i.e.,

$$\omega_K = \cup_{K' \cap K \neq 0} K', \qquad K \in \mathcal{T}_h^m,\ m = 0, \ldots, r. \tag{49}$$

Moreover, we define the functional sub-spaces $V_{D,m} = \{v \in V^\tau : \mathrm{supp}\,(v) \subset \overline{D \times I_m}\}$ for any set $D \subset \Omega$ (cf. (18)) and the corresponding error measures (cf. (23))

$$\mathcal{R}_{D,m}(w) = \sup_{\{0 \neq v \in V_{D,m}\}} \frac{1}{\|v\|_{V^\tau}} \sum_{K,m} b_{K,m}(w, v). \tag{50}$$

Obviously, the definition of $V_{D,m}$ and $\mathcal{R}_{D,m}(u_h^\tau)$ together with the shape regularity implies

$$\sum_{K,m} \mathcal{R}_{K,m}(u_h^\tau) \leq \sum_{K,m} \mathcal{R}_{\omega_K,m}(u_h^\tau) \lesssim \mathcal{R}(u_h^\tau). \tag{51}$$

Moreover, for each space-time element $K \times I_m$, $K \in \mathcal{T}_h^m$, $m = 1, \ldots, r$, we introduce the $L^2(K \times I_m)$-projection of the non-polynomial functions, namely

$$\begin{aligned} &\overline{\vartheta'(u_h^\tau)} \in P_q(I_m, P_{p_K}(K)) : \quad \big(\overline{\vartheta'(u_h^\tau)}, v\big)_{K,m} = \big(\vartheta'(u_h^\tau), v\big)_{K,m} \quad \forall v \in P_q(I_m, P_{p_K}(K)) \\ &\overline{g} \in P_q(I_m, P_{p_K}(K)) : \quad \big(\overline{g}, v\big)_{K,m} = \big(g, v\big)_{K,m} \quad \forall v \in P_q(I_m, P_{p_K}(K)). \end{aligned} \tag{52}$$

Finally, for each vertex a of the mesh $\mathcal{T}_h^m$, we denote by ω_a a patch of elements $K \in \mathcal{T}_h^m$ that share this vertex. By $p_a = \max_{K \in \omega_a} p_K$ we denote the maximal polynomial degree on ω_a. Then, for each a of $K \in \mathcal{T}_h^m$, we define a vector-valued function $\overline{\sigma}_a = \overline{\sigma}_a(u_h^\tau, \nabla u_h^\tau) \in P_q(I_m, \mathrm{RTN}_{p_a}(K))$ (cf. (32)) by

$$\begin{aligned} \big(\overline{\sigma}_a \cdot n_K, v\big)_{\gamma,m} &= \big(\psi_a \langle \sigma(u_h^\tau, \nabla u_h^\tau) \rangle \cdot n_K, v\big)_{\gamma,m} \quad \forall v \in P_q(I_m, P_{p_a}(\gamma)),\ \gamma \subset K \\ (\overline{\sigma}_a \cdot v)_{K,m} &= (\psi_a \sigma(u_h^\tau, \nabla u_h^\tau), v)_{K,m} \quad \forall v \in P_q(I_m, P_{p_a - 1}(K)^d), \end{aligned} \tag{53}$$

where $\langle \cdot \rangle$ denotes the mean value on $\gamma \subset \partial K$ and ψ_a is a continuous piecewise linear function such that $\psi_a(a) = 1$ and it vanishes at the other vertices of K. Finally, we set $\overline{\sigma}|_{K \times I_m} = \sum_{a \in K} \overline{\sigma}_a$.

The proof of the local efficiency of the error estimates presented is based on the choice of a suitable test function in (23). We set

$$w(x,t) = \frac{d_{K,m}^2}{\tau_m} P_h\big(\{\vartheta(u_h^\tau)\}_{m-1}\big)(x) \chi_K(x) \Phi_m(t). \tag{54}$$

where $\chi_K(x)$ is the standard bubble function on K, $\Phi_m(t)$ is the Legendre polynomial of degree $q+1$ on I_m (and vanishing outside) and $P_h\big(\{\vartheta(u_h^\tau)\}_{m-1}\big) \in P_{p_K}(K)$ is the $L^2(K)$-projection weighted by $\chi_K(x)$ given by

$$\big(P_h\big(\{\vartheta(u_h^\tau)\}_{m-1}\big), \chi_K v\big)_K = \big(\{\vartheta(u_h^\tau)\}_{m-1}, \chi_K v\big)_K \qquad \forall v \in P_{p_K}(K). \tag{55}$$

We note that

$$P_h\big(\{\vartheta(u_h^\tau)\}_{m-1}\big) \neq \{\overline{\vartheta(u_h^\tau)}\}_{m-1}, \tag{56}$$

in general, compare with (52).

Using the inverse inequality, the polynomial function w given by (54) can be estimated as

$$\begin{aligned} \|w\|^2_{V_{K,m}} &= d_{K,m}^{-2}\left(h_K^2\|\nabla w\|^2_{K,m} + \tau_m^2\|w'\|^2_{K,m}\right) \lesssim d_{K,m}^{-2}\|w\|^2_{K,m} \\ &\leq \frac{d_{K,m}^2}{\tau_m^2}\left\|P_h\big(\{\vartheta(u_h^\tau)\}_{m-1}\big)\right\|^2_K \int_{I_m}\Phi_m^2(t)\,\mathrm{d}t \lesssim \frac{d_{K,m}^2}{\tau_m}\left\|P_h\big(\{\vartheta(u_h^\tau)\}_{m-1}\big)\right\|^2_K. \end{aligned} \tag{57}$$

Similarly as in [11] or [18], we introduce the oscillation terms

$$\eta_{G,K,m} := d_{K,m}\|\overline{g} - g\|_{K,m}, \quad \eta_{\vartheta,K,m} := \frac{d_{K,m}}{\sqrt{\tau_m}}\left\|\{\vartheta(u_h^\tau)\}_{m-1} - P_h\big(\{\vartheta(u_h^\tau)\}_{m-1}\big)\right\|_K, \tag{58}$$

$$\eta_{\vartheta',K,m} := d_{K,m}\left\|\overline{\vartheta'(u_h^\tau)} - \vartheta'(u_h^\tau)\right\|_{K,m},$$

$$\eta_{\sigma,K,m} := \frac{d_{K,m}}{h_K}\left\|\overline{\sigma} - \sigma(u_h^\tau, \nabla u_h^\tau)\right\|_{K,m} + d_{K,m}\left\|\nabla\cdot\overline{\sigma} - \nabla\cdot\sigma(u_h^\tau, \nabla u_h^\tau)\right\|_{K,m}.$$

The goal is to prove the lower bounds of the proposed error estimates, namely to estimate $\eta_{T,K,m}$, $\eta_{R,K,m}$ and $\eta_{S,K,m}$ by $\mathcal{R}_{K,m}(u_h^\tau)$ and the oscillation terms (58), $K \in \mathcal{T}_h$, $m = 1,\dots,r$.

Theorem 2 *Let $\eta_{T,K,m}$, $K \in \mathcal{T}_h^m$, $m = 1,\dots,r$ be the error estimates given by (42), then*

$$\eta_{T,K,m} \lesssim \mathcal{R}_{K,m}(u_h^\tau) + \eta_{G,K,m} + \eta_{\vartheta',K,m} + \eta_{\vartheta,K,m} + \eta_{S,K,m}. \tag{59}$$

where $\mathcal{R}_{K,m}$ are the local error measures defined by (49)–(50) and the oscillation terms $\eta_{G,K,m}$, $\eta_{\vartheta,K,m}$ and $\eta_{\vartheta',K,m}$ are given by (58).

Proof We start the proof by the putting function w from (54) as the test function in (50), i.e.

$$\mathcal{R}_{K,m}(u_h^\tau) = \sup_{0\neq v\in V_{K,m}} \frac{\sum_{K,m} b_{K,m}(u_h^\tau, v)}{\|v\|_{V^\tau}} \geq \frac{b_{K,m}(u_h^\tau, w)}{\|w\|_{V^\tau}} \tag{60}$$

since $\mathrm{supp}(w) = K \times I_m$, cf. (54). Then, using (20) and the fact that w vanishes on ∂K, we have

$$\mathcal{R}_{K,m}(u_h^\tau) \geq \frac{\big(\vartheta'(u_h^\tau) - g, w\big)_{K,m} + \big(\sigma(u_h^\tau, \nabla u_h^\tau), \nabla w\big)_{K,m} + \big(\{\vartheta(u_h^\tau)\}_{m-1}, w_+^{m-1}\big)_K}{\|w\|_{V_{K,m}}} \tag{61}$$

$$= \frac{\left(\overline{\vartheta'(u_h^\tau)} - \overline{g}, w\right)_{K,m} + \left(\sigma_h^\tau, \nabla w\right)_{K,m}}{\|w\|_{V_{K,m}}} + \frac{\left(\left\{\vartheta(u_h^\tau)\right\}_{m-1}, w_+^{m-1}\right)_K}{\|w\|_{V_{K,m}}} =: \xi_1 + \xi_2$$
$$+ \frac{\left(\overline{g} - g, w\right)_{K,m} + \left(\sigma - \sigma_h^\tau, \nabla w\right)_{K,m} + \left(\vartheta'(u_h^\tau) - \overline{\vartheta'(u_h^\tau)}, w\right)_{K,m}}{\|w\|_{V_{K,m}}}$$
$$=: \xi_3 + \xi_4 + \xi_5.$$

The functions $\overline{\vartheta'(u_h^\tau)}$, $\overline{g}$ and σ_h^τ are polynomials of degree q in time whereas w and ∇w are the (Legendre) polynomial of degree $(q+1)$ in time, cf. (54). Due to the $L^2(I_m)$-orthogonality of the Legendre polynomials, we have $\xi_1 = 0$, since

$$\left(\overline{\vartheta'(u_h^\tau)} - \overline{g}, w\right)_{K,m} + \left(\sigma_h^\tau, \nabla w\right)_{K,m} = 0 \tag{62}$$

Moreover, using inequality (57), relations (54)-(55) and the equivalence of norms on finite dimensional spaces,

we obtain

$$\begin{aligned}\xi_2 &\gtrsim \frac{\left(P_h\left(\left\{\vartheta(u_h^\tau)\right\}_{m-1}\right), \frac{d_{K,m}^2}{\tau_m} P_h\left(\left\{\vartheta(u_h^\tau)\right\}_{m-1}\right)\chi_K\right)_K}{\frac{d_{K,m}}{\sqrt{\tau_m}} \left\| P_h\left(\left\{\vartheta(u_h^\tau)\right\}_{m-1}\right)\right\|_K} \\ &\gtrsim \frac{d_{K,m}}{\sqrt{\tau_m}} \frac{\left(P_h\left(\left\{\vartheta(u_h^\tau)\right\}_{m-1}\right), P_h\left(\left\{\vartheta(u_h^\tau)\right\}_{m-1}\right)\right)_K}{\left\| P_h\left(\left\{\vartheta(u_h^\tau)\right\}_{m-1}\right)\right\|_K} = \frac{d_{K,m}}{\sqrt{\tau_m}} \left\| P_h\left(\left\{\vartheta(u_h^\tau)\right\}_{m-1}\right)\right\|_K.\end{aligned} \tag{63}$$

Furthermore, let $w_{K,m} = \frac{1}{K \times I_m} \int_{K \times I_m} w \, \mathrm{d}x \, \mathrm{d}t$ be the mean value of w on the space-time element $K \times I_m$. Due to (52), the Cauchy–Schwarz inequality and (35), we have

$$\begin{aligned}|\xi_3| &= \frac{\left|\left(\overline{g} - g, w - w_{K,m}\right)_{K,m}\right|}{\|w\|_{V_{K,m}}} \\ &\leq \frac{\|\overline{g} - g\|_{K,m} \left\|w - w_{K,m}\right\|_{K,m}}{\|w\|_{V_{K,m}}} \lesssim d_{K,m} \|\overline{g} - g\|_{K,m} = \eta_{G,K,m},\end{aligned} \tag{64}$$

and

$$|\xi_5| \lesssim d_{K,m} \left\| \vartheta'(u_h^\tau) - \overline{\vartheta'(u_h^\tau)} \right\|_{K,m} = \eta_{\vartheta',K,m}. \tag{65}$$

Similarly, the Cauchy–Schwarz inequality and (22) imply

$$\begin{aligned}|\xi_4| &\leq \frac{d_{K,m}}{h_K} \left\|\sigma(u_h^\tau, \nabla u_h^\tau) - \sigma_h^\tau\right\|_{K,m} \frac{h_K \|\nabla w\|_{K,m}}{d_{K,m} \|w\|_{V_{K,m}}} \\ &\leq \frac{d_{K,m}}{h_K} \left\|\sigma(u_h^\tau, \nabla u_h^\tau) - \sigma_h^\tau\right\|_{K,m} = \eta_{S,K,m}.\end{aligned} \tag{66}$$

Collecting (61)–(66), we have

$$\mathcal{R}_{K,m}(u_h^\tau) \gtrsim \frac{d_{K,m}}{\sqrt{\tau_m}} \left\| P_h\left(\left\{\vartheta(u_h^\tau)\right\}_{m-1}\right)\right\|_K - \eta_{G,K,m} - \eta_{S,K,m} - \eta_{\vartheta',K,m}. \tag{67}$$

Moreover, using (42c), (26), integration by parts, the boundedness of the Radau polynomials, the triangle inequality and (58), we have

$$\begin{aligned}\eta_{T,K,m} &= \frac{d_{K,m}}{\tau_m}\left\|R_h^\tau - \vartheta(u_h^\tau)\right\|_{K,m} = \frac{d_{K,m}}{\tau_m}\left\|\left\{\vartheta(u_h^\tau)\right\}_{m-1} r_m\right\|_{K,m} \\ &= \frac{d_{K,m}}{\tau_m}\left\|\left\{\vartheta(u_h^\tau)\right\}_{m-1}\right\|_K \sqrt{\int_{I_m} r_m^2 \,\mathrm{d}t} \lesssim \frac{d_{K,m}}{\sqrt{\tau_m}}\left\|\left\{\vartheta(u_h^\tau)\right\}_{m-1}\right\|_K \\ &\leq \frac{d_{K,m}}{\sqrt{\tau_m}}\left\|P_h\left(\left\{\vartheta(u_h^\tau)\right\}_{m-1}\right)\right\|_K + \eta_{\vartheta,K,m}.\end{aligned} \tag{68}$$

Hence, (67) and (68)

$$\eta_{T,K,m} \leq \mathcal{R}_{K,m}(u_h^\tau) + \eta_{\vartheta,K,m} + \eta_{G,K,m} + \eta_{\vartheta',K,m} + \eta_{S,K,m}, \tag{69}$$

which proves the theorem. □

Theorem 3 *Let $\eta_{S,K,m}$ and $\eta_{R,K,m}$, $K \in \mathcal{T}_h^m$, $m = 1, \ldots, r$ be the error estimates given by (42), then*

$$\eta_{R,K,m} \lesssim \mathcal{R}_{\omega_K,m}(u_h^\tau) + \eta_{G,K,m} + \eta_{\sigma,K,m} + \eta_{S,K,m}, \tag{70}$$

$$\eta_{S,K,m} \lesssim \mathcal{R}_{\omega_K,m}(u_h^\tau) + \eta_{G,K,m} + \sum_{K\subset\omega_K} \eta_{\sigma,K,m}, \tag{71}$$

where $\mathcal{R}_{\omega_K,m}$ is the local error measures defined by (49)–(50) and the oscillation terms $\eta_{G,K,m}$, $\eta_{\vartheta,K,m}$ and $\eta_{\vartheta',K,m}$ are given by (58).

Proof The proof is in principle identical with the proof [18, Lemmas 7-9], we present the main step for completeness. Let $\overline{g}$ and $\overline{\sigma}$ be the projection given by (52) and (53). Using the triangle inequality, the inverse inequality and (58), we obtain

$$\begin{aligned}\eta_{R,K,m} &= d_{K,m}\left\|(R_h^\tau)' - \nabla\cdot\sigma_h^\tau - g\right\|_{K,m} \\ &\leq d_{K,m}\left\|(R_h^\tau)' - \nabla\cdot\overline{\sigma} - \overline{g}\right\|_{K,m} + d_{K,m}\|\overline{g} - g\|_{K,m} + d_{K,m}\left\|\nabla\cdot\overline{\sigma} - \nabla\cdot\sigma_h^\tau\right\|_{K,m} \\ &\lesssim d_{K,m}\left\|(R_h^\tau)' - \nabla\cdot\overline{\sigma} - \overline{g}\right\|_{K,m} + \eta_{G,K,m} + \frac{d_{K,m}}{h_K}\left\|\overline{\sigma} - \sigma_h^\tau\right\|_{K,m}.\end{aligned} \tag{72}$$

The first term on the right-hand side of (72) can be estimated as in [36, Theorem 4.10] by

$$d_{K,m}\left\|(R_h^\tau)' - \nabla\cdot\overline{\sigma} - \overline{g}\right\|_{K,m} \lesssim Res_{\omega_K,m}(u_h^\tau) + \eta_{G,K,m} + \eta_{\sigma,K,m}, \tag{73}$$

where the resulting oscillation terms are estimated with the aid (58). Moreover, the last term on the right-hand side of (72) together with (42b) and assumption (58), reads

$$\begin{aligned}\frac{d_{K,m}}{h_K}\left\|\overline{\sigma} - \sigma_h^\tau\right\|_{K,m} &\leq \frac{d_{K,m}}{h_K}\left\|\overline{\sigma} - \sigma(u_h^\tau, \nabla u_h^\tau)\right\|_{K,m} + \frac{d_{K,m}}{h_K}\left\|\sigma(u_h^\tau, \nabla u_h^\tau) - \sigma_h^\tau\right\|_{K,m} \\ &\leq \eta_{\sigma,K,m} + \eta_{S,K,m},\end{aligned} \tag{74}$$

which proves (70).

The proof of (71) is based on the decomposition

$$\left\|\sigma_h^\tau - \sigma(u_h^\tau, \nabla u_h^\tau)\right\|_{K,m} \leq \left\|\sigma_h^\tau - \overline{\sigma}\right\|_{K,m} + \left\|\overline{\sigma} - \sigma(u_h^\tau, \nabla u_h^\tau)\right\|_{K,m}. \tag{75}$$

While the second term on the right-hand side of (75) can be estimated by assumption (58), the estimate of the first term is somewhat more technical. It depends on the flux reconstruction used. For the flux reconstruction in Sect. 5.2, the proof is identical to the proof of [18, Lemma 9], which mimics the stationary variant [24, Theorem 3.12]. On the other hand, using the flux

reconstruction from Sect. 5.1, it is possible to apply the technique from [11, Lemma 7.5], where the final relation has to be integrated over I_m. □

5 Spatial Flux Reconstructions and Stopping Criteria

We present two ways of reconstructing the spatial flux $\sigma_h^\tau \in L^2(0,T,H(\mathrm{div},\Omega))$ that satisfies the assumptions (33)–(34). The first one, proposed in [19] for the case of homogeneous Dirichlet boundary condition, is defined by the volume and edge momenta of the Raviart-Thomas-Nedelec (RTN) elements, cf. [7], and is easy to compute. The second approach is based on the solution of local Neumann problems on patches associated with each vertex of the mesh. This idea comes from, e.g., [24], its space-time variant was proposed in [18] for nonlinear convection-diffusion equations. Finally, in Sect. 5.3, we discuss the errors arising from the solution of algebraic systems and introduce a stopping criterion for the appropriate iterative solver.

5.1 Element-Wise Variant

We denote by $p_{K,\max}$ the maximum polynomial degree over the element K and its neighbours that share the entire edge with K and $p_{\gamma,\max}$ the maximum polynomial degree on neighbouring elements having a common edge γ. Let $\mathrm{RTN}_{p_{K,\max}}(K)$ be the space of RTN finite elements of order $p_{K,\max}$ for element $K \in \mathcal{T}_h^m$, cf. (32), and $u_h^\tau \in S_{hp}^{\tau q}$ be the approximate solution. The spatial reconstruction σ_h^τ is defined element-wise: for each $K \in \mathcal{T}_h^m$, find $\sigma_h^\tau|_{K\times I_m} \in P_q(I_m, \mathrm{RTN}_{p_{K,\max}}(K))$ with $\sigma_h^\tau \cdot n|_{\gamma\times I_m} \in P_q(I_m, P_{p_{\gamma,\max}}(\gamma))$ such that

$$\left(\sigma_h^\tau \cdot n, v\right)_{\gamma,m} = \begin{cases} \left(\left\langle \mathbf{K}(u_h^\tau)\nabla u_h^\tau \right\rangle \cdot n - \alpha[u_h^\tau]\cdot n, v\right)_{\gamma,m} & \forall v \in P_q(I_m, P_{p_{\gamma,\max}}(\gamma)),\ \gamma \subset \partial K \setminus \Gamma_{\mathrm{N}} \\ \left(g_N, v\right)_\gamma & \forall v \in P_q(I_m, P_{p_{\gamma,\max}}(\gamma)),\ \gamma \subset \partial K_N \end{cases} \tag{76}$$

$$\begin{aligned} \left(\sigma_h^\tau, v\right)_{K,m} &= \left(\mathbf{K}(u_h^\tau)\nabla u_h^\tau, \nabla v\right)_{K,m} + (\beta - \tfrac{1}{2})\left(\mathbf{K}(u_h^\tau)[u_h^\tau], \nabla v\right)_{\partial K\setminus\Gamma,m} \\ &\quad + (2\beta-1)\left(\mathbf{K}(u_h^\tau)[u_h^\tau], \nabla v\right)_{\partial K\cap\Gamma_{\mathrm{D}},m} \qquad \forall v \in P_q(I_m, P_{p_{K,\max}-1}(K)^d). \end{aligned}$$

The edge momenta in (76) are uniquely defined and since $p_{\gamma,\max} \le p_{K,\max}$, σ_h^τ in (76) is well defined as well. Here, the numerical flux $\left\langle \mathbf{K}(u_h^\tau)\nabla u_h^\tau \right\rangle \cdot n - \alpha[u_h^\tau]\cdot n$ is conservative on interior edges, which implies that $\sigma_h^\tau \cdot n$ are the same on each interior edge γ and therefore the resulting reconstruction $\sigma_h^\tau \in L^2(0,T,H(\mathrm{div},\Omega))$ globally.

Obviously, the first relation in (76) with $p_K \le p_{\gamma,\max}$ directly implies assumption (34). Moreover, using the Green theorem, (76), (12), (31) and $p_K \le p_{\gamma,\max} \le p_{K,\max}$, we obtain

$$\begin{aligned} \left(\nabla\cdot\sigma_h^\tau, v\right)_{K,m} &= -\left(\sigma_h^\tau, \nabla v\right)_{K,m} + \left(\sigma_h^\tau \cdot n_K, v\right)_{\partial K,m} \\ &= -\left(\mathbf{K}(u_h^\tau)\nabla u_h^\tau, \nabla v\right)_{K,m} + \left(\left\langle \mathbf{K}(u_h^\tau)\nabla u_h^\tau \right\rangle \cdot n - \alpha[u_h^\tau]\cdot n, v\right)_{\partial K\setminus\Gamma_{\mathrm{N}},m} \\ &\quad - (\beta-\tfrac{1}{2})\left(\mathbf{K}(u_h^\tau)[u_h^\tau], \nabla v\right)_{\partial K\setminus\Gamma,m} - (2\beta-1)\left(\mathbf{K}(u_h^\tau)[u_h^\tau], \nabla v\right)_{\partial K\cap\Gamma_{\mathrm{D}},m} \\ &\quad + \left(g_N, v\right)_{\partial K_N,m} \\ &= -A_{K,m}(u_h^\tau, v) - \left(g, v\right)_{K,m} = \left((R_h^\tau)' - g, v\right)_{K,m} \\ &\forall v \in P_q(I_m, P_{p_K}(K)),\ K \in \mathcal{T}_h^m, \end{aligned} \tag{77}$$

which justifies the assumption (33).

5.2 Patch-Wise Variant

For each vertex a of the mesh $\mathcal{T}_h^m$, we denote by ω_a a patch of elements $K \in \mathcal{T}_h^m$ sharing this vertex. By $p_a = \max_{K\in\omega_a} p_K$ we denote the maximal polynomial degree on ω_a. Let $P^*_{p_a}(\omega_a)$ be the space of piecewise polynomial discontinuous functions of degree p_a on ω_a with mean value zero for $a \notin \partial\Omega$. We define the space

$$
\begin{aligned}
W^N_{\mathrm{RTN},p_a}(\omega_a) &= \{v \in H(\mathrm{div},\omega_a);\, v|_K \in \mathrm{RTN}_{p_a}(K),\, v\cdot n = 0 \text{ on } \partial\omega_a\}, \quad a \notin \partial\Omega \\
W^N_{\mathrm{RTN},p_a}(\omega_a) &= \{v \in H(\mathrm{div},\omega_a);\, v|_K \in \mathrm{RTN}_{p_a}(K),\, v\cdot n = 0 \text{ on } \partial\omega_a \setminus \partial\Omega, \\
&\qquad \& \ (v\cdot n,\phi)_{\gamma,m} = (g_N,\phi)_{\gamma,m}\ \forall\phi \in P_q(I_m, P_{p_a}(\gamma)) \text{ on } \partial\omega_a \cap \partial K_N\}, a \in \partial\Omega.
\end{aligned}
\tag{78}
$$

We set the local problems on patches ω_a for all vertices a: find $\sigma_h^\tau \in P_q(I_m, W^N_{\mathrm{RTN},p_a}(\omega_a))$ and $r_a^\tau \in P_q(I_m, P^*_{p_a}(\omega_a))$ such that

$$
\begin{aligned}
\left(\sigma_a^\tau, v\right)_{\omega_a,m} - \left(r_a^\tau, \nabla\cdot v\right)_{\omega_a,m} &= \left(\xi_a^1, v\right)_{\omega_a,m} \quad \forall v \in P_q(I_m, W^N_{\mathrm{RTN},p_a}(\omega_a)) \\
\left(\nabla\cdot\sigma_a^\tau, \phi\right)_{\omega_a,m} &= \left(\xi_a^2, \phi\right)_{\omega_a,m} \quad \forall\phi \in P_q(I_m, P^*_{p_a}(\omega_a)),
\end{aligned}
\tag{79}
$$

where

$$
\begin{aligned}
\xi_a^1 &= \psi_a \sigma(u_h^\tau, \nabla u_h^\tau) \\
\xi_a^2 &= \psi_a (R_h^\tau)' - \psi_a g + \nabla\psi_a \cdot \vec{\xi}(u_h^\tau, \nabla u_h^\tau),
\end{aligned}
\tag{80}
$$

with

$$
\vec{\xi}(u_h^\tau, \nabla u_h^\tau) = \sigma(u_h^\tau, \nabla u_h^\tau) + (2\beta - 1) \sum_{\gamma\not\subset\Gamma_N} \ell_{m,\gamma}(u_h^\tau),
\tag{81}
$$

and $\ell_{m,\gamma} : S_{hp,m} \to [S_{h0,m}]^d$ is the lifting operator defined by

$$
\int_\Omega \ell_{m,\gamma}(u_h^\tau)\cdot v \,\mathrm{d}x = \int_\gamma [u_h^\tau] \left\langle \mathbf{K}(u_h^\tau) v \right\rangle \mathrm{d}x \qquad \forall v \in [S_{h0,m}]^d, \quad \gamma \not\subset \Gamma_N.
\tag{82}
$$

Then the final reconstructed flux is obtained by summing up σ_a^τ on each element that contains vertex a, i.e.,

$$
\sigma_h^\tau|_{K,m} = \sum_{a\in K} \sigma_a^\tau|_K.
\tag{83}
$$

The assumption (34) follows directly from (78) and $p_K \le p_a$. Inserting the hat function $\psi_a v$ for $a \notin \partial\Omega$ and $v \in P_q(I_m)$ in (17), using (5), (82) and omitting the zero terms, we have

$$
\begin{aligned}
&\sum_{K\in\mathcal{T}_h^m} A_{K,m}(u_h^\tau, \psi_a v) \\
&\quad = \sum_{K\in\mathcal{T}_h^m} \left(\mathbf{K}(u_h^\tau)\nabla u_h^\tau, \nabla\psi_a v\right)_{K,m} \\
&\qquad + (2\beta-1) \sum_{\gamma\not\subset\partial\Omega} \left([u_h^\tau], \left\langle \mathbf{K}(u_h^\tau)\nabla\psi_a v\right\rangle\right)_{\gamma,m} + (2\beta-1) \sum_{\gamma\subset\Gamma_D} \left([u_h^\tau], \mathbf{K}(u_h^\tau)\nabla\psi_a v\right)_{\gamma,m} \\
&\qquad - \left(g, \psi_a v\right)_{\Omega,m} = \left(\xi_a^2, v\right)_{\omega_a,m} - \left(R_h^\tau, \psi_a v\right)_{\omega_a,m}
\end{aligned}
\tag{84}
$$

Applying (13) and (31), we gain for $a \notin \partial\Omega$ and $v \in P_q(I_m)$

$$\left(\nabla \cdot \sigma_a^\tau, v\right)_{\omega_a,m} = \sum_{K\subset\omega_a} \left(A_{K,m}(u_h^\tau, \psi_a v) + \left(R_h^\tau, \psi_a v\right)_{K,m} \right) = \left(\xi_a^2, v\right)_{\omega_a,m}. \tag{85}$$

From this it follows that the second relation in (79) holds element-wise, i.e.

$$\left(\nabla \cdot \sigma_a^\tau, \phi\right)_{K,m} = \left(\xi_a^2, \phi\right)_{K,m}, \quad \forall \phi \in P_q(I_m, P_{p_a}(K)). \tag{86}$$

Then (33) follows from

$$\begin{aligned}(\nabla \cdot \sigma_h^\tau, \phi)_{K,m} &= \sum_{a\subset K} (\nabla \cdot \sigma_a^\tau, \phi)_{K,m} = \sum_{a\subset K} \left(\xi_a^2, \phi\right)_{K,m} \\ &= ((R_h^\tau)' - g, \phi)_{K,m} \quad \forall \phi \in P_q(I_m, P_{p_a}(K))\end{aligned} \tag{87}$$

and from $p_K \le p_a$.

5.3 Stopping Criteria for Iterative Solvers

The space-time discretization (13) leads to a system of nonlinear algebraic equations for each time level $m = 1, \dots, r$. These systems have to be solved iteratively by a suitable solver, e.g., the Picard method, the Newton method or their variants. Therefore, it is necessary to set a suitable stopping criterion for the iterative solvers so that, on the one hand, the algebraic errors do not affect the quality of the approximate solution and, on the other hand, an excessive number of algebraic iterations is avoided.

However, the error estimates presented in Sect. 4 do not take into account errors arising from the inaccurate solution of these systems. Indeed, the aforementioned reconstructions fulfill assumption (33) only if the systems given by (13) are solved exactly. The full a posteriori error analysis including algebraic errors has been treated, e.g., in [8, 23, 29]. These error estimators are based on additional flux reconstructions that need to be evaluated at each iteration, and therefore, the overall computational time is increased.

To speed up the computations and control the algebraic errors, we adopt the technique of [17]. This approach offers (i) the measurement of algebraic errors by a quantity similar to the error measure (23), (ii) the setting of the stopping criterion for iterative solvers with one parameter corresponding to the relative error, and (iii) a fast evaluation of the required quantities.

For each $m = 1, \dots, r$, we define the estimators (cf. (23))

$$\eta_{\mathrm{alg}}^m(u_h^\tau) = \sup_{0\neq v\in S_{hp}^{\tau q}} \frac{\sum_{K\in\mathcal{T}_h^m} b_{K,m}(u_h^\tau, v)}{\|v\|_{V^\tau}}, \qquad \eta_{\mathrm{spa}}^m(u_h^\tau) = \sup_{0\neq v\in S_{hp+1}^{\tau q+1}} \frac{\sum_{K\in\mathcal{T}_h^m} b_{K,m}(u_h^\tau, v)}{\|v\|_{V^\tau}}, \tag{88}$$

where the norm $\|\cdot\|_{V^\tau}$ is given by (22),

$$\begin{aligned} S_{hp+1}^{\tau q+1} &= \{v \in L^2(\Omega \times (0,T)) : \ v|_{I_m} \in P_{q+1}(I_m, S_{hp+1,m}),\ m = 1, \dots, r\}, \\ \text{and} \quad S_{hp+1,m} &= \{v \in L^2(\Omega) : v|_K \in P_{p_K+1}(K),\ K \in \mathcal{T}_h^m\}, \qquad m = 0, \dots, r. \end{aligned} \tag{89}$$

The space $S_{hp+1}^{\tau q+1}$ is an enrichment space of $S_{hp}^{\tau q}$ by polynomials of the space degree $p_K + 1$ and the time degree $q + 1$ for each $K \times I_m$, $K \in \mathcal{T}_h^m$, $m = 1, \dots, r$. Finally, we define the

global in time quantities

$$\eta_{\mathrm{alg}}(u_h^\tau) = \left(\sum_{m=1}^{r} (\eta_{\mathrm{alg}}^m(u_h^\tau))^2 \right)^{1/2}, \qquad \eta_{\mathrm{spa}}(u_h^\tau) = \left(\sum_{m=1}^{r} (\eta_{\mathrm{spa}}^m(u_h^\tau))^2 \right)^{1/2}. \tag{90}$$

Obviously, if u_h^τ fulfills (13) exactly, then $\eta_{\mathrm{alg}}^m(u_h^\tau) = 0$ for all $m = 0, \dots, r$. Moreover, if u_h^τ is the weak solution (6) then $\eta_{\mathrm{spa}}^m(u_h^\tau) = 0$ for all $m = 0, \dots, r$. Comparing (88) with (23), the quantity $\eta_{\mathrm{spa}}(u_h^\tau)$ exhibits a variant of the error measure $\mathcal{R}(u_h^\tau)$. Nevertheless, $\eta_{\mathrm{spa}}(u_h^\tau)$ is neither lower nor upper bound of $\mathcal{R}(u_h^\tau)$ since $S_{hp+1}^{\tau q+1} \not\subset V^\tau$ and $V^\tau \not\subset S_{hp+1}^{\tau q+1}$.

The quantities (88) can be evaluated very fast since the suprema (maxima) are the sum of the suprema (maxima) for all space-time elements $K \times I_m$, $K \in \mathcal{T}_h^m$, $m = 1, \dots, r$, which are computable separately, cf. [17] for details. Hence, we prescribe the stopping criterion for the corresponding iterative solver as

$$\eta_{\mathrm{alg}}^m(u_h^\tau) \le c_A \eta_{\mathrm{spa}}^m(u_h^\tau), \qquad m = 1, \dots, r, \tag{91}$$

where $c_A \in (0, 1)$ is the user-dependent constant. The justification of this approach and the influence of algebraic errors on the error estimates are studied numerically in Sect. 6.1.1.

6 Numerical Experiments

We present numerical experiments that justify the a posteriori error estimates (41)–(42). Since the error measure (23) is the dual norm of the residual, it is not possible to evaluate the error even if the exact solution is known. Therefore, similarly to [18], we approximate the error by solving the dual problem given for each time interval I_m, $m = 1, \dots, r$: Find $\psi_m \in Y_m^\tau = L^2(I_m, H^1(\Omega))$,

$$\big(\psi_m, \phi\big)_{Y_m^\tau} = \sum_{K,m} b_{K,m}(u_h^\tau, \phi) \qquad \forall \phi \in Y_m^\tau, \tag{92}$$

where (cf. (21a)–(22))

$$\big(u, v\big)_{Y_m^\tau} = \sum\nolimits_{K \in \mathcal{T}_h^m} d_{K,m}^{-2} \left(h_K^2 \big(\nabla u, \nabla v\big)_{K,m} + \tau_m^2 \big(u', v'\big)_{K,m} \right), \qquad m = 1, \dots, r. \tag{93}$$

Then we have $\mathcal{R}(u_h^\tau)^2 = \sum_{m=1}^r \|\psi\|_{Y_m^\tau}^2$. We solve (92) for each $m = 1, \dots, r$ by linear conforming finite element on a global refinement of the space-time mesh $\mathcal{T}_h^m \times I_m$ which is proportional to the space and time polynomial approximation degrees. We denote this quantity by $\widetilde{\mathcal{R}}(u_h^\tau)$. The second error contribution $\mathcal{J}$ given by (24) is computable, so the total error $\mathcal{E}$ (cf. (25)) is approximated by $\widetilde{\mathcal{E}}(u_h^\tau) := \big(\widetilde{\mathcal{R}}(u_h^\tau)^2 + \mathcal{J}(u_h^\tau)\big)^{1/2}$.

Remark 3 Sometimes, this approximate evaluation of the (exact) error is not sufficiently accurate for fine grids and high polynomial approximation degrees. In this case, very fine global refinement is required and then the resulting algebraic systems are too large to be solved in a reasonable time.

All numerical experiments were carried out using the patch-wise reconstruction from Sect. 5.2 using the in-house code ADGFEM [10]. The arising nonlinear algebraic systems are solved iteratively by a Newton-like method, we refer to [14] for details. Each Newton-line iteration leads to a linear algebraic system that is solved by GMRES method with block ILU(0) preconditioner.

6.1 Barenblatt Problems

We consider two nonlinear variants of (3) following from the Barenblatt problem [4] where the analytical solution exists. The first variant reads

$$\partial_t \vartheta(u) - \Delta u = 0, \qquad \vartheta(u) = u^{1/m}, \qquad m \in (0,1), \tag{94}$$

where the analytical solution is

$$u(x_1,x_2,t) = \frac{1}{1+t}\left(\left[1 - \frac{m-1}{4m^2}\frac{x_1^2+x_2^2}{(1+t)^{1/m}}\right]_+\right)^{\frac{m}{m-1}}, \qquad \lfloor z \rfloor_+ := \max(z,0),\ z \in \mathbb{R} \tag{95}$$

Using the substitution $v := u^{1/m}$, we have the second variant

$$\partial_t v - \nabla \cdot (m|v|^{m-1}\nabla v) = 0, \qquad m > 1, \tag{96}$$

having the solution

$$v(x_1,x_2,t) = \left\{\frac{1}{1+t}\left(\left[1 - \frac{m-1}{4m^2}\frac{x_1^2+x_2^2}{(1+t)^{1/m}}\right]_+\right)^{\frac{m}{m-1}}\right\}^{1/m}. \tag{97}$$

For both problems ((94)–(95) and (96)–(97)), the computational domain is $\Omega = (-6,6)^2$ and the Dirichlet boundary condition is prescribed on all boundaries by (95) or (97). The final time is $T = 1$.

We carried out computation using a sequence of uniform triangular grids (having 288, 1152, 4608 and 18432 triangles) with several combinations of polynomial approximation degrees with respect to space (p) and time (q). The time step has been chosen constant $\tau = 0.01$. Besides the error quantities ($\widetilde{\mathcal{R}}(u_h^\tau)$ and $\mathcal{J}(u_h^\tau)$) and its estimators η, $\eta_R := \sum_{K,m} \eta_{R,K,m}$, $\eta_S := \sum_{K,m} \eta_{S,K,m}$ and $\eta_T := \sum_{K,m} \eta_{T,K,m}$, we evaluate the effectivity indices

$$i_{\text{eff}} = \frac{\eta}{\widetilde{\mathcal{R}}(u_h^\tau)}, \qquad i_{\text{eff}}^{\text{tot}} = \frac{\left(\eta^2 + \mathcal{J}(u_h^\tau)\right)^{1/2}}{\widetilde{\mathcal{E}}(u_h^\tau)}. \tag{98}$$

In addition, we present the experimental order of convergence (EoC) of the errors and the estimators for each pair of successive meshes.

Tables 1–4 show the results for both Barenblatt problems ((94)–(95) with $m = 0.25$ and (96)–(97) with $m = 2$) with two variants of the scaling parameter $d_{K,m}$, $K \in \mathcal{T}_h^m$, $m = 1,\dots,r$ given by (21a) and (21b). The quantity #DoF represents the number of degrees of freedom in the space, that is, #DoF $= \dim S_{hp,m}$, $m = 1,\dots,r$. We observe a good correspondence between $\widetilde{\mathcal{R}}(u_h^\tau)$ and η, the effectivity index i_{eff} varies between 1 and 2.5 for all tested values of p and q and both variants of $d_{K,m}$ ((21a) and (21b)).

Finally, we note that the experimental orders of convergence EoC in Tables 1–4) of the error $\widetilde{\mathcal{R}}(u_h^\tau)$ and its estimate η are $O(h^p)$ for the choice (21b) of the scaling parameter $d_{K,m}$ but only $O(h^{p-1})$ for the choice (21a). This follows from the fact that $\tau_m \ll h_K$ for the computations of the Barenblatt problem and then the dominant part of $d_{K,m}$ is $\tau_m^{-2} T \left\| \frac{\mathrm{d}\vartheta}{\mathrm{d}u} \right\|_{K,m,\infty}$, cf. (21a), which implies that $d_{K,m} = O(h^0)$ (the time step is the same for all computations). The dominant part of the error estimator is $\eta_{S,K,m}$, hence if $\left\| \sigma_h^\tau - \sigma(u_h^\tau, \nabla u_h^\tau) \right\|_{K,m} = O(h^p)$ then $\eta_{S,K,m} = O(h^{p-1})$, cf. (42b). Nevertheless, comparing the pairs of Tables 1–2 and Tables 3–4, we found that the effectivity indexes are practically independent of the choice of $d_{K,m}$.

Table 1 Barenblatt problem (94)–(95), $m = 0.25$, scaling parameter $d_{K,m}$ given by (21a), approximation of the error and the error estimators, EOC in parenthesis

h	#DoF	$\widetilde{\mathcal{R}}(u_h^\tau)$	η	$\mathcal{J}(u_h^\tau)$	η_R	η_S	η_T	i_{eff}	$i_{\text{eff}}^{\text{tot}}$
p = 1 & q = 1									
1.41	864	8.42×10^{-4}	1.46×10^{-3}	4.01×10^{-3}	7.67×10^{-6}	1.22×10^{-3}	7.94×10^{-4}	1.73	1.13
0.71	3456	7.31×10^{-4}	1.29×10^{-3}	3.69×10^{-3}	7.68×10^{-6}	1.16×10^{-3}	5.38×10^{-4}	1.76	1.13
		(0.20)	(0.18)	(0.12)	(0.00)	(0.07)	(0.56)		
0.35	13824	4.98×10^{-4}	1.04×10^{-3}	2.95×10^{-3}	8.56×10^{-6}	1.02×10^{-3}	1.55×10^{-4}	2.09	1.16
		(0.55)	(0.30)	(0.33)	(−0.16)	(0.18)	(1.80)		
0.18	55296	4.40×10^{-4}	1.01×10^{-3}	2.81×10^{-3}	9.00×10^{-6}	1.01×10^{-3}	3.22×10^{-5}	2.31	1.18
		(0.18)	(0.04)	(0.07)	(−0.07)	(0.02)	(2.26)		
p = 2 & q = 2									
1.41	1728	2.06×10^{-4}	3.49×10^{-4}	1.32×10^{-3}	5.60×10^{-7}	3.17×10^{-4}	1.46×10^{-4}	1.70	1.09
0.71	6912	1.16×10^{-4}	1.86×10^{-4}	7.91×10^{-4}	4.59×10^{-8}	1.74×10^{-4}	6.64×10^{-5}	1.60	1.08
		(0.82)	(0.91)	(0.74)	(3.61)	(0.86)	(1.14)		
0.35	27648	6.51×10^{-5}	8.69×10^{-5}	4.33×10^{-4}	5.46×10^{-8}	8.57×10^{-5}	1.44×10^{-5}	1.34	1.04
		(0.84)	(1.10)	(0.87)	(−0.25)	(1.02)	(2.20)		
0.18	110592	3.51×10^{-5}	4.34×10^{-5}	2.28×10^{-4}	6.54×10^{-8}	4.32×10^{-5}	2.46×10^{-6}	1.23	1.03
		(0.89)	(1.00)	(0.92)	(−0.26)	(0.99)	(2.56)		

Table 1 continued

h	#DoF	$\widetilde{\mathcal{R}}(u_h^\tau)$	η	$\mathcal{J}(u_h^\tau)$	η_R	η_S	η_T	i_{eff}	$i_{\text{eff}}^{\text{tot}}$
p = 3 & q = 2									
1.41	2880	6.32×10^{-5}	1.16×10^{-4}	3.74×10^{-4}	4.39×10^{-8}	9.31×10^{-5}	6.83×10^{-5}	1.83	1.12
0.71	11520	2.03×10^{-5}	3.02×10^{-5}	1.45×10^{-4}	3.48×10^{-8}	2.82×10^{-5}	1.08×10^{-5}	1.49	1.06
		(1.64)	(1.94)	(1.37)	(0.33)	(1.73)	(2.66)		
0.35	46080	5.77×10^{-6}	8.16×10^{-6}	4.11×10^{-5}	5.31×10^{-8}	8.03×10^{-6}	1.20×10^{-6}	1.41	1.05
		(1.81)	(1.89)	(1.81)	(−0.61)	(1.81)	(3.17)		
0.18	184320	1.35×10^{-6}	2.09×10^{-6}	9.47×10^{-6}	6.53×10^{-8}	2.04×10^{-6}	8.52×10^{-8}	1.55	1.07
		(2.10)	(1.97)	(2.12)	(−0.30)	(1.98)	(3.82)		

Table 2 Barenblatt problem (94)–(95), $m = 0.25$, scaling parameter $d_{K,m}$ given by (21b), approximation of the error and the error estimators, EOC in parenthesis

h	#DoF	$\widetilde{\mathcal{R}}(u_h^\tau)$	η	$\mathcal{J}(u_h^\tau)$	η_R	η_S	η_T	i_{eff}	$i_{\text{eff}}^{\text{tot}}$
p = 1 & q = 1									
1.41	864	2.34×10^{-1}	4.06×10^{-1}	1.12	2.14×10^{-3}	3.39×10^{-1}	2.21×10^{-1}	1.73	1.13
0.71	3456	1.02×10^{-1}	1.79×10^{-1}	5.14×10^{-1}	1.07×10^{-3}	1.62×10^{-1}	7.50×10^{-2}	1.76	1.13
		(1.20)	(1.18)	(1.12)	(1.00)	(1.07)	(1.56)		
0.35	13824	3.47×10^{-2}	7.26×10^{-2}	2.05×10^{-1}	5.96×10^{-4}	7.14×10^{-2}	1.08×10^{-2}	2.09	1.16
		(1.55)	(1.30)	(1.33)	(0.84)	(1.18)	(2.80)		
0.18	55296	1.53×10^{-2}	3.54×10^{-2}	9.80×10^{-2}	3.14×10^{-4}	3.51×10^{-2}	1.12×10^{-3}	2.31	1.18
		(1.18)	(1.04)	(1.07)	(0.93)	(1.02)	(3.26)		
0.09	221184	7.63×10^{-3}	1.76×10^{-2}	4.85×10^{-2}	1.59×10^{-4}	1.75×10^{-2}	1.44×10^{-4}	2.31	1.18
		(1.01)	(1.00)	(1.01)	(0.98)	(1.00)	(2.97)		
p = 2 & q = 2									
1.41	1728	5.73×10^{-2}	9.73×10^{-2}	3.68×10^{-1}	1.56×10^{-4}	8.83×10^{-2}	4.08×10^{-2}	1.70	1.09
0.71	6912	1.62×10^{-2}	2.60×10^{-2}	1.10×10^{-1}	6.40×10^{-6}	2.43×10^{-2}	9.24×10^{-3}	1.60	1.08
		(1.82)	(1.91)	(1.74)	(4.61)	(1.86)	(2.14)		
0.35	27648	4.53×10^{-3}	6.06×10^{-3}	3.02×10^{-2}	3.81×10^{-6}	5.97×10^{-3}	1.01×10^{-3}	1.34	1.04
		(1.84)	(2.10)	(1.87)	(0.75)	(2.02)	(3.20)		
0.18	110292	1.22×10^{-3}	1.51×10^{-3}	7.96×10^{-3}	2.28×10^{-6}	1.51×10^{-3}	8.56×10^{-5}	1.23	1.03
		(1.89)	(2.00)	(1.92)	(0.74)	(1.99)	(3.55)		

Table 2 continued

h	#DoF	$\widetilde{\mathcal{R}}(u_h^\tau)$	η	$\mathcal{J}(u_h^\tau)$	η_R	η_S	η_T	i_{eff}	$i_{\text{eff}}^{\text{tot}}$
p = 3 & q = 2									
1.41	2880	1.76×10^{-2}	3.22×10^{-2}	1.04×10^{-1}	1.22×10^{-5}	2.59×10^{-2}	1.90×10^{-2}	1.83	1.12
0.71	11520	2.83×10^{-3}	4.20×10^{-3}	2.02×10^{-2}	4.85×10^{-6}	3.92×10^{-3}	1.51×10^{-3}	1.49	1.06
		(2.64)	(2.94)	(2.37)	(1.33)	(2.73)	(3.66)		
0.35	46080	4.42×10^{-4}	5.68×10^{-4}	2.87×10^{-3}	3.70×10^{-6}	5.59×10^{-4}	8.38×10^{-5}	1.29	1.04
		(2.68)	(2.89)	(2.81)	(0.39)	(2.81)	(4.17)		
0.18	184320	4.71×10^{-5}	7.28×10^{-5}	3.30×10^{-4}	2.27×10^{-6}	7.10×10^{-5}	2.97×10^{-6}	1.55	1.07
		(3.23)	(2.97)	(3.12)	(0.70)	(2.98)	(4.82)		

Table 3 Barenblatt problem (96)–(97), $m = 2$, scaling parameter $d_{K,m}$ given by (21a), approximation of the error and the error estimators, EOC in parenthesis

h	#DoF	$\widetilde{\mathcal{R}}(u_h^\tau)$	η	$\mathcal{J}(u_h^\tau)$	η_R	η_S	η_T	i_{eff}	$i_{\text{eff}}^{\text{tot}}$
p = 1 & q = 1									
1.41	864	3.35×10^{-3}	5.76×10^{-3}	1.59×10^{-2}	6.44×10^{-7}	4.58×10^{-3}	3.50×10^{-3}	1.72	1.13
0.71	3456	1.71×10^{-3}	3.19×10^{-3}	7.67×10^{-3}	5.90×10^{-7}	2.74×10^{-3}	1.64×10^{-3}	1.87	1.16
		(0.97)	(0.85)	(1.05)	(0.13)	(0.74)	(1.09)		
0.35	13824	1.09×10^{-3}	2.27×10^{-3}	4.57×10^{-3}	4.81×10^{-7}	2.17×10^{-3}	6.69×10^{-4}	2.07	1.21
		(0.64)	(0.49)	(0.75)	(0.29)	(0.34)	(1.29)		
0.18	55296	9.17×10^{-4}	2.04×10^{-3}	3.60×10^{-3}	3.62×10^{-7}	2.02×10^{-3}	2.41×10^{-4}	2.22	1.25
		(0.26)	(0.15)	(0.34)	(0.41)	(0.10)	(1.48)		
p = 2 & q = 2									
1.41	1728	6.33×10^{-4}	1.27×10^{-3}	4.52×10^{-3}	1.98×10^{-7}	9.68×10^{-4}	8.18×10^{-4}	2.00	1.12
0.71	6912	3.26×10^{-4}	5.98×10^{-4}	2.20×10^{-3}	1.63×10^{-7}	4.70×10^{-4}	3.71×10^{-4}	1.84	1.11
		(0.96)	(1.08)	(1.04)	(0.28)	(1.04)	(1.14)		
0.35	27648	1.91×10^{-4}	2.99×10^{-4}	1.17×10^{-3}	1.24×10^{-7}	2.60×10^{-4}	1.47×10^{-4}	1.57	1.08
		(0.77)	(1.00)	(0.91)	(0.40)	(0.85)	(1.33)		
0.18	110592	1.23×10^{-4}	1.71×10^{-4}	7.18×10^{-4}	8.96×10^{-8}	1.63×10^{-4}	5.26×10^{-5}	1.39	1.06
		(0.63)	(0.81)	(0.71)	(0.47)	(0.68)	(1.48)		

Table 3 continued

h	#DoF	$\widetilde{\mathcal{R}}(u_h^\tau)$	η	$\mathcal{J}(u_h^\tau)$	η_R	η_S	η_T	i_{eff}	$i_{\text{eff}}^{\text{tot}}$
p = 3 & q = 2									
1.41	2880	2.57×10^{-4}	5.65×10^{-4}	2.00×10^{-3}	1.26×10^{-7}	3.69×10^{-4}	4.28×10^{-4}	2.20	1.14
0.71	11520	1.51×10^{-4}	2.70×10^{-4}	1.06×10^{-3}	8.66×10^{-8}	2.07×10^{-4}	1.74×10^{-4}	1.79	1.10
		(0.77)	(1.06)	(0.91)	(0.54)	(0.84)	(1.30)		
0.35	46080	1.00×10^{-4}	1.50×10^{-4}	6.70×10^{-4}	6.12×10^{-8}	1.33×10^{-4}	6.91×10^{-5}	1.50	1.06
		(0.59)	(0.85)	(0.66)	(0.50)	(0.63)	(1.33)		
0.18	184320	6.92×10^{-5}	9.37×10^{-5}	4.48×10^{-4}	4.30×10^{-8}	9.04×10^{-5}	2.46×10^{-5}	1.35	1.05
		(0.53)	(0.68)	(0.58)	(0.51)	(0.56)	(1.49)		

Table 4 Barenblatt problem (96)–(97), $m = 2$, scaling parameter $d_{K,m}$ given by (21b), approximation of the error and the error estimators, EOC in parenthesis

h	#DoF	$\widetilde{\mathcal{R}}(u_h^\tau)$	η	$\mathcal{J}(u_h^\tau)$	η_R	η_S	η_T	i_{eff}	$i_{\text{eff}}^{\text{tot}}$
p = 1 & q = 1									
1.41	864	3.35×10^{-1}	5.76×10^{-1}	1.59	6.44×10^{-5}	4.58×10^{-1}	3.50×10^{-1}	1.72	1.13
0.71	3456	8.54×10^{-2}	1.60×10^{-1}	3.84×10^{-1}	2.95×10^{-5}	1.37×10^{-1}	8.19×10^{-2}	1.87	1.16
		(1.97)	(1.85)	(2.05)	(1.13)	(1.74)	(2.09)		
0.35	13824	2.74×10^{-2}	5.68×10^{-2}	1.14×10^{-1}	1.20×10^{-5}	5.43×10^{-2}	1.68×10^{-2}	2.07	1.21
		(1.64)	(1.49)	(1.75)	(1.29)	(1.34)	(2.29)		
0.18	55296	1.15×10^{-2}	2.57×10^{-2}	4.53×10^{-2}	4.56×10^{-6}	2.55×10^{-2}	3.03×10^{-3}	2.22	1.25
		(1.25)	(1.15)	(1.34)	(1.40)	(1.09)	(2.47)		
0.09	221184	5.54×10^{-3}	1.27×10^{-2}	2.13×10^{-2}	1.64×10^{-6}	1.26×10^{-2}	6.05×10^{-4}	2.29	1.27
		(1.06)	(1.02)	(1.09)	(1.48)	(1.01)	(2.32)		
p = 2 & q = 2									
1.41	1728	6.33×10^{-2}	1.27×10^{-1}	4.52×10^{-1}	1.98×10^{-5}	9.68×10^{-2}	8.18×10^{-2}	2.00	1.12
0.71	6912	1.63×10^{-2}	2.99×10^{-2}	1.10×10^{-1}	8.16×10^{-6}	2.35×10^{-2}	1.85×10^{-2}	1.84	1.11
		(1.96)	(2.08)	(2.04)	(1.28)	(2.04)	(2.14)		
0.35	27648	4.78×10^{-3}	7.49×10^{-3}	2.94×10^{-2}	3.10×10^{-6}	6.52×10^{-3}	3.69×10^{-3}	1.57	1.08
		(1.77)	(2.00)	(1.90)	(1.40)	(1.85)	(2.33)		
0.18	110592	1.55×10^{-3}	2.15×10^{-3}	9.04×10^{-3}	1.13×10^{-6}	2.04×10^{-3}	6.62×10^{-4}	1.39	1.06
		(1.63)	(1.80)	(1.70)	(1.46)	(1.67)	(2.48)		
p = 3 & q = 2									
1.41	2880	2.57×10^{-2}	5.65×10^{-2}	2.00×10^{-1}	1.26×10^{-5}	3.69×10^{-2}	4.28×10^{-2}	2.20	1.14
0.71	11520	7.54×10^{-3}	1.35×10^{-2}	5.31×10^{-2}	4.33×10^{-6}	1.03×10^{-2}	8.70×10^{-3}	1.79	1.10
		(1.77)	(2.06)	(1.91)	(1.54)	(1.84)	(2.30)		
0.35	46080	2.51×10^{-3}	3.76×10^{-3}	1.68×10^{-2}	1.53×10^{-6}	3.34×10^{-3}	1.73×10^{-3}	1.50	1.06
		(1.59)	(1.85)	(1.66)	(1.50)	(1.63)	(2.33)		
0.18	184320	8.71×10^{-4}	1.18×10^{-3}	5.64×10^{-3}	5.42×10^{-7}	1.14×10^{-3}	3.09×10^{-4}	1.36	1.05
		(1.53)	(1.67)	(1.57)	(1.50)	(1.55)	(2.48)		

6.1.1 Justification of the Algebraic Stopping Criterion (91)

We present the numerical study of the stopping criterion (91) which is used in the iterative solution of algebraic systems given by (13). We consider again the Barenblatt problem (94)–(95) with $m = 0.25$ and (96)–(97) with $m = 2$. The user-dependent constant c_A in (91) has been chosen as 10^{-1}, 10^{-2}, 10^{-3} and 10^{-4}. Tables 5 and 6 show the estimators η, $\mathcal{J}(u_h^\tau)$, η_{alg} and η_{alg}, cf. (90), for selected meshes and polynomial approximation degrees and the scaling parameter $d_{K,m}$ chosen by (21a).

Additionally, we present the total number of steps of the Newton-like solver N_{non}, the total number of GMRES iterations N_{lin} and the computational time in seconds. The computational time has only an informative character.

We observe that the error estimators η, $\mathcal{J}(u_h^\tau)$ and also η_{spa} converge to the limit values for decreasing c_A in (91) which mimic the case when the algebraic errors are negligible. Moreover, the relative differences between the actual values η and $\mathcal{J}(u_h^\tau)$ and their limits correspond more or less to the value of c_A. Obviously, smaller values of c_A cause prolongation of the computational time, due to a higher number of iterations, with a negligible effect on accuracy. Thus, the choice $c_A = 10^{-2}$ seems to be optimal in order to balance accuracy and efficiency.

The presented numerical experiments indicate that the estimator $\eta_{\text{spa}}(u_h^\tau)$ gives an upper bound of $\mathcal{R}(u_h^\tau)$, however, this observation is not supported by the theory. The quantity $\eta_{\text{spa}}(u_h^\tau)$ is used only in the stopping criterion (91).

6.2 Tracy Problem

Tracy problem represents a standard benchmark, where the analytical solutions of the Richards equation are available [35]. We consider the Gardners constitutive relations [26]

$$\mathbf{K}(u) = \begin{cases} \mathbf{K}_s \exp(-\alpha\psi) & \text{if } \psi > 0 \\ \mathbf{K}_s & \text{if } \psi \le 0 \end{cases}, \qquad \vartheta(u) = \begin{cases} \theta_r + (\theta_s - \theta_r)\exp(-\alpha\psi) & \text{if } \psi > 0 \\ \theta_s & \text{if } \psi \le 0 \end{cases} \tag{99}$$

where $\psi = u - z$ is the pressure head, z is the vertical coordinate and the material parameters $\mathbf{K}_s = 1.2\mathbb{I}$, $\theta_s = 0.5$, $\theta_r = 0.0$, and $\alpha = 0.1$ are the isotropic conductivity, saturated water content, residual water content, and the soil index parameter related to the pore-size distribution, respectively.

The computational domain is $\Omega = (0, 1)^2$, the initial condition is set $u = u_r := -10$ in Ω where u_r corresponds to the hydraulic head when the porous medium is dry. On the top part of the boundary $\Gamma_1 := \{(x, z),\ x \in (0, 1),\ z = 1\}$, we prescribe the boundary condition

$$u(x) = \frac{1}{\alpha}\log\Big(\exp(\alpha u_r) + (1 - \exp(\alpha u_r))\sin(\pi x)\Big), \qquad x \in (0, 1) \tag{100}$$

and on the rest of boundary Γ we set $u = u_r$. We note that this benchmark poses an inconsistency between the initial and boundary conditions on Γ_1. Hence, the most challenging part is the computation close to $t = 0$. In order to avoid the singularity at $t = 0$, we investigate the error only on the interval $t \in [1.0 \times 10^{-5}, 1.1 \times 10^{-4}]$ with the fixed time step τ is 1.0×10^{-6}.

We perform a computation using a sequence of uniform triangular grids with several combinations of polynomial approximation degrees and the choice (21b), the results are shown in Table 7. We observe reasonable values of the effectivity indices except for the finest

Table 5 Barenblatt problem (94)–(95), $m = 0.25$, scaling parameter $d_{K,m}$ given by (21a), numerical study of the algebraic stopping criterion (91)

c_A	η	$\mathcal{J}(u_h^\tau)$	η_{alg}	η_{spa}	N_{non}	N_{lin}	time(s)
$h = 0.35$, $p = 1$ & $q = 1$, #DoF = 13824							
1.0×10^{-1}	1.2475×10^{-3}	3.0760×10^{-3}	8.1322×10^{-4}	1.3804×10^{-2}	202	14148	422.1
1.0×10^{-2}	1.0559×10^{-3}	2.9565×10^{-3}	7.6470×10^{-5}	1.3483×10^{-2}	362	21589	606.8
1.0×10^{-3}	1.0435×10^{-3}	2.9468×10^{-3}	7.9268×10^{-6}	1.3458×10^{-2}	529	26545	693.6
1.0×10^{-4}	1.0423×10^{-3}	2.9457×10^{-3}	7.3279×10^{-7}	1.3456×10^{-2}	579	27766	705.1
$h = 0.35$, $p = 2$ & $q = 2$, #DoF = 27648							
1.0×10^{-1}	1.0443×10^{-4}	4.3586×10^{-4}	5.7369×10^{-5}	1.1019×10^{-3}	406	10581	1968.3
1.0×10^{-2}	8.8249×10^{-5}	4.3375×10^{-4}	6.1984×10^{-6}	1.0961×10^{-3}	536	12059	2119.4
1.0×10^{-3}	8.7054×10^{-5}	4.3350×10^{-4}	6.0680×10^{-7}	1.0956×10^{-3}	576	12541	2030.1
1.0×10^{-4}	8.6948×10^{-5}	4.3347×10^{-4}	5.1172×10^{-8}	1.0955×10^{-3}	618	13580	2132.1
$h = 0.35$, $p = 3$ & $q = 2$, #DoF = 46080							
1.0×10^{-1}	9.9098×10^{-6}	4.1201×10^{-5}	5.3825×10^{-6}	1.0693×10^{-4}	534	10480	6610.2
1.0×10^{-2}	8.2946×10^{-6}	4.1156×10^{-5}	6.0342×10^{-7}	1.0670×10^{-4}	602	11479	6998.3
1.0×10^{-3}	8.1647×10^{-6}	4.1150×10^{-5}	4.5285×10^{-8}	1.0669×10^{-4}	636	12288	7181.7
1.0×10^{-4}	8.1577×10^{-6}	4.1150×10^{-5}	4.5439×10^{-9}	1.0669×10^{-4}	668	13178	7566.5

Table 6 Barenblatt problem (96)–(97), $m = 2$, scaling parameter $d_{K,m}$ given by (21a), numerical study of the algebraic stopping criterion (91)

c_A	η	$\mathcal{J}(u_h^\tau)$	η_{alg}	η_{spa}	N_{non}	N_{lin}	time(s)
$h = 0.35$, $p = 1$ & $q = 1$, #DoF = 13824							
1.0×10^{-1}	2.3055×10^{-3}	4.6659×10^{-3}	4.0757×10^{-4}	9.3084×10^{-3}	100	2199	224.6
1.0×10^{-2}	2.2689×10^{-3}	4.5703×10^{-3}	1.3712×10^{-5}	9.2211×10^{-3}	200	4957	284.5
1.0×10^{-3}	2.2688×10^{-3}	4.5694×10^{-3}	3.5878×10^{-6}	9.2216×10^{-3}	299	7715	352.6
1.0×10^{-4}	2.2688×10^{-3}	4.5696×10^{-3}	3.9773×10^{-7}	9.2220×10^{-3}	378	9790	413.7
$h = 0.35$, $p = 2$ & $q = 2$, #DoF = 27648							
1.0×10^{-1}	3.0332×10^{-4}	1.1855×10^{-3}	7.0279×10^{-5}	1.7702×10^{-3}	201	4065	1535.3
1.0×10^{-2}	2.9940×10^{-4}	1.1753×10^{-3}	7.2859×10^{-6}	1.7634×10^{-3}	286	5820	1675.9
1.0×10^{-3}	2.9916×10^{-4}	1.1747×10^{-3}	8.0717×10^{-7}	1.7619×10^{-3}	393	7803	1759.7
1.0×10^{-4}	2.9916×10^{-4}	1.1747×10^{-3}	8.9211×10^{-8}	1.7620×10^{-3}	529	10172	1984.6
$h = 0.35$, $p = 3$ & $q = 2$, #DoF = 46080							
1.0×10^{-1}	1.5523×10^{-4}	6.8813×10^{-4}	7.1710×10^{-5}	1.1705×10^{-3}	202	4222	5539.6
1.0×10^{-2}	1.5037×10^{-4}	6.6961×10^{-4}	5.7493×10^{-6}	1.1586×10^{-3}	316	6538	6068.2
1.0×10^{-3}	1.5026×10^{-4}	6.6968×10^{-4}	6.3387×10^{-7}	1.1580×10^{-3}	453	8880	6615.7
1.0×10^{-4}	1.5026×10^{-4}	6.6967×10^{-4}	5.9895×10^{-8}	1.1580×10^{-3}	591	11150	7101.3

Algorithm 1: Space-time mesh adaptive algorithm.

1: inputs: tolerance $\delta > 0$, initial time step τ_1, initial mesh $\mathcal{T}_h^1$ and space $S_{hp,1}$
2: set $m \leftarrow 1, t \leftarrow 0$
3: **while** $t < T$ **do**
4: **repeat**
5: solve $\sum_K B_{K,m}(u_h^\tau, v) = 0 \quad \forall v \in S_{hp}^{\tau q}$ for the given m, cf. (13)
6: evaluate $\eta_m(u_h^\tau)$ and set δ_m, cf. (41) and (102)
7: **if** $\eta_m \le \delta_m$ **then**
8: set $t \leftarrow t + \tau_m$, propose the new size of the time step τ_{m+1}
9: set $\mathcal{T}_h^{m+1} \leftarrow \mathcal{T}_h^m$, $S_{hp,m+1} \leftarrow S_{hp,m}$, $m \leftarrow m+1$
10: **else**
11: generate a new mesh $\mathcal{T}_h^m$ & space $S_{hp,m}$
12: reduce the size of the time step τ_m
13: **end if**
14: **until** $\eta_m \le \delta_m$
15: **end while**

grids and the higher degrees of polynomial approximation, where the effectivity indices i_{eff} are below 1. Based on the values of EoC, we suppose that i_{eff} below 1 is not caused by the failure of the error estimator but due to an inaccurate approximation $\widetilde{\mathcal{R}}(u_h^\tau)$ of the exact error; see Remark 3.

7 Mesh Adaptive Algorithm

We introduce the mesh adaptive algorithm which is based on the a posteriori error estimates η, cf. (41). Let $\delta > 0$ be the given tolerance, the goal of the algorithm is to define the sequence of time steps τ_m, meshes $\mathcal{T}_h^m$ and spaces $S_{hp,m}$, $m = 1, \dots, r$ such that the corresponding approximate solution $u_h^\tau \in S_{hp}^{\tau q}$ given by (13) satisfies the condition

$$\eta = \eta(u_h^\tau) \le \delta. \tag{101}$$

Another possibility is to require $\left(\eta^2 + \mathcal{J}(u_h^\tau)\right)^{1/2} \le \delta$, then the following considerations have to be modified appropriately.

The mesh adaptation strategy is built on the equi-distribution principle, namely the sequences $\{\tau_m,\ \mathcal{T}_h^m,\ S_{hp,m}\}_{m=1}^r$ should be generated such that

$$\eta_m \le \delta_m := \delta\sqrt{\tau_m/T} \qquad \forall m = 1, \dots, r, \tag{102a}$$

$$\eta_{K,m} \le \delta_{K,m} := \delta_m\sqrt{1/\#\mathcal{T}_h^m} \qquad \forall K \in \mathcal{T}_h^m\ \forall m = 1, \dots, r, \tag{102b}$$

where $\eta_m := (\sum_{K\in\mathcal{T}_h^m} \eta_{K,m}^2)^{1/2}$ is the error estimate corresponding to the time interval I_m, $m = 1, \dots, r$ and $\#\mathcal{T}_h^m$ denotes the number of elements of $\mathcal{T}_h^m$. Obviously, if all the conditions in (102) are valid, then the criterion (101) is achieved.

Based on (101)–(102), we introduce the abstract Algorithm 1. The size of $\tau_m, m = 1, \dots, r$ (step 8 of the algorithm) are chosen to equilibrate estimates of the spatial and temporal reconstruction, $\eta_{S,m} := (\sum_{K\in\mathcal{T}_h^m}(\eta_{S,K,m})^2)^{1/2}$ and $\eta_{T,m} := (\sum_{K\in\mathcal{T}_h^m}(\eta_{T,K,m})^2)^{1/2}$, cf. (42). Particularly, we set the new time step according to the formula

$$\tau_{m+1} = \tau_m c_F \left(\frac{\eta_{S,m}}{\eta_{T,m}}\right)^{1/(q+1)}, \qquad m = 1, \dots, r, \tag{103}$$

Table 7 Tracy problem scaling parameter $d_{K,m}$ given by (21b), approximation of the error and the error estimators, EOC in parenthesis

h	#DoF	$\widetilde{\mathcal{R}}(u_h^\tau)$	η	$\mathcal{J}(u_h^\tau)$	η_R	η_S	η_T	i_{eff}	$i_{\text{eff}}^{\text{tot}}$
p = 1 & q = 1									
0.18	384	2.43×10^{-1}	2.95×10^{-1}	6.62×102	2.25×10^{-3}	2.90×10^{-1}	4.83×10^{-2}	1.22	1.00
0.09	1536	8.77×10^{-2}	1.19×10^{-1}	2.47×102	9.92×10^{-4}	1.10×10^{-1}	4.36×10^{-2}	1.35	1.00
		(1.47)	(1.32)	(1.43)	(1.18)	(1.40)	(0.15)		
0.04	6144	1.50×10^{-2}	2.51×10^{-2}	4.68×101	1.33×10^{-4}	2.39×10^{-2}	7.31×10^{-3}	1.67	1.00
		(2.55)	(2.24)	(2.40)	(2.90)	(2.20)	(2.58)		
0.02	24576	7.34×10^{-3}	1.22×10^{-2}	2.27×101	8.11×10^{-5}	1.20×10^{-2}	1.86×10^{-3}	1.66	1.00
		(1.03)	(1.04)	(1.05)	(0.71)	(0.99)	(1.98)		
p = 2 & q = 2									
0.18	768	4.88×10^{-2}	6.18×10^{-2}	1.92×102	3.35×10^{-4}	6.04×10^{-2}	1.22×10^{-2}	1.27	1.00
0.09	3072	1.75×10^{-2}	1.98×10^{-2}	6.03×101	1.44×10^{-5}	1.95×10^{-2}	3.35×10^{-3}	1.13	1.00
		(1.48)	(1.65)	(1.67)	(4.54)	(1.63)	(1.86)		
0.04	12288	5.59×10^{-3}	6.28×10^{-3}	1.94×101	4.68×10^{-6}	6.08×10^{-3}	1.56×10^{-3}	1.12	1.00
		(1.64)	(1.65)	(1.64)	(1.62)	(1.68)	(1.11)		
0.02	49,152	1.90×10^{-3}	1.33×10^{-3}	4.37	2.41×10^{-6}	1.31×10^{-3}	1.87×10^{-4}	0.70	1.00
		(1.56)	(2.24)	(2.15)	(0.96)	(2.21)	(3.06)		
p = 3 & q = 2									
0.18	1280	2.24×10^{-2}	2.64×10^{-2}	9.60×101	2.42×10^{-5}	2.61×10^{-2}	4.48×10^{-3}	1.18	1.00
0.09	5120	6.26×10^{-3}	7.94×10^{-3}	2.77×101	1.21×10^{-5}	7.13×10^{-3}	3.48×10^{-3}	1.27	1.00
		(1.84)	(1.74)	(1.79)	(1.00)	(1.87)	(0.37)		
0.04	20480	1.40×10^{-3}	4.60×10^{-4}	1.63	2.87×10^{-6}	4.49×10^{-4}	8.90×10^{-5}	0.33	1.00
		(2.16)	(4.11)	(4.08)	(2.08)	(3.99)	(5.29)		
0.02	81920	1.37×10^{-3}	8.85×10^{-5}	3.08×10^{-1}	2.37×10^{-6}	8.59×10^{-5}	1.09×10^{-5}	0.06	1.00
		(0.03)	(2.38)	(2.41)	(0.28)	(2.39)	(3.03)		

Table 8 Barenblatt problem (96)–(97), scaling parameter $d_{K,m}$ given by (21b), the error estimators obtained by the adaptive computations using Algorithm 1

hp adaptation						
δ	#DoF	η	$\mathcal{J}(u_h^\tau)^{1/2}$	η_R	η_S	η_T
2.0E−03	4 543	7.82×10^{-4}	1.69×10^{-3}	2.18×10^{-4}	5.40×10^{-4}	3.23×10^{-4}
1.0E−03	6 244	4.57×10^{-4}	1.17×10^{-3}	1.48×10^{-4}	3.13×10^{-4}	1.43×10^{-4}
5.0E−04	9 071	2.10×10^{-4}	7.02×10^{-4}	6.75×10^{-5}	1.38×10^{-4}	7.79×10^{-5}

The quantity #DoF is the average number of space degrees of freedom per one time step

where $c_F \in (0, 1)$ is the security factor and $q \geq 0$ is the polynomial degree with respect to time. Therefore, $q + 1$ corresponds to the temporal order of convergence.

The construction of the new mesh (step 11 in Algorithm 1) is based on the modification of the anisotropic *hp*-mesh adaptation method from [15, 20]. Having the actual mesh $\mathcal{T}_h^m$, for each $K \in \mathcal{T}_h^m$ we set the new volume of K according the formula

$$v_K = |K| \Lambda(\delta_{K,m}/\eta_{K,m}), \qquad K \in \mathcal{T}_h^m, \tag{104}$$

where $\delta_{K,m}$ is the local tolerance from (102b), $|K|$ is the volume of $|K|$ and $\Lambda : \mathbb{R}^+ \to \mathbb{R}+$ is a suitable increasing function such that $\Lambda(1) = 1$. For particular variants of Λ, we refer to [15, 20].

When the new volume of mesh elements is established by (104), the new shape of K and a new polynomial approximation degree p_K are optimized by minimizing the interpolation error. This optimization is done locally for each mesh element. In one adaptation level, we admit the increase or decrease of p_K by one. Setting the new area, shape, and polynomial approximation degree for each element of the current mesh, we define the continuous mesh model [16] and carry out a remeshing using the code ANGENER [9].

The generated meshes are completely non-nested and non-matching, hence the evaluation of the time-penalty term (cf. Remark 1) is delicate. We refer to [20] where this aspect is described in detail and numerically verified. The presented numerical analysis takes into account the errors arising from the re-meshing in the temporal reconstruction R_h^τ, which contains term $\left\{\vartheta(u_h^\tau)\right\}_{m-1}$, cf. (26). The following numerical experiments show that the error estimator is under the control also after each remeshing.

7.1 Barenblatt Problem

We apply Algorithm 1 to the Barenblatt problem (96) with $m = 2$. Table 8 shows the error estimators obtained by adaptive computation for three different tolerances δ. Compared with the error estimators from Table 4, we observe that the adaptive computations achieve significantly smaller error estimates using a significantly smaller number of degrees of freedom. We note that we are not able to present the quantity $\widetilde{\mathcal{R}}$ (cf. (92)–(93)) approximating the error since the finite element code used for the evaluation of $\widetilde{\mathcal{R}}$ supports only uniform grids.

Figure 1 shows the performance of Algorithm 1, where each dot corresponds to one time step $m = 1, \ldots, r$. We plot the values of the accumulated estimators $\overline{\eta}_m = \sum_{i=1}^m \eta_i$ for all $m = 1, \ldots, r$. The red nodes correspond to all computed time steps, including the rejected ones (steps 11–12 of Algorithm 1) whereas the blue nodes mark only the accepted time steps. The rejected time steps indicate the re-meshing. Moreover, we plot the "accumulated" tolerance $\delta(t_m/T)^{1/2}$, cf. (101) and (102a). We observe that the resulting estimator η at $t = T$

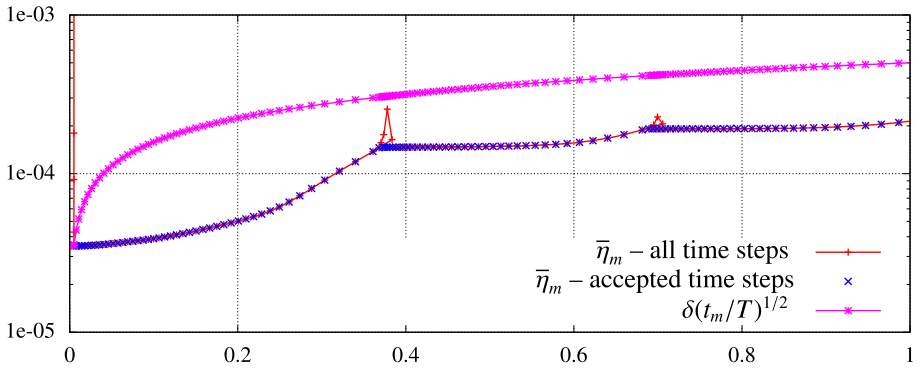


Fig. 1 Barenblatt problem, (96)–(97), $m = 2$, performance of Algorithm 1, accumulated error estimator $\overline{\eta}_m$ and the "accumulated" tolerance $\delta(t_m/T)^{1/2}$ for $m = 1, \ldots, r$

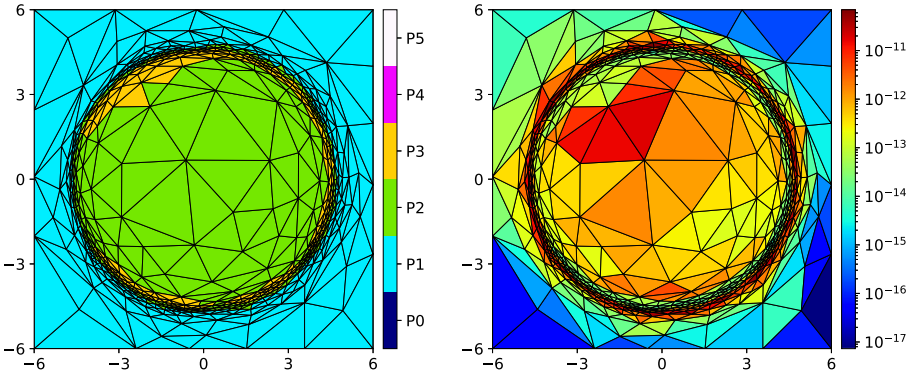


Fig. 2 Barenblatt problem, hp-mesh obtained by Algorithm 1 (left) and the error estimators $\eta_{K,m}$, $K \in \mathcal{T}_h^m$ at $T = 1$

is below the tolerance δ by a factor of approximately 2.5 since conditions (102) are stronger than (101).

Figure 2, left, shows the hp-mesh obtained by Algorithm 1 at the final time $T = 1$, each triangle is highlighted by a color corresponding to the polynomial degree used p_K, $K \in \mathcal{T}_h^m$. We observe a strong anisotropic refinement about the circular singularity of the solution when $u \to 0^+$, see the analytical formula (97). Outside of this circle, large triangles with the smallest polynomial degree ($p = 1$) are generated. On the other hand, due to the regularity of the solution in the interior of the circle, the polynomial degrees $p = 2$ or $p = 3$ are generated.

Moreover, Fig. 2, right, shows the error estimator $\eta_{K,m}$, $K \in \mathcal{T}_h^m$ at $T = 1$. The elements in the exterior of the circle have small values of $\eta_{K,m} \approx 10^{-17} - 10^{-14}$ due to a constant solution and negligible errors. On the other hand, the values of $\eta_{K,m}$ for the rest of elements $K \in \mathcal{T}_h^m$ are in the range 10^{-13}–10^{-11} due to the equidistant principle used.

7.2 Single Ring Infiltration

We deal with the numerical solution of the single ring infiltration experiment, which is frequently used for the identification of saturated hydraulic conductivity, cf. [32, 39] for example. We consider the Richards equation (3) where the active pore volume ϑ is given by (2), the water content function θ is given by the van Genuchten's law [27] and the conductivity $\mathbf{K}(u) = \mathbf{K}_s \mathcal{K}_r(u)$ is given by the Mualem function [31], namely

$$\theta(u) = \begin{cases} \frac{\theta_s - \theta_r}{(1+(-\alpha\psi)^n)^m} + \theta_r & \text{for } \psi < 0, \\ \theta_s & \text{for } \psi \geq 0, \end{cases}$$

$$\mathcal{K}_r(u) = \begin{cases} \frac{(1-(-\alpha\psi)^{m\,n}(1+(-\alpha\psi)^n)^{-m})^2}{(1+(-\alpha\psi)^n)^{m/2}} & \text{for } \psi < 0, \\ 1 & \text{for } \psi \geq 0, \end{cases} \tag{105}$$

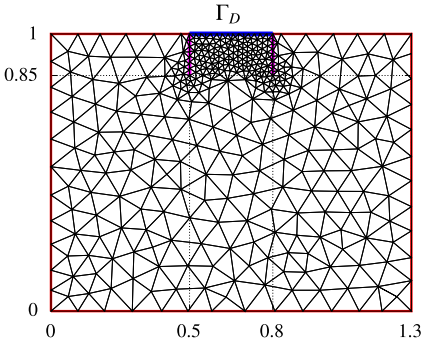
where $\psi = u - z$ is the pressure head, z is the vertical coordinate and the material parameters $\mathbf{K}_s = 0.048\,\mathbb{I}\ \mathrm{m} \cdot \mathrm{hours}^{-1}$, $\theta_s = 0.55$, $\theta_r = 0.0$, $\alpha = 0.8\,\mathrm{m}^{-1}$, $n = 1.2$, $m = 1/6$ and $S_s = 10^{-3}\,\mathrm{m}^{-1}$ (cf. (2)).

The computational domain together with the boundary parts is sketched in Fig. 3a. On the boundary part Γ_D we set the Dirichlet boundary condition $u = 1.05\,\mathrm{m}$, and on $\Gamma_\mathrm{N} = \Gamma \setminus \Gamma_\mathrm{D}$ we consider the homogeneous Neumann boundary condition. The smaller "magenta" vertical lines starting at Γ_D belong to Γ_N. At $t = 0$, a dry medium with $u = \psi + z = -2\,\mathrm{m}$ is prescribed. We carried out the computation until the physical time $T = 2\,\mathrm{hours}$. The inconsistency of the initial and boundary condition on Γ_D makes the computation quite difficult for $t \approx 0$.

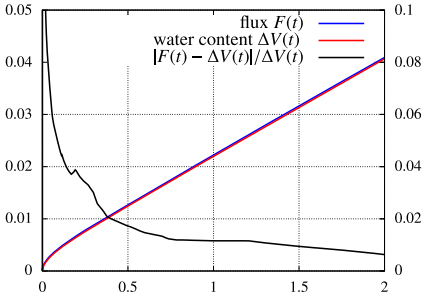
Figure 3b verifies the conservativity of the adaptive method. We plot the quantities

$$F(t) = \int_0^t \int_\Gamma \mathbf{K}(u)\nabla u \cdot n \,\mathrm{d}S\,\mathrm{d}t,$$
$$\Delta V(t) = V(t) - V(0), \quad V(t) = \int_\Omega \vartheta(u(\cdot,t))\,\mathrm{d}x, \ t \in [0,T], \tag{106}$$

where $F(t)$ is the total flux of the water through the boundary Γ till time t and $\Delta V(t)$ is the changes of the water content in the domain between times 0 and t. From equation (3) and the



(a) Initial mesh, the Dirichlet boundary Γ_D and the Neumann boundary $\Gamma_\mathrm{N} = \Gamma \setminus \Gamma_\mathrm{D}$ (red and magenta).



(b) Quantities $F(t)$ and $\Delta V(t)$, $t \in (0,2)$, cf. (106) and their relative difference (the right-hand side vertical axis).

Fig. 3 Single ring infiltration problem

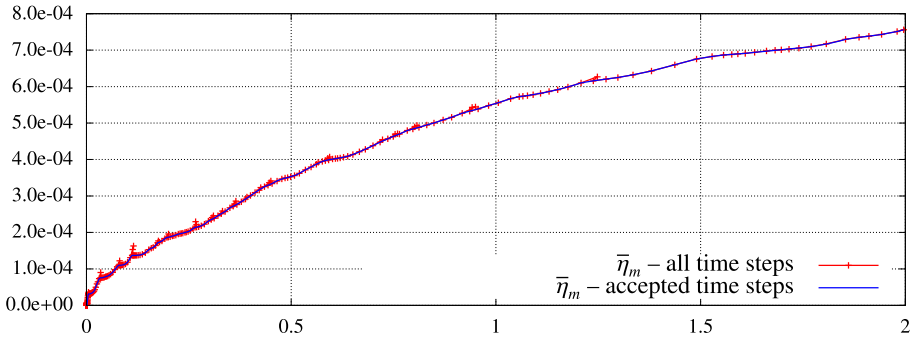


Fig. 4 Single ring infiltration, performance of Algorithm 1, accumulated error estimator $\overline{\eta}_m$ with respect to t_m, $m = 1, \ldots, r$

Stokes theorem, we have the conservation law $F(t) = \Delta V(t)$ for all $t \in [0, T]$. Therefore, we also show the relative difference between these quantities $|F(t) - \Delta V(t)|/\Delta V(t)$ for $t > 0$ in Fig. 3b the vertical label on the right. We observe that, except for the time close to zero, where the inconsistency between initial and boundary conditions is problematic, the relative difference is at the level of several percent.

Furthermore, Fig. 4 shows the accumulated estimators $\overline{\eta}_m = \sum_{i=1}^{m} \eta_i$ for time levels t_m, $m = 1, \ldots, m$. The red nodes correspond to all computed time steps, including the rejected steps whereas the blue line connects only the accepted time steps. The rejected time steps are followed by the remeshing which is carried out namely for small t. We observe that the elimination of the rejected time steps causes that the errors arising from the remeshing do not essentially affect the total error estimate η.

Moreover, Fig. 5 shows the hp-meshes, the hydraulic head and the error estimator $\eta_{K,m}$, $K \in \mathcal{T}_h^m$ at selected time levels obtained from Algorithm 1 with $\delta = 5.0 \times 10^{-3}$. We observe the mesh adaptation namely at the (not sharp) interface between the saturated and non-saturated medium and also in the vicinity of the domain singularities. The error estimators $\eta_{K,m}$, $K \in \mathcal{T}_h^m$ indicate an equi-distribution of the error.

8 Conclusion

We derived reliable and efficient a posteriori error estimates in the residual-based norm for the Richards equation discretized by the space-time discontinuous Galerkin method. The numerical verification indicates the effectivity indexes between 1 and 2.5 for the tested examples. Moreover, we introduced the hp-mesh adaptive method handling varying non-nested and non-matching meshes and demonstrated its efficiency for simple test benchmark and its applicability for the numerical solution of the single ring infiltration experiment.

It will be possible to generalize the presented approach to genuinely space-time hp-adaptive method, where the (local) polynomial order q in time is varied as well. However, the question is of potential benefit. Based on our experience, the setting $q = 1$ gives sufficiently accurate approximation for the majority of tested problems.

On the other hand, the choice $q = 0$ would be sufficient only in subdomains of Ω where the solution is almost constant in time. Therefore, we suppose that the benefit of local varying of polynomial order in time will be low.

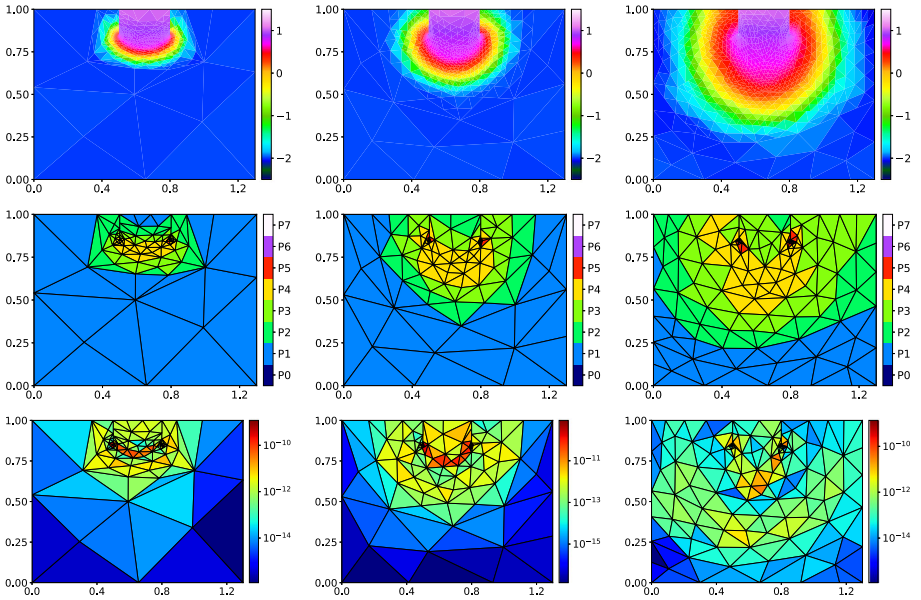


Fig. 5 Single ring infiltration, hydraulic head (top), the corresponding hp-meshes obtained by Algorithm 1 (center) and the error estimators $\eta_{K,m}$, $K \in \mathcal{T}_h^m$ (bottom) at $t = 0.4$, $t = 0.8$ and $t = 2$ hours (from left to right)

Although the presented numerical examples are two-dimensional, it would be possible to apply the presented error estimates and mesh adaptation to three-dimensional problems as well. We refer, e.g., to [1] and the references therein, where the anisotropic mesh adaptation techniques are developed for time-dependent 3D problems.

Funding Open access publishing supported by the National Technical Library in Prague. This work has been supported by the Czech Science Foundation Grant No. 20-01074 S (V.D.), the Charles University grant SVV-2023-260711, and the Grant Agency of Charles University Project No. 28122 (H.S.), European Development Fund-Project "Center for Advanced Aplied Science" No. CZ.02.1.01/0.0/0.0./16 019/0000778 (M.V.). V.D. acknowledges the membership in the Nečas Center for Mathematical Modeling ncmm.karlin.mff.cuni.cz.

Data Availability No datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Alauzet, F., Loseille, A., Olivier, G.: Time-accurate multi-scale anisotropic mesh adaptation for unsteady flows in CFD. J. Comput. Phys. **373**, 28–63 (2018)
2. Alt, H.W., Luckhaus, S.: Quasilinear elliptic-parabolic differential equations. Math. Z. **183**(3), 311–341 (1983)
3. Baradi, M., Difonzo, F.V.: Strong solutions for Richards' equation with Cauchy conditions and constant pressure. Environ. Fluid Mech. **20**, 165–174 (2020)
4. Barenblatt, G.I.: On some unsteady motions of a liquid and gas in a porous medium. Akad. Nauk SSSR. Prikl. Mat. Meh. **16**, 67–78 (1952)
5. Baron, V., Coudière, Y., Sochala, P.: Adaptive multistep time discretization and linearization based on a posteriori error estimates for the Richards equation. Appl. Numer. Math. **112**, 104–125 (2017)
6. Bernardi, C., El Alaoui, L., Mghazli, Z.: A posteriori analysis of a space and time discretization of a nonlinear model for the flow in partially saturated porous media. IMA J. Numer. Anal. **34**(3), 1002–1036 (2014)
7. Brezzi, F., Fortin, M.: Mixed and Hybrid Finite Element Methods, Springer Series in Computational Mathematics, vol. 15. Springer, New York (1991)
8. Di Pietro, D.A., Vohralík, M., Yousef, S.: An a posteriori-based, fully adaptive algorithm with adaptive stopping criteria and mesh refinement for thermal multiphase compositional flows in porous media. Compu. Math. Appl. **68**(12, Part B), 2331–2347 (2014)
9. Dolejší, V.: ANGENER – Anisotropic mesh generator, in-house code. Charles University, Prague, Faculty of Mathematics and Physics (2000). https://msekce.karlin.mff.cuni.cz/~dolejsi/angen/
10. Dolejší, V.: ADGFEM – Adaptive discontinuous Galerkin finite element method, in-house code. Charles University, Prague, Faculty of Mathematics and Physics (2020). https://msekce.karlin.mff.cuni.cz/~dolejsi/adgfem/
11. Dolejší, V., Ern, A., Vohralík, M.: A framework for robust a posteriori error control in unsteady nonlinear advection-diffusion problems. SIAM J. Numer. Anal. **51**(2), 773–793 (2013)
12. Dolejší, V., Ern, A., Vohralík, M.: *hp*-adaptation driven by polynomial-degree-robust a posteriori error estimates for elliptic problems. SIAM J. Sci. Comput. **38**(5), A3220–A3246 (2016)
13. Dolejší, V., Feistauer, M.: Discontinuous Galerkin Method: Analysis and Applications to Compressible Flow. Springer Series in Computational Mathematics, vol. 48. Springer, Cham (2015)
14. Dolejší, V., Kuráž, M., Solin, P.: Adaptive higher-order space-time discontinuous Galerkin method for the computer simulation of variably-saturated porous media flows. Appl. Math. Model. **72**, 276–305 (2019)
15. Dolejší, V., May, G.: Anisotropic *hp*-Mesh Adaptation Methods. Birkhäuser, Cham (2022)
16. Dolejší, V., May, G., Rangarajan, A.: A continuous *hp*-mesh model for adaptive discontinuous Galerkin schemes. Appl. Numer. Math. **124**, 1–21 (2018)
17. Dolejší, V., Roskovec, F., Vlasák, M.: Residual based error estimates for the space-time discontinuous Galerkin method applied to the compressible flows. Comput. Fluids **117**, 304–324 (2015)
18. Dolejší, V., Roskovec, F., Vlasák, M.: A posteriori error estimates for higher order space-time Galerkin discretizations of nonlinear parabolic problems. SIAM J. Numer. Anal. **59**(3), 1486–1509 (2021)
19. Dolejší, V., Shin, H.G.: A posteriori error estimate and mesh adaptation for the numerical solution of the richards equation. In: J.M. Melenk, et al. (eds.) Spectral and High Order Methods for Partial Differential Equations ICOSAHOM 2020+1, Lecture Notes in Computational Science and Engineering 137 (2023)
20. Dolejší, V., May, G.: An anisotropic *hp*-mesh adaptation method for time-dependent problems based on interpolation error control. J. Sci. Comput. **95**(2) (2023)
21. Ern, A., Guermond, J.L.: Theory and Practice of Finite Elements. Springer, Berlin (2004)
22. Ern, A., Smears, I., Vohralík, M.: Guaranteed, locally space-time efficient, and polynomial-degree robust a posteriori error estimates for high-order discretizations of parabolic problems. SIAM J. Numer. Anal. **55**(6), 2811–2834 (2017)
23. Ern, A., Vohralík, M.: Adaptive inexact Newton methods with a posteriori stopping criteria for nonlinear diffusion PDEs. SIAM J. Sci. Comput. **35**(4), A1761–A1791 (2013)
24. Ern, A., Vohralík, M.: Polynomial-degree-robust a posteriori estimates in a unified setting for conforming, nonconforming, discontinuous Galerkin, and mixed discretizations. SIAM J. Numer. Anal. **53**(2), 1058–1081 (2015)
25. Farthing, M., Ogden, F.: Numerical solution of Richards' equation: a review of advances and challenges. Soil Sci. Soc. Am. J. **81**(6), 1257–1269 (2017)
26. Gardner, W.R.: Some steady state solutions of the unsaturated moisture flow equation with application to evaporation from a water table. Soil Sci. **85**, 228–232 (1958)
27. van Genuchten, M.T.: Closed-form equation for predicting the hydraulic conductivity of unsaturated soils. Soil Sci. Soc. Am. J. **44**(5), 892–898 (1980)

28. Kim, I.C., Požár, N.: Nonlinear elliptic-parabolic problems. Arch. Ration. Mech. Anal. **210**, 975–1020 (2013)
29. Mallik, G., Vohralík, M., Yousef, S.: Goal-oriented a posteriori error estimation for conforming and nonconforming approximations with inexact solvers. J. Comput. Appl. Math. **366** (2020)
30. Mitra, K., Vohralík, M.: A posteriori error estimates for the Richards equation. Tech. Rep. hal-03328944v2, INRIA (2022)
31. Mualem, Y.: A new model for predicting the hydraulic conductivity of unsaturated porous media. Water Resour. Res. **12**(3), 513–522 (1976)
32. Nakhaei, M., Šimůnek, J.: Parameter estimation of soil hydraulic and thermal property functions for unsaturated porous media using the hydrus-2d code. J. Hydrol. Hydromech. **62**(1) (2014)
33. Richards, L.A.: Capillary conduction of liquids through porous mediums. J. Appl. Phys. **1**(5), 318–333 (1931)
34. Rudin, W.: Real and Complex Analysis, 3rd edn. McGraw-Hill, New York (1987)
35. Tracy, F.T.: Clean two- and three-dimensional analytical solutions of Richards equation for testing numerical solvers. Water Resour Res **42**(8) (2006)
36. Verfürth, R.: A Posteriori Error Estimation Techniques for Finite Element Methods. Numerical Mathematics and Scientific Computation. Oxford University Press, Oxford (2013)
37. Vohralík, M.: A posteriori error estimates for efficiency and error control in numerical simulations. Universite Pierre et Marie Curie – Paris 6 (2018). Lecture notes Course NM497
38. Vohralík, M., Yousef, S.: A simple a posteriori estimate on general polytopal meshes with applications to complex porous media flows. Comput. Methods Appl. Mech. Eng. **331**, 728–760 (2018)
39. Xu, X., Lewis, C., Liu, W., Albertson, J., Kiely, G.: Analysis of single-ring infiltrometer data for soil hydraulic properties estimation: comparison of BEST and Wu methods. Agric. Water Manag. **107**, 34–41 (2012)