SPACE–TIME ADAPTATION FOR PURELY DIFFUSIVE PROBLEMS IN AN ANISOTROPIC FRAMEWORK

STEFANO MICHELETTI AND SIMONA PEROTTO

(Communicated by Paul Houston)

This paper is dedicated to Mario

Abstract. The main goal of this work is the proposal of an efficient space-time adaptive procedure for a cGdG approximation of an unsteady diffusion problem. We derive a suitable a posteriori error estimator where the contribution of the spatial and of the temporal discretization is kept distinct. In particular our interest is addressed to phenomena characterized by temporal multiscale as well as strong spatial directionalities. On the one hand we devise a sound criterion to update the time step, able to follow the evolution of the problem under investigation. On the other hand we exploit an anisotropic triangular adapted grid. The reliability and the efficiency of the proposed error estimator are assessed numerically.

Key Words. space-time adaptation, anisotropic meshes, heat equation, space-time finite elements.

1. Introduction and motivations

We are interested in devising an effective procedure for selecting both the time step and the grid size to approximate unsteady diffusion problems. Such models are employed to describe applications of interest in, e.g., heat flow problems, hydrogeology, particle diffusion phenomena, etc.

We propose an adaptation algorithm based on both an adaptive choice of the temporal step and an anisotropic mesh adaptation strategy. Indeed the applications we have in mind often exhibit both temporal multiscale phenomena and spatial heterogeneities. The first issue calls for a time step, fitting the evolution of the phenomenon at hand; the second occurs in the presence of lower dimensional features of the domain, of the problem data or of the solution.

The proposed adaptation procedure relies on a theoretical tool, i.e., an a posteriori error estimator, driving the automatic choice of the spatial and temporal steps. In more detail, the key point is to identify, in the error estimator, a space and a time contribution. This is aligned with other works such as, e.g., R. Verfürth [51], J.M. Cascón, L. Ferragut & M.I. Asensio [9], D. Meidner & B. Vexler [33], M. Schmich & B. Vexler [44]. The additional value of our work is the possibility of dealing also with problems characterized by strong spatial directional features by means of an economic mesh from the computational viewpoint. It is in fact well known

that, by better orienting the mesh elements according to the main features of the solution, it is possible to maximize the solution accuracy for a fixed number of elements, rather than reduce the number of degrees of freedom for a fixed solution accuracy (see, e.g., M.J. Castro–Díaz, F. Hecht, B. Mohammadi & O. Pironneau [10], J. Dompierre, M.-G. Vallet, Y. Bourgault, M. Fortin & W.G. Habashi [18], T. Apel [2], D.L. Darmofal & D.A. Venditti [15], L. Formaggia & S. Perotto [22], K.G. Siebert [47], E.H. Georgoulis, E. Hall & P. Houston [24]).

In addition, the proposed adaptive algorithm carries out the spatial adaptation via an optimization procedure, in contrast to the more familiar mark-refine approach (Z. Chen & J. Feng [11], [9, 33, 44]). On the other hand the temporal step is adaptively updated driven by the time residual contribution rather than by a standard fixed-ratio reduction, as, e.g., in A. Schmidt & K.G. Siebert [45], [11, 9, 33], M. Picasso [41]. Alternative methods to drive the space-time adaptation are provided, for instance, in B. Cockburn & C.-W. Shu [14], R. Hartmann & P. Houston [25], D. Kröner & M. Ohlberger [30], J.J. Sudirham, J.J.W. van der Vegt & R.M.J. van Damme [48].

To approximate the problem under investigation we adopt a discretization scheme appropriate to manage space and time in parallel, i.e., based on space-time finite elements. In particular we choose finite elements continuous in space but discontinuous in time, i.e., the so-called cGdG scheme (K. Eriksson, C. Johnson & V. Thomée [21], K. Eriksson, D. Estep, P. Hansbo & C. Johnson [19]).

In compliance with G. Akrivis, C. Makridakis & R. Nochetto [1], M. Picasso [40], K. Eriksson & C. Johnson [20] we are interested in controlling a suitable global norm of the space-time error discretization, in our case through an anisotropic residual-based like a posteriori estimator. Even though the discretization framework that we provide is quite general and suited for a generic cGdG method, the anisotropic a posteriori analysis is confined to the cG(1)dG(0) scheme due to the particular anisotropic setting employed ([22], L. Formaggia & S. Perotto [23]).

The main difference with respect to a corresponding isotropic error estimator (R. Verfürth [50], [51]) is the presence of an anisotropic contribution weighting the standard residuals, and depending on the actual error. To preserve the anisotropic information without giving up the computability of the estimator, we resort to a suitable recovery approach: as for the space, we employ the well-performing Zienkiewicz-Zhu gradient recovery procedure (O.C. Zienkiewicz & J.Z. Zhu [52, 53, 54]); the time recovery is based on the idea in [33].

The choice of resorting to the Zienkiewicz-Zhu methodology can be attributed to various factors: the method is rather independent of the problem, of the governing equations and of most details of the finite element formulation (except for the finite element space); it is cheap to compute, easy to implement and works very well in practice (see, e.g., J.C. Bruch [6], K.L. Lawrence & R.V. Nambiar [31], T.P. Pawlak, M.J. Wheeler & S.M. Yunus [39]). Moreover we have already exploited such an approach also in an anisotropic framework with successful results (see, e.g., G. Maisano, S. Micheletti, S. Perotto & C.L. Bottasso [32], S. Micheletti & S. Perotto [34], L. Dedè, S. Micheletti & S. Perotto [17]).

As far as we know, the only paper available in the current literature dealing with an anisotropic management of triangular grids in a time-dependent framework is [41]. However in this last work the time adaptation issue is completely skipped. The author simply introduces the successive halving of the time step so that the time discretization contribution becomes negligible with respect to the spatial one. Moreover in this case a standard backward Euler scheme is adopted to discretize

the time. No particular care is finally devoted to the interplay between successive temporal intervals.

The paper is structured as follows. After presenting in Section 2 the reference problem and the corresponding cGdG formulation, we introduce in Section 3 the leading ideas of the anisotropic setting we refer to. Section 4 is devoted to the derivation of an anisotropic estimate for the space-time discretization error associated with the cG(1)dG(0) scheme, and to the proposal of a corresponding a posteriori (i.e., fully computable) error estimator. Then an effective space-time adaptive procedure, "converting" the anisotropic error estimator into a criterion to select both the space and time discretization, is introduced; the reliability and the efficiency of such a procedure is numerically investigated on three different test cases in Section 5. Finally some conclusions are drawn in the last section.

2. The reference problem

We introduce the model problem we shall be referring to in devising the proposed space-time adaptive procedure. For this purpose we first set up the notation adopted throughout the paper for the function spaces. With reference to the spatial independent variable only, we introduce the space $L^2(\omega)$ of the Lebesgue square-integrable functions on $\omega \subset \mathbb{R}^d$, for $d = 1, 2$, and with norm $\| \cdot \|_{L^2(\omega)}$. Then we identify the Sobolev space $H^1(\omega)$ of the functions which are in $L^2(\omega)$ along with the corresponding distributional derivatives of order one. As a suitable subspace of $H^1(\omega)$ we consider the space $H^1_\Gamma(\omega)$ of the functions in $H^1(\omega)$ vanishing (in the sense of trace) on a subset $\Gamma \neq \emptyset$ of $\partial\omega$. We define apart the space $L^\infty(\omega)$ of the functions bounded a.e. in ω. For a generic Hilbert space V, let V' be the corresponding dual space, and let ${}_{V'}\langle \cdot, \cdot \rangle_V$ denote the duality pairing between V' and V.

Moving to space-time functions $v = v(\mathbf{x}, t)$, we introduce the spaces

$$L^2(0, T; V) = \left\{ v : (0, T) \to V \; : \; \int_0^T \|v(t)\|_V^2 \, dt < +\infty \right\},$$

$$H^1(0, T; V) = \left\{ v, \; \frac{\partial v}{\partial t} \in L^2(0, T; V) \right\},$$

$$C^0([0, T]; V) = \{ v : [0, T] \to V \text{ continuous} \; : \; \forall t \in [0, T] \; \|v(t)\|_V < +\infty \},$$

with V a given Hilbert space, combining requirements on both the space and time smoothness. For a complete treatment, we refer to R. Dautray & J.-L. Lions [16].

We identify the reference model with the parabolic problem

$$(1) \quad \begin{cases} \alpha \dfrac{\partial u}{\partial t} - \nabla \cdot (D \nabla u) = f & (\mathbf{x}, t) \in \mathcal{Q} = \Omega \times J, \\[2mm] u(\mathbf{x}, t) = 0 & (\mathbf{x}, t) \in \partial\mathcal{Q}_D = \Gamma_D \times J, \\[2mm] D \nabla u \cdot \mathbf{n} = g & (\mathbf{x}, t) \in \partial\mathcal{Q}_N = \Gamma_N \times J, \\[2mm] u(\mathbf{x}, 0) = u_0(\mathbf{x}) & \mathbf{x} \in \Omega, \end{cases}$$

where Ω is a polygonal domain in $\mathbb{R}^2$ with Lipschitz continuous boundary $\partial\Omega$, Γ_D and Γ_N are measurable nonoverlapping partitions of $\partial\Omega$ such that $\partial\Omega = \overline{\Gamma}_D \cup \overline{\Gamma}_N$ and $\overset{\circ}{\Gamma}_D \cap \overset{\circ}{\Gamma}_N = \emptyset$, $J = (0, T)$ is the time interval, and $D \nabla u \cdot \mathbf{n}$ is the conormal derivative of u, with $\mathbf{n}$ the unit outward normal vector to $\partial\Omega$. The following assumptions are advanced on the data: the source $f \in L^2(0, T; L^2(\Omega))$; the conductivity tensor

$D = [d_{ij}] \in [L^\infty(\Omega)]^{2 \times 2}$ satisfies the uniform ellipticity condition

$$\sum_{i,j=1}^{2} d_{ij}(\mathbf{x}) \xi_i \, \xi_j \geq \gamma |\boldsymbol{\xi}|^2 \quad \forall \boldsymbol{\xi} = (\xi_1, \xi_2)^T \in \mathbb{R}^2 \quad \text{a.e. } \mathbf{x} \text{ in } \Omega$$

with $\gamma > 0$; the storage capacity $\alpha \in L^\infty(\Omega)$, with $\alpha > 0$ a.e. in Ω; the initial datum $u_0 \in L^2(\Omega)$ while the Neumann one $g \in L^2(0, T; L^2(\Gamma_N))$.
Problem (1) models, for instance, the heat diffusion in heat transfer theory or the drawdown of the hydraulic head observed at a well in an aquifer, under Darcy's law (we refer to Section 5 and S. Micheletti & S. Perotto [35] for possible examples). The weak formulation associated with (1) can thus be set up: find $u \in U = L^2(0, T; H^1_{\Gamma_D}(\Omega)) \cap H^1(0, T; (H^1_{\Gamma_D}(\Omega))')$ such that

$$(2) \quad \int_{\mathcal{Q}} \left\{ \alpha \frac{\partial u}{\partial t} v + D \nabla u \cdot \nabla v \right\} d\mathbf{x} \, dt = \int_{\mathcal{Q}} f v \, d\mathbf{x} \, dt + \int_{\partial \mathcal{Q}_N} g v \, ds \, dt \quad \forall v \in U,$$

with $u(\mathbf{x}, 0) = u_0(\mathbf{x})$.
It turns out that U is continuously embedded in $C^0([0, T]; L^2(\Omega))$ (see, e.g., [16]), thus assuring the temporal continuity to the weak solution u in (2). Moreover, the Hilbert spaces $H^1_{\Gamma_D}(\Omega)$ and $L^2(\Omega)$ constitute with the dual space $(H^1_{\Gamma_D}(\Omega))'$ a Gelfand triple, i.e., the continuous injections $H^1_{\Gamma_D}(\Omega) \hookrightarrow L^2(\Omega) \hookrightarrow (H^1_{\Gamma_D}(\Omega))'$ hold. The inner product on $L^2(\Omega)$ is consequently an equivalent representation of the duality pairing between $H^1_{\Gamma_D}(\Omega)$ and its dual $(H^1_{\Gamma_D}(\Omega))'$.

A weak formulation alternative to (2) is provided in (5) with a view to the discretization.

2.1. The space-time discretization. The most straightforward discretization scheme with a view to deriving an error estimator discriminating the spatial from the temporal contribution resorts to space-time finite elements (see, the pioneer papers P. Jamet [29], [21], A.K. Aziz & P. Monk [3], as well as reference works like, e.g., [19], V. Thomée [49]). In particular we select finite elements continuous in space but discontinuous in time (a so-called cGdG scheme). This discontinuity allows a thorough freedom in choosing the spatial grids as well as a reduction of the computational costs.

We first manage the time discretization. We partition the t-axis via the time levels

$$(3) \qquad 0 \equiv t_0 < t_1 < \ldots < t_{n-1} < t_n < \ldots < t_{N-1} < t_N \equiv T,$$

thus identifying the generic time interval $J_n = (t_{n-1}, t_n]$, of width $k_n = t_n - t_{n-1}$, and the n-th space-time slab $S_n = \Omega \times J_n$, with $n = 1, \ldots, N$. Notice that partition (3) is not necessarily uniform.
In the spirit of a cGdG approximation, we look for an approximate solution to (2) coinciding, on each subinterval J_n, with a polynomial of degree at most q in t and with coefficients in $H^1_{\Gamma_D}(\Omega)$, i.e., a function belonging to the space

$$\mathcal{S} = \left\{ v : (0, T] \to H^1_{\Gamma_D}(\Omega) : v(\mathbf{x}, t)\big|_{J_n} = \sum_{j=0}^{q} v_j(\mathbf{x}) \, t^j, \text{ with } v_j \in H^1_{\Gamma_D}(\Omega) \right\},$$

where $q \geq 0$ is a given integer. These functions are allowed to be discontinuous at each time level, with continuity from the left. Since $0 \notin J_1$, the value $v(\mathbf{x}, 0)$ has to

be specified separately, for any $v \in \mathcal{S}$. Moreover while $\mathcal{S} \not\subset U$, $\mathcal{S}\big|_{S_n} \subset U\big|_{S_n}$ where

$$\mathcal{S}\big|_{S_n} = \left\{ v(\mathbf{x}, t) = \sum_{j=0}^{q} v_j(\mathbf{x})\, t^j, \text{ with } v_j \in H^1_{\Gamma_D}(\Omega) \text{ and } t \in J_n \right\}$$

and $U\big|_{S_n} = L^2(J_n; H^1_{\Gamma_D}(\Omega)) \cap H^1(J_n; (H^1_{\Gamma_D}(\Omega))')$.

The possible time discontinuity leads us to distinguish between the two values

$$v_m^+ = \lim_{t \to 0^+} v(\mathbf{x}, t_m + t) \quad \text{and} \quad v_m^- = \lim_{t \to 0^+} v(\mathbf{x}, t_m - t),$$

and to define the *temporal jump*

(4)
$$[v]_m = v_m^+ - v_m^- \quad \text{with} \quad m = 0, \dots, N-1.$$

We stress that the jumps $[v]_m$ in (4) are identically equal to zero when $v \in U$. Thanks to this remark and by splitting the time integral onto the subintervals J_n's, we provide an alternative weak formulation, equivalent to (2): find $u \in U$ such that

(5)
$$\sum_{n=1}^{N} \int_{S_n} \left\{ \alpha \frac{\partial u}{\partial t} v + D\nabla u \cdot \nabla v \right\} d\mathbf{x}\, dt + \sum_{m=1}^{N-1} \int_{\Omega} \alpha [u]_m\, v_m^+\, d\mathbf{x} + \int_{\Omega} \alpha u_0^+\, v_0^+\, d\mathbf{x}$$

$$= \sum_{n=1}^{N} \left\{ \int_{S_n} fv\, d\mathbf{x}\, dt + \int_{\partial \mathcal{Q}_N^n} gv\, ds\, dt \right\} + \int_{\Omega} \alpha u_0^-\, v_0^+\, d\mathbf{x} \quad \forall v \in U,$$

where $u_0^+ = u_0^- = u_0(\mathbf{x})$ and $\partial \mathcal{Q}_N^n = \Gamma_N \times J_n$, for $n = 1, \dots, N$. The left-hand side of (5) identifies on the space $W = U \cup \mathcal{S}$ the bilinear form

$$B_{\mathrm{DG}}(w, z) = \sum_{n=1}^{N} \int_{S_n} \left\{ \alpha \frac{\partial w}{\partial t} z + D\nabla w \cdot \nabla z \right\} d\mathbf{x}\, dt + \sum_{m=1}^{N-1} \int_{\Omega} \alpha [w]_m\, z_m^+\, d\mathbf{x} + \int_{\Omega} \alpha w_0^+\, z_0^+\, d\mathbf{x},$$

for any $w, z \in W$. With plain calculations it can be verified that

(6)
$$B_{\mathrm{DG}}(w, w) = \sum_{n=1}^{N} \int_{J_n} \| D^{1/2} \nabla w \|^2_{L^2(\Omega)}\, dt + \frac{1}{2} \sum_{m=1}^{N-1} \| \alpha^{1/2} (w_m^+ - w_m^-) \|^2_{L^2(\Omega)}$$

$$+ \frac{1}{2} \| \alpha^{1/2} w_0^+ \|^2_{L^2(\Omega)} + \frac{1}{2} \| \alpha^{1/2} w_N^- \|^2_{L^2(\Omega)} \quad \forall w \in W,$$

i.e., the bilinear form $B_{\mathrm{DG}}(\cdot, \cdot)$ induces the norm $||| \cdot |||_{\mathrm{DG}} = (B_{\mathrm{DG}}(\cdot, \cdot))^{1/2}$ on the space W. Notice that the second summation in (6) vanishes when $w \in U$.

Let us settle now the spatial discretization. We resort to a family of conformal decompositions of $\overline{\Omega}$ into triangles, such that there exists always a vertex of the triangulation at the interface between Γ_D and Γ_N (see, e.g., Ph. Ciarlet [12]). The temporal discontinuity allows the employment of different families of meshes on each space-time slab S_n, for $n = 1, \dots, N$ (see Figure 1, left). In particular, for any time interval J_n, we set $\mathcal{T}_{h_n} = \{K_n\}$, with K_n triangle of diameter h_{K_n} and $h_n = \max_{K_n} h_{K_n}$, while we denote with $S_{K_n} = K_n \times J_n$ and $L_{K_n} = \partial K_n \times J_n$ the space-time prism associated with the element K_n and the corresponding lateral surface, respectively (see Figure 1, right).

We are in a position to introduce the discrete space

$$\mathcal{S}_h = \left\{ v_h \in \mathcal{S} : v_h(\mathbf{x}, t)\big|_{J_n} = \sum_{j=0}^{q} \psi_j(\mathbf{x})\, t^j, \text{ with } \psi_j \in X^r_{h_n} \cap H^1_{\Gamma_D}(\Omega) \right\} \subset \mathcal{S},$$

i.e., the cG(r)dG(q) space (see [21, 19]), where $X_{h_n}^r$ is the space of the finite elements of degree r associated with the mesh $\mathcal{T}_{h_n}$ (see, e.g., [12]). We highlight that the functions $v_h \in \mathcal{S}_h$ preserve the spatial continuity as well as the time discontinuity proper of $\mathcal{S}$.
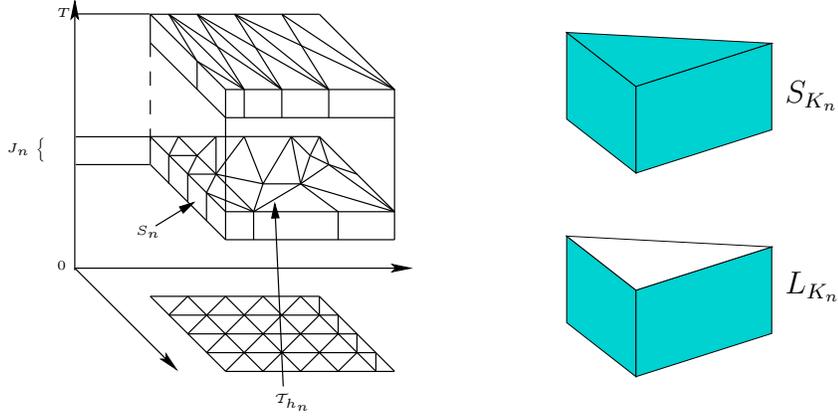

FIGURE 1. Temporal discontinuity of the spatial meshes (left); the space-time prism S_{K_n} and the corresponding lateral surface L_{K_n} (right).

The cG(r)dG(q) formulation of problem (1) can thus be stated as the discrete counterpart of the weak form (5): find $u_h \in \mathcal{S}_h$ such that

$$
\sum_{n=1}^{N} \int_{S_n} \left\{ \alpha \frac{\partial u_h}{\partial t} v_h + D\nabla u_h \cdot \nabla v_h \right\} d\mathbf{x}\, dt + \sum_{m=1}^{N-1} \int_{\Omega} \alpha [u_h]_m \, (v_h)_m^+ \, d\mathbf{x}
$$

$$
(7) \qquad + \int_{\Omega} \alpha (u_h)_0^+ (v_h)_0^+ \, d\mathbf{x} = \sum_{n=1}^{N} \left\{ \int_{S_n} f v_h \, d\mathbf{x}\, dt + \int_{\partial \mathcal{Q}_N^n} g v_h \, ds\, dt \right\}
$$

$$
+ \int_{\Omega} \alpha u_h^0 (v_h)_0^+ \, d\mathbf{x} \quad \forall v_h \in \mathcal{S}_h,
$$

with $u_h^0 \in X_{h_1}^r \cap H_{\Gamma_D}^1(\Omega)$ a suitable finite element approximation of the initial datum $u_0(\mathbf{x})$. Notice that, while u_h^0 is known, the corresponding right value $(u_h)_0^+$ is an unknown.

We complete the section by recovering the well-known *Galerkin orthogonality*, crucial tool with a view to any a posteriori error analysis. Firstly such a property can be proved slabwise due to the time discontinuity.
For this purpose we restrict the weak formulation (5) to the slab S_n, for any $n = 1, \ldots, N$: find $u \in U\big|_{S_n}$ such that
(8)
$$
\int_{S_n} \left\{ \alpha \frac{\partial u}{\partial t} v + D\nabla u \cdot \nabla v \right\} d\mathbf{x}\, dt + \int_{\Omega} \alpha [u]_{n-1} \, v_{n-1}^+ \, d\mathbf{x} = \int_{S_n} f v \, d\mathbf{x}\, dt + \int_{\mathcal{Q}_N^n} g v \, ds\, dt,
$$

for any $v \in U\big|_{S_n}$, and where $[u]_{n-1} = 0$. As confined to the single slab S_n, we can choose $v \in \mathcal{S}\big|_{S_n}$, being $\mathcal{S}\big|_{S_n} \subset U\big|_{S_n}$. This would not have been the case on the whole space-time domain $\mathcal{Q}$.
Now the temporal discontinuity characterizing the space $\mathcal{S}_h$ allows us to select the

values of v_h in (7) independently on each J_n. We pick v_h to vanish outside J_n: the cGdG form (7) thus reduces to the unique equation on J_n

$$
\text{(9)} \quad
\begin{aligned}
&\int_{S_n} \left\{ \alpha \frac{\partial u_h}{\partial t} v_h + D \nabla u_h \cdot \nabla v_h \right\} d\mathbf{x}\, dt + \int_\Omega \alpha [u_h]_{n-1} (v_h)_{n-1}^+ \, d\mathbf{x} \\
&= \int_{S_n} f v_h \, d\mathbf{x}\, dt + \int_{\mathcal{Q}_N^n} g v_h \, ds\, dt \quad \forall n = 1, \dots, N,
\end{aligned}
$$

with $[u_h]_0 = (u_h)_0^+ - u_h^0$. If we subtract (9) from (8) after identifying v with v_h, we get the desired orthogonality relation: for any $n = 1, \dots, N$,

$$
\text{(10)} \quad \int_{S_n} \left\{ \alpha \frac{\partial e_h}{\partial t} v_h + D \nabla e_h \cdot \nabla v_h \right\} d\mathbf{x}\, dt + \int_\Omega \alpha [e_h]_{n-1} (v_h)_{n-1}^+ \, d\mathbf{x} = 0,
$$

with $e_h = u - u_h \in W$ the *space-time error* associated with the cGdG approximation u_h. Identity (10) can now be generalized to an arbitrary function $v_h \in \mathcal{S}_h$ by suitably summing through the slabs S_n's.

3. Looking for anisotropic information

When modeling a physical phenomenon exhibiting strong directional features, to sharply detect these characteristics it is advisable to take into account not only the size but also the orientation and the shape of each mesh element via a suitable mesh adaptation procedure. This demand leads us to leave the standard isotropic adaptive techniques and to move towards an anisotropic grid adaptation strategy (see, for instance, [10, 18, 2, 15, 22, 47], and W. Cao [7]), even if characterized by a more complex analysis with respect to the standard isotropic one. On the other hand the attained reduction on the computational costs often justifies the sophisticated theoretical analysis involved.

We provide here the basic ideas of the anisotropic setting founded in [22]. We focus on the generic slab S_n, with $\mathcal{T}_{h_n} = \{K_n\}$ the associated conformal mesh. The source of the anisotropic information is the standard invertible affine map $T_{K_n} : \widehat{K} \to K_n$ from the reference triangle $\widehat{K}$ to the general element $K_n \in \mathcal{T}_{h_n}$, identified by the relation

$$
\mathbf{x} = (x_1, x_2)^T = T_{K_n}(\widehat{\mathbf{x}}) = M_{K_n} \widehat{\mathbf{x}} + \mathbf{t}_{K_n} \quad \forall \mathbf{x} \in K_n,
$$

with $\widehat{\mathbf{x}} \in \widehat{K}$, and where $M_{K_n} \in \mathbb{R}^{2 \times 2}$ and $\mathbf{t}_{K_n} \in \mathbb{R}^2$. In particular we exploit the spectral properties of the Jacobian M_{K_n} via two successive factorizations: we first introduce the polar decomposition $M_{K_n} = B_{K_n} Z_{K_n}$ of M_{K_n} into the symmetric positive definite matrix $B_{K_n} \in \mathbb{R}^{2 \times 2}$ and the orthogonal matrix $Z_{K_n} \in \mathbb{R}^{2 \times 2}$ corresponding to a stretching and a rotation, respectively. Then we further factorize B_{K_n} in terms of its eigenvectors $\mathbf{r}_{i,K_n}$ and eigenvalues λ_{i,K_n}, for $i = 1, 2$, as $B_{K_n} = R_{K_n}^T \Lambda_{K_n} R_{K_n}$, with $\Lambda_{K_n} = \operatorname{diag}(\lambda_{1,K_n}, \lambda_{2,K_n})$ and $R_{K_n}^T = [\mathbf{r}_{1,K_n}, \mathbf{r}_{2,K_n}]$. Notice that Z_{K_n} and $\mathbf{t}_{K_n}$ do not play any crucial role as associated with a rigid rotation and a shift, respectively.[1]

As reference element $\widehat{K}$ we adopt the equilateral triangle inscribed in the unit circle, with barycenter located at the origin. For this choice we can completely describe the shape and the orientation of each element K_n through the quantities $\mathbf{r}_{i,K_n}$ and λ_{i,K_n}. The unit circle circumscribed to $\widehat{K}$ is mapped into an ellipse circumscribing K_n: the eigenvectors $\mathbf{r}_{i,K_n}$ and the eigenvalues λ_{i,K_n} provide us with the directions and the lengths of the semi-axes of such an ellipse, respectively (see Figure 2). In

---

[1]Actually the overall decomposition is equivalent to the standard SVD decomposition: $M_K = R_{K_n}^T \Lambda_{K_n} P_{K_n}$, with $P_{K_n} = R_{K_n} Z_{K_n}$.

particular, we introduce the so-called stretching factor $s_{K_n} = \lambda_{1,K_n}/\lambda_{2,K_n}$ which quantifies the deformation of the element K_n. Without loosing generality, we assume $\lambda_{1,K_n} \geq \lambda_{2,K_n}$, so that $s_{K_n} \geq 1$, s_K being identically equal to 1 in the case of an equilateral K.
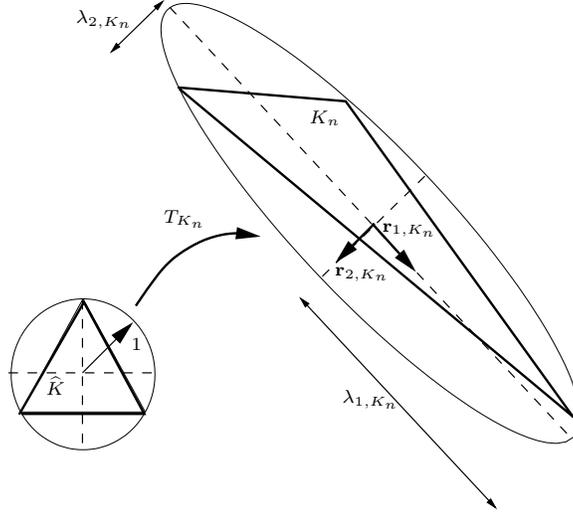

FIGURE 2. Geometrical interpretation of the map T_{K_n} and main anisotropic quantities.

With a view to the a posteriori analysis below, we introduce a suitable interpolation operator. In particular, since we often deal with functions not necessarily continuous, we refer to the Clément interpolant (Ph. Clément [13]). In the case of affine finite elements the Clément interpolant is defined as

$$(11) \qquad I_{h_n}^1 v = \sum_{N_j} P_j v(N_j) \varphi_j \quad \forall v \in L^2(\Omega),$$

where $\varphi_j \in X_{h_n}^1$ is the Lagrangian basis function associated with the node N_j, while $P_j v$ denotes the plane associated with the patch Δ_j of the elements sharing N_j, identified by the relations

$$\int_{\Delta_j} (P_j v - v)\psi \, d\mathbf{x} = 0 \quad \text{with } \psi = 1, x_1, x_2.$$

The sum in (11) runs on all the mesh vertices except the N_j's belonging to the Dirichlet boundary.

The actual goal is to generalize to the time dependent framework some of the anisotropic error estimates proved for this interpolant in [22, 23]. In more detail, in the spirit of a standard mesh adaptation procedure, we are interested in local interpolation error estimates, i.e., associated with the space-time prism S_{K_n}.

Lemma 3.1. *Let Δ_{K_n} be the patch of the elements sharing at least a vertex with the triangle $K_n \in \mathcal{T}_{h_n}$, and let v be a function in $H^1(\Omega)$. Then if*

$$(12) \qquad card(\Delta_{K_n}) \leq \mathcal{N} \quad and \quad diam(\Delta_{\widehat{K}}) \leq C_\Delta \simeq O(1),$$

with $C_\Delta \geq h_{\widehat{K}}$ constant and $\Delta_{\widehat{K}} = T_{K_n}^{-1}(\Delta_{K_n})$ the reference patch, there exist two constants $C_i = C_i(\mathcal{N}, C_\Delta)$, with $i = 1, 2$, such that, $\forall K_n \in \mathcal{T}_{h_n}$,

$$(13) \qquad \|v - I_{h_n}^1 v\|_{L^2(S_{K_n})} \leq C_1 \left[\sum_{i=1}^2 \lambda_{i,K_n}^2 \left(\mathbf{r}_{i,K_n}^T G_{K_n}^n(v) \mathbf{r}_{i,K_n} \right) \right]^{1/2},$$

(14)

$$\|v - I_{h_n}^1 v\|_{L^2(L_{K_n})} \leq C_2 \left(\frac{h_{K_n}}{\lambda_{1,K_n} \lambda_{2,K_n}} \right)^{1/2} \left[\sum_{i=1}^2 \lambda_{i,K_n}^2 \left(\mathbf{r}_{i,K_n}^T G_{K_n}^n(v) \mathbf{r}_{i,K_n} \right) \right]^{1/2},$$

where $G_{K_n}^n(v) \in \mathbb{R}^{2\times 2}$ is the symmetric positive semi-definite matrix defined by

$$(15) \qquad \left(G_{K_n}^n(v) \right)_{ij} = \int_{\Delta_{K_n} \times J_n} \frac{\partial v}{\partial x_i} \frac{\partial v}{\partial x_j} \, d\mathbf{x} \, dt \quad with \quad i, j = 1, 2,$$

S_{K_n} and L_{K_n} being as in Figure 1.

Proof. Results (13) and (14) follow by integrating on J_n relation (30) in [22] and (2.15) in [23], respectively and thanks to the time independence of the $\mathbf{r}_{i,K_n}$'s and of the λ_{i,K_n}'s. $\qquad\square$

Remark 3.2. *The reference patch $\Delta_{\widehat{K}}$ is obtained by extending the inverse map $T_{K_n}^{-1} : K_n \to \widehat{K}$ to all of the elements of Δ_{K_n}. Conditions (12) avoid excessively distorted patches in the reference framework. Essentially they do not limit the anisotropic features (stretching factor and orientation) of each $T \in \Delta_{K_n}$, but rather the variation over Δ_{K_n} of the anisotropic quantities. We refer to S. Micheletti, S. Perotto & M. Picasso [37] for more details.*

We highlight the explicit dependence of the estimates (13)-(14) on the anisotropic quantities λ_{i,K_n} and $\mathbf{r}_{i,K_n}$ in contrast with standard isotropic interpolation estimates, where only the diameter h_{K_n} of K_n plays a role. In more detail, via matrix $G_{K_n}^n(v)$ the information provided by the first-order partial derivatives is projected along the directions $\mathbf{r}_{1,K_n}$ and $\mathbf{r}_{2,K_n}$ rather than lumped into the H^1-seminorm $|v|_{H^1(\Delta_{K_n} \times J_n)}$ as in the isotropic case. As particular case when $\lambda_{1,K_n} \simeq \lambda_{2,K_n} \simeq h_{K_n}$, i.e., when the triangle K_n is equilateral, we recover the corresponding isotropic results (see, for instance, [12]).

Estimates (13)-(14) hold also for a Clément-like interpolant (L.R. Scott & S. Zhang [46]), provided that the definition of the patch Δ_{K_n} is accordingly modified. On the other hand we refer to [22, 23] for the corresponding anisotropic interpolation error estimates associated with the standard Lagrange interpolant.
Finally we stress that, in contrast to other well-established anisotropic frameworks (see, e.g., [2]), the approach above is free from any maximal angle condition. The sharpness of the estimates is not compromised even though one of the angles grows to π (see Section 2.1 in [22] and [7] for further details).

4. The "driving force" for the space-time adaptation procedure

This section represents the hardcore of the whole paper. Here we build the theoretical tool at the basis of the space-time adaptation procedure for problem (1). The final outcome is a residual-based type a posteriori error estimator for the norm $|||e_h|||_{\mathrm{DG}}$, enriched with anisotropic information. An anisotropic control

for the DG-norm of the error e_h represents the actual novelty of the a posteriori analysis below in the current literature pertaining to parabolic problems.

We need to anticipate some useful notation. We define the *local residuals* distinguishing between the spatial and the temporal ones. For a fixed time interval J_n, with $n = 1, \ldots, N$, and for any $K_n \in \mathcal{T}_{h_n}$, let

$$(16) \qquad \rho_{K_n} = \left[f - \alpha \frac{\partial u_h}{\partial t} + \nabla \cdot (D\nabla u_h) \right]\Big|_{S_{K_n}}$$

and

$$(17) \qquad j_{K_n} = \begin{cases} 0 & \text{on } (\partial K_n \cap \Gamma_D) \times J_n, \\ 2(g - D\nabla u_h \cdot \mathbf{n})\big|_{S_{K_n}} & \text{on } (\partial K_n \cap \Gamma_N) \times J_n, \\ -[D\nabla u_h \cdot \mathbf{n}] & \text{on } (\partial K_n \cap \mathcal{E}_h^n) \times J_n, \end{cases}$$

be the interior and the boundary residual associated with the cGdG-approximation u_h and with the prism S_{K_n}, respectively, where $\mathcal{E}_h^n$ is the skeleton of $\mathcal{T}_{h_n}$ while $[D\nabla u_h \cdot \mathbf{n}] = D\nabla u_h \cdot \mathbf{n}_{K_n} + D\nabla u_h \cdot \mathbf{n}_{K_n'}$ denotes the jump of the diffusive flux across the internal interfaces of K_n, for any $K_n' \in \mathcal{T}_{h_n}$ such that $K_n' \cap K_n \in \mathcal{E}_h^n$.

We introduce now the temporal residual

$$(18) \qquad \mathcal{J}_n = \alpha[-u_h]_n = \alpha\big(- (u_h)_n^+ + (u_h)_n^- \big)$$

associated with u_h and with the time level t_n, together with the initial error

$$(19) \qquad e_0^- = \alpha\big(u_0 - u_h^0 \big).$$

The quantities ρ_{K_n} and j_{K_n} in (16)-(17) are related to the space discretization, whereas $\mathcal{J}_n$ and e_0^- depend on the discontinuous temporal scheme.

An important role in the a posteriori analysis below is also played by the time projection operator $T_n : \mathcal{S}\big|_{S_n} \to H^1_{\Gamma_D}(\Omega)$, for $n = 1, \ldots, N$, such that

$$(20) \qquad T_n w = k_n^{-1} \int_{J_n} w(\mathbf{x}, t)\, dt \quad \forall w \in \mathcal{S}$$

satisfying the estimate (see, e.g., [19])

$$(21) \qquad \|w - T_n w\|_{L^2(J_n)} \leq k_n \left\| \frac{\partial w}{\partial t} \right\|_{L^2(J_n)}.$$

Notice that no constant appears in such an estimate. Moreover, by definition of T_n, the projection error $(w - T_n w)$ is orthogonal to any constant function, i.e.,

$$(22) \qquad \int_{J_n} (w - T_n w)\, c\, dt = 0 \quad \forall w \in \mathcal{S}\big|_{S_n} \text{ and } \forall\, c = \text{constant in time.}$$

We finally define the time-averaged residuals

$$(23) \qquad \overline{\rho}_{K_n} = T_n \rho_{K_n} \quad \text{and} \quad \overline{j}_{K_n} = T_n j_{K_n}.$$

Proposition 4.1. *Let $u \in U$ be the solution to the weak problem (2) and let $u_h \in \mathcal{S}_h$ be the corresponding cGdG-approximation solving (7) for the choices $r = 1$ and $q = 0$. Let the approximation u_h^0 of the initial datum u_0 coincide with the corresponding α-weighted L^2-projection $P_{h_1}^1 u_0$ onto the space $X_{h_1}^1 \cap H^1_{\Gamma_D}(\Omega)$, such that*

$$(24) \qquad \int_\Omega \alpha\big(P_{h_1}^1 u_0 - u_0 \big) v\, d\mathbf{x} = 0 \quad \forall v \in X_{h_1}^1 \cap H^1_{\Gamma_D}(\Omega).$$

Then, under the hypotheses of Lemma 3.1, there exists a constant $C = C(\mathcal{N}, C_\Delta)$ such that

$$
(25) \quad |||e_h|||_{\mathrm{DG}}^2 \leq C \sum_{n=1}^{N} \sum_{K_n \in \mathcal{T}_{h_n}} \left(\alpha_{K_n}^S R_{K_n}^S \omega_{K_n}^S + \alpha_{K_n}^{T1} R_{K_n}^{T1} \omega_{K_n}^{T1} + \alpha_{K_n}^{T2} R_{K_n}^{T2} \omega_{K_n}^{T2} \right),
$$

where $\alpha_{K_n}^S = |\widehat{K}| \lambda_{1,K_n}^{3/2} \lambda_{2,K_n}^{3/2}$, $\alpha_{K_n}^{T1} = \alpha_{K_n}^{T2} = k_n^2$,

$$
R_{K_n}^S = \frac{1}{|K_n|^{1/2}} \left\{ \frac{1}{2} \|\overline{j}_{K_n}\|_{L^2(L_{K_n})} \left(\frac{h_{K_n}}{\lambda_{1,K_n} \lambda_{2,K_n}} \right)^{1/2} \right.
$$
$$
\left. + \left[\|\overline{\rho}_{K_n}\|_{L^2(S_{K_n})} + \frac{1}{k_n^{1/2}} \left(\|\mathcal{J}_{n-1}\|_{L^2(K_n)} + \delta_{1n} \|e_0^-\|_{L^2(K_n)} \right) \right] \right\},
$$

$$
R_{K_n}^{T1} = \frac{1}{k_n^{1/2}} \left[\|\rho_{K_n} - \overline{\rho}_{K_n}\|_{L^2(S_{K_n})} + \frac{1}{k_n^{1/2}} \left(\|\mathcal{J}_{n-1}\|_{L^2(K_n)} + \delta_{1n} \|e_0^-\|_{L^2(K_n)} \right) \right],
$$

$$
R_{K_n}^{T2} = \frac{1}{2 k_n^{1/2}} \|j_{K_n} - \overline{j}_{K_n}\|_{L^2(L_{K_n})},
$$

$$
\omega_{K_n}^{T1} = \frac{1}{k_n^{1/2}} \left\| \frac{\partial e_h}{\partial t} \right\|_{L^2(S_{K_n})}, \quad \omega_{K_n}^{T2} = \frac{1}{k_n^{1/2}} \left\| \frac{\partial e_h}{\partial t} \right\|_{L^2(L_{K_n})},
$$

$$
\omega_{K_n}^S = \left[s_{K_n} \left(\mathbf{r}_{1,K_n}^T \widetilde{G}_{K_n}^n(e_h) \mathbf{r}_{1,K_n} \right) + \frac{1}{s_{K_n}} \left(\mathbf{r}_{2,K_n}^T \widetilde{G}_{K_n}^n(e_h) \mathbf{r}_{2,K_n} \right) \right]^{1/2},
$$

δ_{1n} *is the Kronecker symbol identifying the first slab* S_1, $\widetilde{G}_{K_n}^n(\cdot) = |K_n|^{-1} G_{K_n}^n(\cdot)$ *is the scaled counterpart of the matrix in (15), notation (16)-(23) being additionally adopted.*

Proof. We move from the definition of the bilinear form $B_{\mathrm{DG}}(\cdot, \cdot)$ applied to the discretization error $e_h \in W$ and to a generic function $v \in W$. By adding and subtracting the values $\int_\Omega \alpha(u_0^- - u_h^0) v_0^+ \, d\mathbf{x}$ and $\int_\Omega \alpha(u_0^- - u_h^0)(v_h)_0^+ \, d\mathbf{x}$, with $u_h^0 = P_{h_1}^1 u_0$, and exploiting the orthogonality relation (10), we get

$$
(26) \quad
\begin{aligned}
B_{\mathrm{DG}}(e_h, v) = \sum_{n=1}^{N} \Bigg[& \int_{S_n} \left\{ \alpha \frac{\partial e_h}{\partial t}(v - v_h) + D \nabla e_h \cdot \nabla(v - v_h) \right\} d\mathbf{x} dt \\
& + \int_\Omega \alpha [e_h]_{n-1} (v - v_h)_{n-1}^+ \, d\mathbf{x} \Bigg] + \int_\Omega \alpha(u_0^- - u_h^0)(v_0^+ - (v_h)_0^+) \, d\mathbf{x},
\end{aligned}
$$

relation (24) being used with $v = (v_h)_0^+ \in X_{h_1}^1 \cap H_{\Gamma_D}^1(\Omega)$. With a view to a local a posteriori estimator, we split the integrals in (26) on the whole Ω onto each triangle of the mesh. Taking advantage of relation (8) and after integrating by parts the diffusive term associated with u_h, we have

$$
\begin{aligned}
B_{\mathrm{DG}}(e_h, v) = \sum_{n=1}^{N} \sum_{K_n \in \mathcal{T}_{h_n}} \Bigg[& \int_{J_n} \left\{ \int_{K_n} \left(f - \alpha \frac{\partial u_h}{\partial t} + \nabla \cdot (D \nabla u_h) \right)(v - v_h) \, d\mathbf{x} \right. \\
& \left. + \int_{\partial K_n \cap \Gamma_N} g(v - v_h) \, ds - \int_{\partial K_n} (D \nabla u_h \cdot \mathbf{n})(v - v_h) \, ds \right\} dt \\
& + \int_{K_n} \alpha [-u_h]_{n-1} (v - v_h)_{n-1}^+ \, d\mathbf{x} \Bigg] + \sum_{K_1 \in \mathcal{T}_{h_1}} \int_{K_1} \alpha(u_0^- - u_h^0)(v_0^+ - (v_h)_0^+) \, d\mathbf{x},
\end{aligned}
$$

i.e., thanks to the definitions (16)-(19),

$$B_{DG}(e_h, v) = \sum_{n=1}^{N} \sum_{K_n \in \mathcal{T}_{h_n}} \left\{ \underbrace{\int_{S_{K_n}} \rho_{K_n}(v - v_h)\, d\mathbf{x}dt}_{(I)} + \underbrace{\frac{1}{2} \int_{L_{K_n}} j_{K_n}(v - v_h)\, dsdt}_{(II)} \right.$$

$$\left. + \underbrace{\int_{K_n} \mathcal{J}_{n-1}(v - v_h)_{n-1}^{+}\, d\mathbf{x}}_{(III)} \right\} + \sum_{K_1 \in \mathcal{T}_{h_1}} \underbrace{\int_{K_1} e_0^{-}\left(v_0^{+} - (v_h)_0^{+}\right) d\mathbf{x}}_{(IV)}.$$

The factor $1/2$ takes into account that each internal edge of $\mathcal{T}_{h_n}$ is shared by two elements. We analyze separately the four terms (I)-(IV) after identifying

$$v_h \equiv T_n(I_{h_n}^1 v) = \frac{1}{k_n} \int_{J_n} I_{h_n}^1 v(\mathbf{x}, t)\, dt \in X_{h_n}^1 \cap H_{\Gamma_D}^1(\Omega),$$

i.e., $v_h \in \mathcal{S}|_{S_n}$ for $q = 0$, extended to zero outside J_n when considered as a function in $\mathcal{S}_h$ for $q = 0$ and $r = 1$. To keep distinct the spatial contribution from the temporal one in the error estimator we resort to the splitting

(27) $$v - T_n(I_{h_n}^1 v) = \left(v - T_n(v)\right) + \left(T_n(v) - T_n(I_{h_n}^1 v)\right).$$

The idea is to exploit the projection estimate (21) on the first term (the temporal one) in (27), while applying the interpolation estimates in Lemma 3.1 on the second one (the spatial contribution) to introduce the desired anisotropic information. We consider the first term. Thanks to the definitions of the averaged residual $\overline{\rho}_{K_n}$ and of the projection operator T_n, and by suitably combining the orthogonality relation (22) and the projection estimate (21) together with the Cauchy-Schwarz inequality, we write (I) as

$$|(I)| = \left| \int_{S_{K_n}} \rho_{K_n}\left(v - T_n(v)\right) d\mathbf{x}dt + \int_{S_{K_n}} \rho_{K_n}\left(T_n(v) - T_n(I_{h_n}^1 v)\right) d\mathbf{x}dt \right|$$

$$= \left| \int_{S_{K_n}} \left(\rho_{K_n} - \overline{\rho}_{K_n}\right)\left(v - T_n(v)\right) d\mathbf{x}dt + \int_{K_n} \left[T_n\left(v - I_{h_n}^1 v\right) \int_{J_n} \rho_{K_n}\, dt \right] d\mathbf{x} \right|$$

$$\leq \int_{K_n} \|\rho_{K_n} - \overline{\rho}_{K_n}\|_{L^2(J_n)} \|v - T_n(v)\|_{L^2(J_n)}\, d\mathbf{x} + \left| \int_{S_{K_n}} \overline{\rho}_{K_n}\left(v - I_{h_n}^1 v\right) d\mathbf{x}dt \right|$$

$$\leq k_n \left\| \frac{\partial v}{\partial t} \right\|_{L^2(S_{K_n})} \|\rho_{K_n} - \overline{\rho}_{K_n}\|_{L^2(S_{K_n})} + \|\overline{\rho}_{K_n}\|_{L^2(S_{K_n})} \|v - I_{h_n}^1 v\|_{L^2(S_{K_n})}.$$

Via the anisotropic interpolation estimate (13) we consequently have

(28)
$$|(I)| \leq k_n \left\| \frac{\partial v}{\partial t} \right\|_{L^2(S_{K_n})} \|\rho_{K_n} - \overline{\rho}_{K_n}\|_{L^2(S_{K_n})}$$

$$+ C_1 \|\overline{\rho}_{K_n}\|_{L^2(S_{K_n})} \left[\sum_{i=1}^{2} \lambda_{i,K_n}^2 \left(\mathbf{r}_{i,K_n}^T G_{K_n}^n(v) \mathbf{r}_{i,K_n} \right) \right]^{1/2}.$$

The second term can be bounded mimicking exactly the computations performed on (I), now confined to the lateral surface L_{K_n} of S_{K_n}. This yields

$$|(II)| \leq \frac{k_n}{2} \left\| \frac{\partial v}{\partial t} \right\|_{L^2(L_{K_n})} \|j_{K_n} - \overline{j}_{K_n}\|_{L^2(L_{K_n})} + \frac{1}{2} \|\overline{j}_{K_n}\|_{L^2(L_{K_n})} \|v - I_{h_n}^1 v\|_{L^2(L_{K_n})}$$

i.e., after introducing the anisotropic information via (14),

$$
\begin{aligned}
|\,(\text{II})\,| \quad &\leq \frac{k_n}{2}\left\|\frac{\partial v}{\partial t}\right\|_{L^2(L_{K_n})}\|j_{K_n}-\bar{j}_{K_n}\|_{L^2(L_{K_n})}\\
&+\frac{C_2}{2}\|\bar{j}_{K_n}\|_{L^2(L_{K_n})}\left(\frac{h_{K_n}}{\lambda_{1,K_n}\lambda_{2,K_n}}\right)^{1/2}\Big[\sum_{i=1}^{2}\lambda_{i,K_n}^2\left(\mathbf{r}_{i,K_n}^T G_{K_n}^n(v)\mathbf{r}_{i,K_n}\right)\Big]^{1/2}.
\end{aligned}
$$

Concerning the quantity (III), first we still apply the splitting (27) to get

$$
(\text{III}) = \int_{K_n}\mathcal{J}_{n-1}\left(v-T_n(v)\right)_{n-1}^{+}d\mathbf{x} + \int_{K_n}\mathcal{J}_{n-1}\left(T_n(v)-T_n(I_{h_n}^1 v)\right)_{n-1}^{+}d\mathbf{x}.
$$

To control the first of these two terms, we exploit the mean value theorem (in time) yielding

$$
\left(v-T_n(v)\right)_{n-1}^{+} = v_{n-1}^{+}-v_{t_n^*} = -\int_{t_{n-1}}^{t_n^*}\frac{\partial v}{\partial t}(s)\,ds,
$$

for a suitable $t_n^* \in (t_{n-1}, t_n)$. Thanks to this equality and using the definition of T_n, the relation $\|\mathcal{J}_{n-1}\|_{L^2(S_{K_n})} = k_n^{1/2}\|\mathcal{J}_{n-1}\|_{L^2(K_n)}$ and the Cauchy-Schwarz inequality, we have

$$
\begin{aligned}
|\,(\text{III})\,| &\leq \int_{S_{K_n}}|\mathcal{J}_{n-1}|\left|\frac{\partial v}{\partial t}\right|d\mathbf{x}dt + \frac{1}{k_n}\int_{S_{K_n}}|\mathcal{J}_{n-1}|\left|v-I_{h_n}^1 v\right|d\mathbf{x}dt\\
&\leq \|\mathcal{J}_{n-1}\|_{L^2(S_{K_n})}\left\|\frac{\partial v}{\partial t}\right\|_{L^2(S_{K_n})} + \frac{1}{k_n}\|\mathcal{J}_{n-1}\|_{L^2(S_{K_n})}\|v-I_{h_n}^1 v\|_{L^2(S_{K_n})}\\
&\leq k_n^{1/2}\|\mathcal{J}_{n-1}\|_{L^2(K_n)}\left\|\frac{\partial v}{\partial t}\right\|_{L^2(S_{K_n})} + \frac{1}{k_n^{1/2}}\|\mathcal{J}_{n-1}\|_{L^2(K_n)}\|v-I_{h_n}^1 v\|_{L^2(S_{K_n})}.
\end{aligned}
$$

The anisotropic estimate (13) leads to

$$
\begin{aligned}
|\,(\text{III})\,| \quad &\leq \|\mathcal{J}_{n-1}\|_{L^2(K_n)}\Big\{k_n^{1/2}\left\|\frac{\partial v}{\partial t}\right\|_{L^2(S_{K_n})}\\
&+\frac{1}{k_n^{1/2}}C_1\Big[\sum_{i=1}^{2}\lambda_{i,K_n}^2\left(\mathbf{r}_{i,K_n}^T G_{K_n}^n(v)\mathbf{r}_{i,K_n}\right)\Big]^{1/2}\Big\}.
\end{aligned}
$$

Finally, (IV) can be bounded by repeating the same steps employed for (III), provided that the initial error e_0^- replaces the temporal residual $\mathcal{J}_{n-1}$. We obtain

$$
\begin{aligned}
(29)\qquad |\,(\text{IV})\,| \quad &\leq \|e_0^-\|_{L^2(K_1)}\Big\{k_1^{1/2}\left\|\frac{\partial v}{\partial t}\right\|_{L^2(S_{K_1})}\\
&+\frac{1}{k_1^{1/2}}C_1\Big[\sum_{i=1}^{2}\lambda_{i,K_1}^2\left(\mathbf{r}_{i,K_1}^T G_{K_1}^n(v)\mathbf{r}_{i,K_1}\right)\Big]^{1/2}\Big\}.
\end{aligned}
$$

By summing (28)-(29), we deduce
(30)

$$
\begin{aligned}
|B_{\mathrm{DG}}(e_h, v)| \leq \sum_{n=1}^{N} \sum_{K_n \in \mathcal{T}_{h_n}} & \left\{ \left[k_n \left\| \frac{\partial v}{\partial t} \right\|_{L^2(S_{K_n})} \| \rho_{K_n} - \overline{\rho}_{K_n} \|_{L^2(S_{K_n})} \right. \right. \\
& + \frac{k_n}{2} \left\| \frac{\partial v}{\partial t} \right\|_{L^2(L_{K_n})} \| j_{K_n} - \overline{j}_{K_n} \|_{L^2(L_{K_n})} + k_n^{1/2} \left\| \frac{\partial v}{\partial t} \right\|_{L^2(S_{K_n})} \\
& \left(\| \mathcal{J}_{n-1} \|_{L^2(K_n)} + \| e_0^- \|_{L^2(K_1)} \delta_{1n} \right) \right] + \left[\sum_{i=1}^{2} \lambda_{i,K_n}^2 \left(\mathbf{r}_{i,K_n}^T G_{K_n}^n(v) \mathbf{r}_{i,K_n} \right) \right]^{1/2} \\
& \left[C_1 \| \overline{\rho}_{K_n} \|_{L^2(S_{K_n})} + \frac{C_2}{2} \| \overline{j}_{K_n} \|_{L^2(L_{K_n})} \left(\frac{h_{K_n}}{\lambda_{1,K_n} \lambda_{2,K_n}} \right)^{1/2} \right. \\
& \left. \left. + \frac{1}{k_n^{1/2}} C_1 \left(\| \mathcal{J}_{n-1} \|_{L^2(K_n)} + \| e_0^- \|_{L^2(K_1)} \delta_{1n} \right) \right] \right\}.
\end{aligned}
$$

Result (25) can now be derived by suitably scaling all the quantities on the right-hand side of (30): the first three terms are scaled with respect to k_n; the others with respect to $|K_n|$. Eventually the choice $v = e_h$ is made. $\square$

Some remarks concerning result (25) are in order. First we point out the specific choice made for the discretization scheme, i.e., for the cGdG space. We use affine finite elements for the spatial discretization due to the anisotropic framework we are interested in, so far confined to the degree one case. On the other hand the choice of functions piecewise constant in time is essentially justified by the interest for temporal multiscale physical phenomena whose solutions are in general nonsmooth. As an example in this direction we refer to the test case in Section 5.3. Moreover the time discontinuity makes the adopted numerical scheme similar to a time-marching procedure, with the related computational advantages. In particular the dG(0) scheme in time is equivalent to a modified backward Euler method, wherein the time averaged source term replaces the contribution of the source term at the new time level [49, 4].

We expect a generalization of result (25) to the cG(r)dG(q) scheme to be feasible. For instance one can resort to the anisotropic framework in W. Cao [8] or [24], to deal with polynomials of arbitrary degree r in space. On the other hand a higher order in time can be managed by suitably re-defining the projection operator T_n in (20) (see, e.g., R. Becker & R. Rannacher [5]).

We also highlight the presence of the time averaged residuals $\overline{\rho}_{K_n}$ and $\overline{j}_{K_n}$ on the right-hand side of (25) instead of ρ_{K_n} and j_{K_n}, respectively. The introduction of these mean values provides us with a more reliable control of the norm $|||e_h|||_{\mathrm{DG}}$, since $\| \overline{v} \|_{L^2(J_n)} \leq \| v \|_{L^2(J_n)}$ as well as $\| v - \overline{v} \|_{L^2(J_n)} \leq \| v \|_{L^2(J_n)}$, for any $v \in L^2(J_n)$. Finally we remark that the residuals $R_{K_n}^S$ and $R_{K_n}^{Ti}$ as well as the weights $\omega_{K_n}^S$ and $\omega_{K_n}^{Ti}$, with $i = 1, 2$, are scaled with respect to the space and time, respectively. All the spatial and temporal dimensional information are thus collected into the coefficients $\alpha_{K_n}^S$ and $\alpha_{K_n}^{Ti}$, respectively while all the anisotropic information are lumped into the $\omega_{K_n}^S$'s. The rationale behind this scaling and the splitting of the spatial and temporal contributions will be clear with a view to the adaptive procedure in Section 4.2.

4.1. The error estimator. The right-hand side of (25) is not yet actually useful to drive an adaptive procedure since it explicitly depends on the (unknown) space-time discretization error e_h. On the other hand the presence of such an error is expected to guarantee the convergence to zero of the upper bound in (25) as the mesh gets finer and finer.

To preserve this nice property while making estimate (25) handy, we need to replace the weights $\omega_{K_n}^S$ and $\omega_{K_n}^{Ti}$ with suitable computable quantities. We have to distinguish between the spatial and temporal weights. Concerning the spatial ones, it turns out to be convenient resorting to the well-known Zienkiewicz-Zhu recovery procedure ([52, 53, 54]), due to the dependence of the $\omega_{K_n}^S$'s on the first-order derivatives of the error. Denoting the recovered gradient of u_h with $\nabla^{ZZ} u_h = \left((\nabla^{ZZ} u_h)_1, (\nabla^{ZZ} u_h)_2 \right)^T$, we substitute the matrix $\widetilde{G}_{K_n}^n(e_h)$ in the definition of $\omega_{K_n}^S$ with the computable one $\widetilde{G}_{K_n}^n(e_h^{*,S})$, given by

$$(31) \qquad \left(\widetilde{G}_{K_n}^n(e_h^{*,S}) \right)_{i,j} = |K_n|^{-1} \int_{\Delta_{K_n} \times J_n} \left((\nabla^{ZZ} u_h)_i - \frac{\partial u_h}{\partial x_i} \right) \left((\nabla^{ZZ} u_h)_j - \frac{\partial u_h}{\partial x_j} \right) d\mathbf{x} \, dt,$$

for $i, j = 1, 2$. We have already employed successfully such an approach also in an anisotropic framework (see, e.g., [32, 34, 17]). In particular, among all the recipes available in the literature for the recovered gradient $\nabla^{ZZ} u_h$, we resort to the one proposed by R. Rodriguez in [43].

As far as the temporal weights are concerned, we use a suitable linear reconstruction in time, following what done in [33]. In more detail, the time derivative $\partial e_h / \partial t$ in $\omega_{K_n}^{T1}$ and $\omega_{K_n}^{T2}$ is evaluated via the quantity

$$(32) \qquad \frac{\partial e_h^{*,T}}{\partial t} = \frac{\partial u_h^{*,T}}{\partial t} - \frac{\partial u_h}{\partial t},$$

where $u_h^{*,T}$ is the piecewise linear function interpolating the left values $(u_h)_n^-$, for any t_n with $n = 0, \ldots, N$ (see Figure 3).
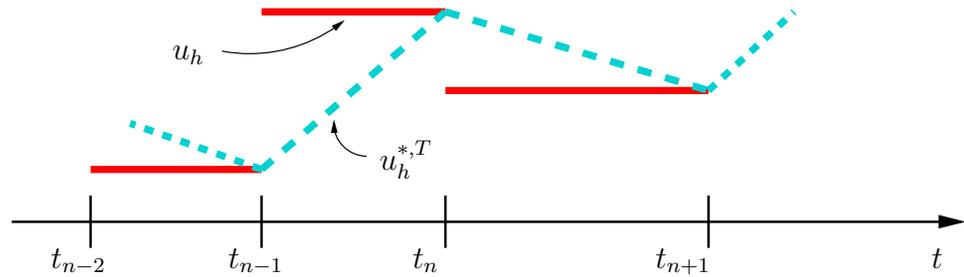

FIGURE 3. Temporal recovery procedure: u_h (solid line), $u_h^{*,T}$ (dotted line).

Finally a particular care has to be taken in computing the residual $\mathcal{J}_n$ in (18) since it merges information coming from the different meshes $\mathcal{T}_{h_n}$ and $\mathcal{T}_{h_{n+1}}$. In practice we evaluate $\mathcal{J}_n$ via

$$(33) \qquad \mathcal{J}_n^{*,S} = \alpha \left(-(u_h)_n^+ + \Pi_{n \to n+1}(u_h)_n^- \right),$$

where the operator $\Pi_{n \to n+1}$ interpolates from the d.o.f.'s of $\mathcal{T}_{h_n}$ onto the d.o.f.'s of $\mathcal{T}_{h_{n+1}}$. The value $\mathcal{J}_n^{*,S}$ is related to the so-called coarsening error (see, e.g., [20], R. Nochetto, A. Schmidt & C. Verdi [38], [11]).

Both the recovery procedures (31) and (32) and the approximation (33) would deserve a deeper analysis which is beyond the purpose of this paper.

We are now in a position to define the a posteriori error estimator associated with result (25). It is represented by the right-hand side of such an estimate provided that the replacements $\mathcal{J}_n \leftarrow \mathcal{J}_n^{*,S}$,

$$\omega_{K_n}^{T1} \leftarrow \omega_{K_n}^{*,T1} = \frac{1}{k_n^{1/2}} \left\| \frac{\partial e_h^{*,T}}{\partial t} \right\|_{L^2(S_{K_n})}, \quad \omega_{K_n}^{T2} \leftarrow \omega_{K_n}^{*,T2} = \frac{1}{k_n^{1/2}} \left\| \frac{\partial e_h^{*,T}}{\partial t} \right\|_{L^2(L_{K_n})},$$

$$\omega_{K_n}^{S} \leftarrow \omega_{K_n}^{*,S} = \left[s_{K_n} \left(\mathbf{r}_{1,K_n}^T \widetilde{G}_{K_n}^n (e_h^{*,S}) \mathbf{r}_{1,K_n} \right) + \frac{1}{s_{K_n}} \left(\mathbf{r}_{2,K_n}^T \widetilde{G}_{K_n}^n (e_h^{*,S}) \mathbf{r}_{2,K_n} \right) \right]^{1/2}$$

are carried out.

Definition 4.2. *Let $u_h \in \mathcal{S}_h$ be the cG(1)dG(0)-approximation of the solution $u \in U$ to the weak problem (2). Then, under the hypotheses of Proposition 4.1, we can assume as a posteriori error estimator for the norm $|||e_h|||_{\mathrm{DG}}$ the value*

$$(34) \qquad \eta = \Big[\sum_{n=1}^{N} \sum_{K_n \in \mathcal{T}_{h_n}} \left(\eta_{K_n}^S + \eta_{K_n}^T \right) \Big]^{1/2}$$

where, in compliance with the notation above, the local space and time error indicators are defined by

$$(35) \qquad \eta_{K_n}^S = \alpha_{K_n}^S R_{K_n}^S \omega_{K_n}^{*,S}, \quad \eta_{K_n}^T = \alpha_{K_n}^{T1} R_{K_n}^{T1} \omega_{K_n}^{*,T1} + \alpha_{K_n}^{T2} R_{K_n}^{T2} \omega_{K_n}^{*,T2},$$

respectively. It is understood that in both $R_{K_n}^S$ and $R_{K_n}^{T1}$ the temporal residual $\mathcal{J}_n$ is replaced by $\mathcal{J}_n^{*,S}$.

We now describe how to exploit the estimator η in (34) to drive an actual space-time adaptive procedure.

4.2. The adaptive algorithm. We look for an effective iterative procedure based on both an adaptive choice of the temporal step and an anisotropic mesh adaptation strategy, suitably combined to contain the computational costs involved in a cG(1)dG(0) approximation of the model problem (1).
The goal is to determine the time steps and the meshes to guarantee a uniform slab-wise distribution of the error e_h. In more detail we fix a global tolerance τ to be ensured on each space-time slab S_n. We split τ into a space (τ^S) and a time (τ^T) contribution. Then the time step and the spatial mesh are successively adapted till both the time and space estimators $\eta_n^T = \sum_{K_n \in \mathcal{T}_{h_n}} \eta_{K_n}^T$ and $\eta_n^S = \sum_{K_n \in \mathcal{T}_{h_n}} \eta_{K_n}^S$ are to within the corresponding tolerances. The same estimators, suitably exploited in a predictive fashion, drive the choice of the temporal and spatial partitions.

Thus the overall procedure can be summarized as follows (see Figure 4 for a sketch). Let us focus on the slab S_n. Firstly a loop of time adaptivity is executed: moving from an initial guess for both the time step (k_n^A) and the mesh $(\mathcal{T}_{h_n}^A)$, the cGdG approximation u_h, solution to (7), is computed along with the corresponding temporal error indicator η_n^T. Then if this estimator is below the desired threshold τ^T, we move to the second step of the whole adaptive procedure; otherwise the time step is iteratively updated until relation $\eta_n^T < \tau^T$ is met. The criterion followed to update k_n^A takes advantage of the definition of the time error estimator. We rewrite it as

$$\eta_n^T = k_n^2 \sum_{K_n \in \mathcal{T}_{h_n}^A} \left(R_{K_n}^{T1} \omega_{K_n}^{*,T1} + R_{K_n}^{T2} \omega_{K_n}^{*,T2} \right).$$
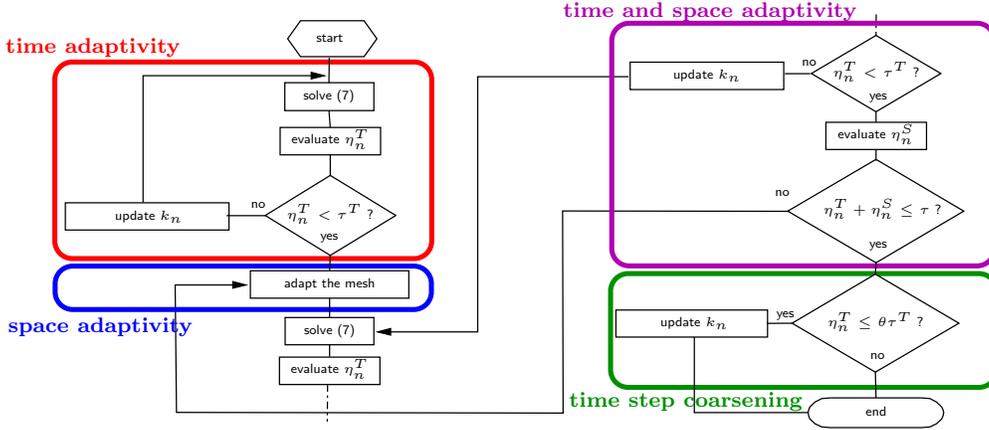
FIGURE 4. Flowchart of the space-time adaptive algorithm.

Now as both the residuals and the weights are scaled with respect to time (i.e., are independent of the time step), we can compute the new value for k_n simply by imposing the equality $\eta_n^T = \tau^T$, i.e., as

$$(36) \qquad k_n = \left[\frac{\tau^T}{\displaystyle\sum_{K_n \in \mathcal{T}_{h_n}^A} \left(R_{K_n}^{T1} \omega_{K_n}^{*,T1} + R_{K_n}^{T2} \omega_{K_n}^{*,T2} \right)} \right]^{1/2}.$$

This defines k_n through a nonlinear relation (as $R_{K_n}^{Ti}$ and $\omega_{K_n}^{Ti}$ depend on k_n). Thus in the spirit of a predictive procedure (a fixed-point map), we evaluate $R_{K_n}^{Ti}$ and $\omega_{K_n}^{Ti}$ at the previous iteration, predict k_n, and solve again (7), until $\eta_n^T < \tau^T$.

The second step of the space-time adaptive algorithm deals with the spatial anisotropic mesh adaptivity. The goal is to identify the mesh with the least number of elements, which satisfies a given constraint (τ^S) on the accuracy of the numerical solution and an error equidistribution criterion on the mesh elements. We employ a metric-based iterative algorithm driven by the estimator η_n^S, still used in a predictive way. In more detail, at each step j of this iterative procedure, we evaluate the estimator on the actual mesh $\mathcal{T}_{h_n}^{(j)}$, we devise a metric $\widetilde{\mathcal{M}}^{(j+1)}$ and then we build the new mesh $\mathcal{T}_{h_n}^{(j+1)}$, matching the metric $\widetilde{\mathcal{M}}^{(j+1)}$ (see [17]). Notice that $\mathcal{T}_{h_n}^{(0)} = \mathcal{T}_{h_n}^A$.

The crucial step of this procedure is the derivation of a metric from the error estimator η_n^S. This is performed via suitable independent local optimization problems. We first recall that a metric is represented by a symmetric positive definite tensor field $\widetilde{\mathcal{M}} : \Omega \to \mathbb{R}^{2 \times 2}$. For our purposes it is convenient to diagonalize $\widetilde{\mathcal{M}}$ as $\widetilde{\mathcal{M}} = \widetilde{R}^T \widetilde{\Lambda}^{-2} \widetilde{R}$, with $\widetilde{\Lambda} = \operatorname{diag}(\widetilde{\lambda}_1, \widetilde{\lambda}_2)$ and $\widetilde{R}^T = [\widetilde{\mathbf{r}}_1, \widetilde{\mathbf{r}}_2]$ positive diagonal and orthogonal matrices, respectively; then we approximate the quantities $\widetilde{\lambda}_i$ and $\widetilde{\mathbf{r}}_i$ by piecewise constant functions over the triangulation $\mathcal{T}_{h_n}^{(j)}$, such that $\widetilde{\mathbf{r}}_i|_{K_n} = \widetilde{\mathbf{r}}_{i,K_n}$, $\widetilde{\lambda}_i|_{K_n} = \widetilde{\lambda}_{i,K_n}$, for any $K_n \in \mathcal{T}_{h_n}^{(j)}$ and with $i = 1, 2$.

We consider now the local error estimator $\eta_{K_n}^S$ in (35). The spatial scaling applied to the residual and the weight essentially lumps all the information related to the measure of the triangle K_n into the coefficient $\alpha_{K_n}^S$ only, at least asymptotically (i.e., when the mesh is sufficiently fine). We observe that minimizing the number

of mesh elements is equivalent to maximizing the area $|K_n|$ of each element. This consequently amounts to solving the local constrained minimization problem: find s_{K_n}, $\mathbf{r}_{1,K_n}$ such that

$$(37) \quad \Theta(s_{K_n}, \mathbf{r}_{1,K_n}) = s_{K_n}\left(\mathbf{r}_{1,K_n}^T \widetilde{G}_{K_n}^n(e_h^{*,S})\mathbf{r}_{1,K_n}\right) + \frac{1}{s_{K_n}}\left(\mathbf{r}_{2,K_n}^T \widetilde{G}_{K_n}^n(e_h^{*,S})\mathbf{r}_{2,K_n}\right)$$

is minimum, with $s_{K_n} \geq 1$, $\|\mathbf{r}_{1,K_n}\|_2 = \|\mathbf{r}_{2,K_n}\|_2 = 1$ and $\mathbf{r}_{1,K_n} \cdot \mathbf{r}_{2,K_n} = 0$, $\|\cdot\|_2$ denoting the standard Euclidean norm. As proved in [34] the solution of (37) exists unique. It is given by $\widetilde{s}_{K_n} = \left[\sigma_{1,K_n}/\sigma_{2,K_n}\right]^{1/2}$ and $\widetilde{\mathbf{r}}_{1,K_n} = \mathbf{p}_{2,K_n}$, where σ_{1,K_n} and σ_{2,K_n} are the maximum and minimum eigenvalues of $\widetilde{G}_{K_n}^n(e_h^{*,S})$, while $\mathbf{p}_{1,K_n}$ and $\mathbf{p}_{2,K_n}$ denote the associated eigenvectors. To fully characterize the metric $\widetilde{\mathcal{M}}^{(j+1)}$ we still need to compute separately the values $\widetilde{\lambda}_{1,K_n}$ and $\widetilde{\lambda}_{2,K_n}$, now "merged" in $\widetilde{s}_{K_n}$. This is achieved by imposing the spatial error equidistribution, i.e., via the identity $\eta_{K_n}^S = \tau^S/N_{h_n}^{(j)}$, for any $K_n \in \mathcal{T}_{h_n}^{(j)}$ and with $N_{h_n}^{(j)}$ the cardinality of $\mathcal{T}_{h_n}^{(j)}$. The metric $\widetilde{\mathcal{M}}^{(j+1)}$ is thus identified by the optimal values $\widetilde{\mathbf{r}}_{i,K_n}$ and $\widetilde{\lambda}_{i,K_n}$, with $i = 1, 2$, for any $K_n \in \mathcal{T}_{h_n}^{(j)}$. The new mesh $\mathcal{T}_{h_n}^{(j+1)}$ matching the metric $\widetilde{\mathcal{M}}^{(j+1)}$ is finally constructed. The mesh generator employed for this purpose is BAMG (F. Hecht [26]). We refer to [17] for the definition of matching condition between a mesh and a metric and for further details.

The third step of the whole adaptive algorithm alternates between time and space adaptivity following the criteria above, till the global tolerance is guaranteed, i.e. $\eta_n^T + \eta_n^S \leq \tau$.

Finally the fourth step can be thought of as a coarsening of the time step. In particular if the time tolerance is amply satisfied, i.e., $\eta_n^T < \theta\,\tau^T$, with $\theta \in (0, 1)$, a new (larger) time step is guessed for the next slab S_{n+1}.

The space-time algorithm above is similar to the one in [45, 11, 9]. However while in our case the space adaptivity is carried out via an optimization strategy, in [45, 11, 9] both the spatial and the temporal adaptation are driven by a compute-estimate-mark-refine procedure. The choice of favoring the time adaptivity with respect to the space adaptation is essentially dictated by the time-marching nature of the dG(0) scheme. Actually, the time-adaptivity step in the flowchart in Figure 4 is suited for a system of ODE's in time as well.

Remark 4.3. *It is worth pointing out that the optimal value $\widetilde{\Theta}$ of Θ in (37) is $\widetilde{\Theta} = 2\left[\sigma_{1,K_n}\sigma_{2,K_n}\right]^{1/2}$. Namely, $\widetilde{\Theta}$ does not depend on the ratio $\sigma_{1,K_n}/\sigma_{2,K_n}$ as the general Θ, but on the product of the two eigenvalues of $\widetilde{G}_{K_n}^n(e_h^{*,S})$. Consequently, $\widetilde{\Theta}$ does not blow up for a small σ_{2,K_n}, but rather it goes to 0.*
Therefore, as σ_{1,K_n} and σ_{2,K_n} are approximate measures of the maximum and minimum curvature of the error $e_h^{*,S}$, respectively, we are able to guarantee a negligible error even if σ_{1,K_n} is large provided that σ_{2,K_n} is small, i.e., when the error exhibits strong anisotropic features.

5. Numerical validation

We assess the reliability and the efficiency of the above procedure on some numerical tests. In particular the last one deals with a case of interest in hydrogeology. Here the anisotropic nature of problem (1) is emphasized by a heterogeneous scalar diffusivity varying anisotropically in Ω.
A rigorous reliability and efficiency analysis for the estimator η is beyond the purposes of the present work. We defer the interested reader to [34] for a possible

track. Nevertheless in the test cases below we check numerically the robustness of η and of the adaptive procedure by evaluating the corresponding effectivity index. At this stage the choice of the tolerances τ^T and τ^S of the adaptive algorithm is driven by a purely heuristic criterion. In particular we aim to yield an adaptive spatial mesh consisting of a number of elements of the order of 10^3. Of course this issue deserves a more rigorous investigation (see Section 6 for some preliminary comments).

5.1. An "ice-breaker" test case. We first assess the proposed adaptive procedure on an academic problem exhibiting clear anisotropic features.
We consider problem (1) on the space-time cylinder $\Omega \times J = (0,1)^3$, with $\alpha = 1$, $D = I$ (identity tensor), $\Gamma_N = \emptyset$ and f and u_0 chosen such that the exact solution is $u = \exp\left(-(r - r_c)^2/\sigma^2\right)$, where $r = \sqrt{(x_1 - 0.5)^2 + (x_2 - 0.5)^2}$, $r_c = 0.1 + 0.2t$, and $\sigma = 0.01$. The exact solution is thus localized in an annular region of thickness $\mathcal{O}(\sigma)$, while the median radius of the ring expands at a constant speed equal to 0.2. We apply the adaptive algorithm sketched in Figure 4 by choosing $\tau^S = 1$ and $\tau^T = 0.1$. Figure 5 shows the adapted meshes at time $t = k_1$ (top-left) and at the final time (top-right), together with two details of the last adapted mesh (bottom). The mesh is correctly detecting the anisotropic features of the solution u: the triangles are more stretched along the circumferential direction than they are in the radial one.
Figure 6 gathers the time evolution of k_n (left) and of the cardinality N_{h_n} of $\mathcal{T}_{h_n}$ (right). As the ring expands, the number of mesh elements, after a first abrupt decrease possibly biased by the choice of an inappropriate initial mesh, grows because of the enlargement of the area of the annular region (the thickness is constant but the perimeter grows). On the other hand, the time step decreases steadily with time, approaching a limiting value for large t. The annular region asymptotically tends to a straight line advancing at a constant speed. A constant time step is consequently expected.

To check the robustness of the estimator (34), we compute the effectivity index E.I. $= \eta/|||e_h|||_{\mathrm{DG}}$. This quantity provides us with a measure of both the reliability and the efficiency of the estimator and of the whole adaptive procedure. We obtain $N = 734$, $|||e_h|||^2_{\mathrm{DG}} = 9.21169$, $\sum_{n=1}^N \eta_n^T = 0.212242$, $\sum_{n=1}^N \eta_n^S = 111.454$, i.e., via (34), E.I. $= 3.4817$. This value is not so unusual in our experience with anisotropically adapted meshes. Similar values can be found for instance in [34] in the case of stationary purely diffusion problems. Likely, the smallness of the time error contribution with respect to the spatial one does not modify substantially the results obtained for the steady case. Overall, the whole adaptive procedure turns out to be efficient and reliable moving from the effectivity index. On the other hand, a more rigorous theoretical analysis would be advisable.

To further assess the effectiveness of the whole adaptive procedure, we repeat the same test case by enforcing an isotropic mesh adaptation. This is achieved by skipping the solution to (37), and letting $s_{K_n} = 1$, i.e., $\widetilde{\lambda}_{1,K_n} = \widetilde{\lambda}_{2,K_n}$. Figure 7 gathers the corresponding adapted meshes at time $t = k_1$ (top-left) and at the final time (top-right), along with two details of the last adapted mesh (bottom). The time evolution of k_n and of the cardinality N_{h_n} of $\mathcal{T}_{h_n}$ are shown in Figure 8. On comparing these last two figures with Figures 5 and 6, we see that also in the isotropic case the mesh is correctly detecting the expansion of the ring, while both the trends of the time step and of the number of mesh elements match qualitatively the corresponding anisotropic ones. However, the cardinality of the isotropic spatial meshes outnumbers the anisotropic one thrice. Concerning the
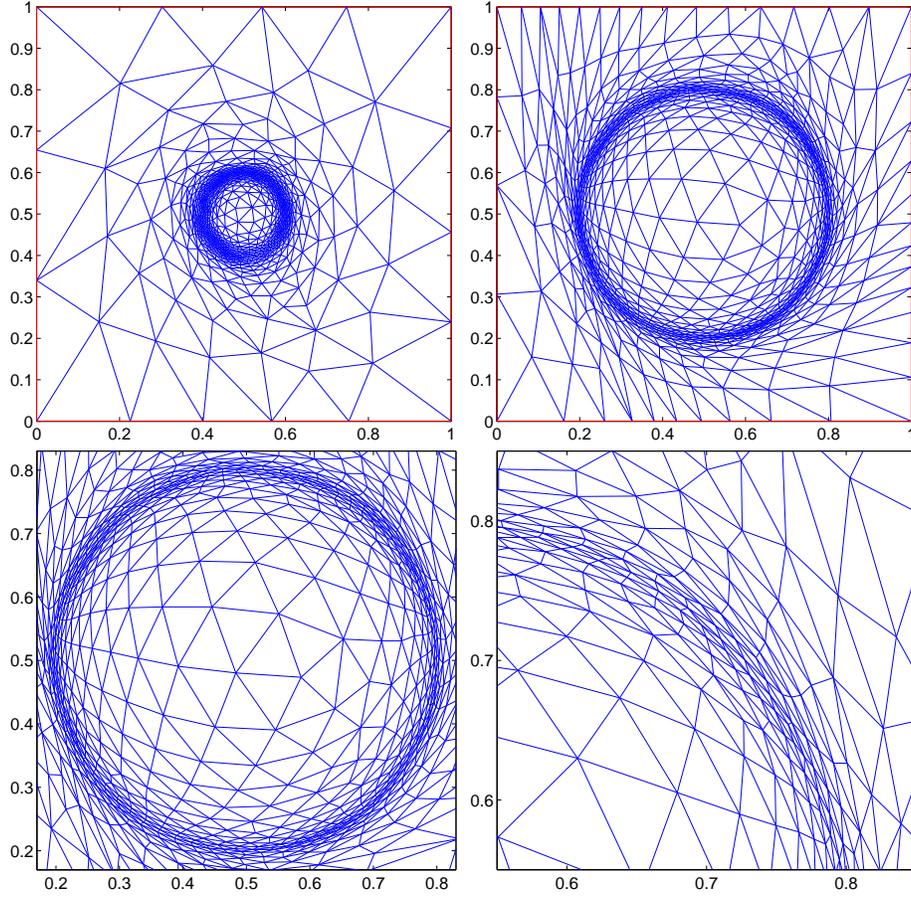
FIGURE 5. Adapted mesh at $t = k_1$ (top-left) and $t = T$ (top-right); details of the last adapted mesh (bottom).
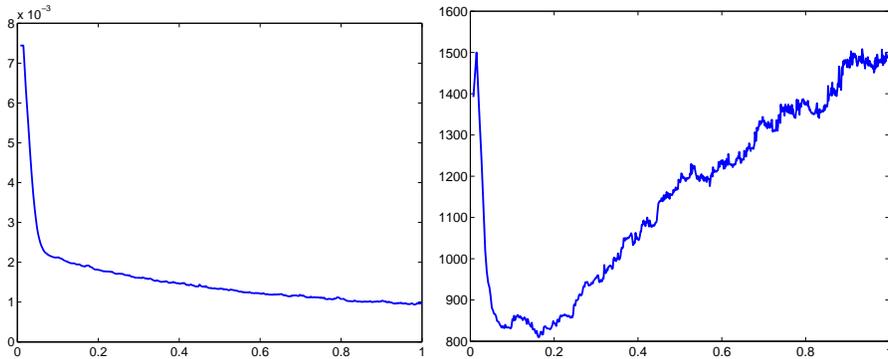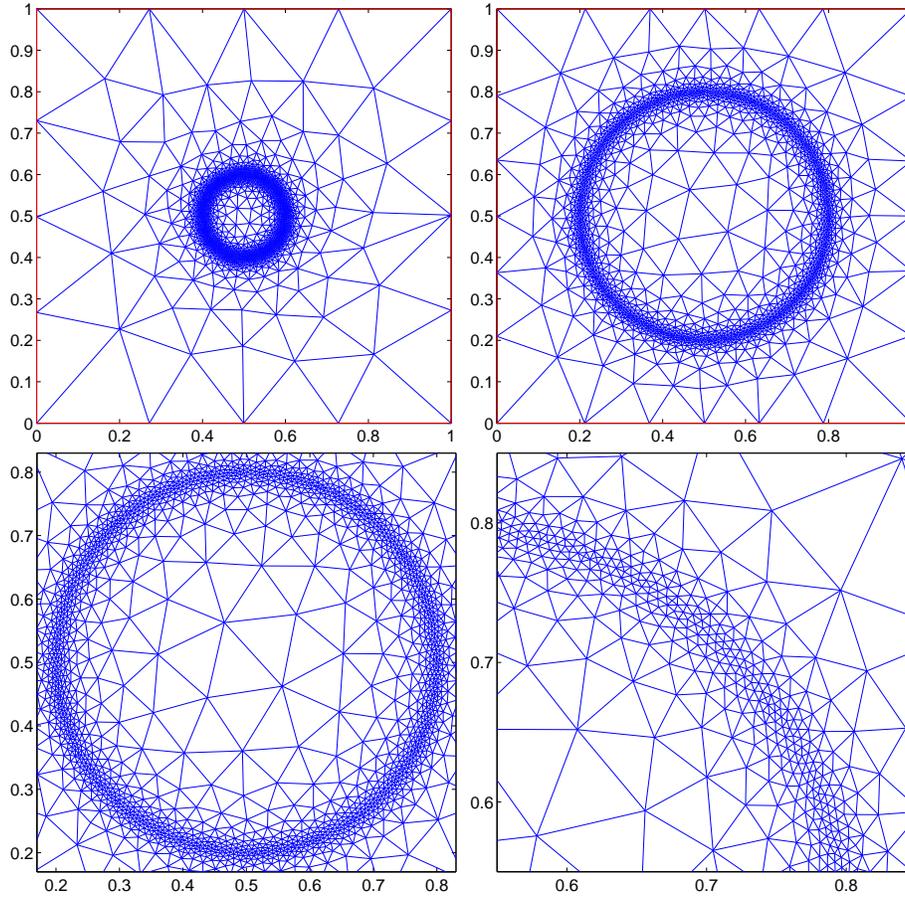

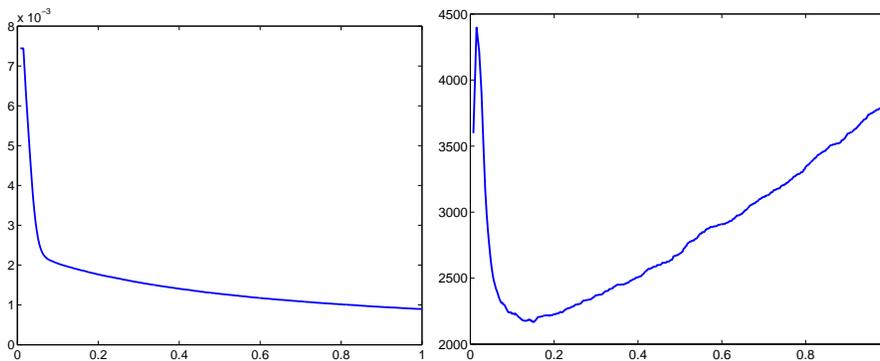FIGURE 6. Time evolution of the time step (left) and of the number of mesh elements (right).

time discretization, we get $N = 765$, while as for the exact error and the time and spatial error estimators we obtain $|||e_h|||^2_{\mathrm{DG}} = 11.4864, \sum_{n=1}^{N} \eta_n^T = 0.206169$,

FIGURE 7. Isotropic adapted mesh at $t = k_1$ (top-left) and $t = T$ (top-right); details of the last adapted mesh (bottom).

FIGURE 8. Time evolution of the time step (left) and of the number of the isotropic mesh elements (right).

$\sum_{n=1}^{N} \eta_n^S = 255.642$, implying E.I. = 4.7195. The error in the isotropic case is very close to the isotropic one. To summarize, the computational effort associated with
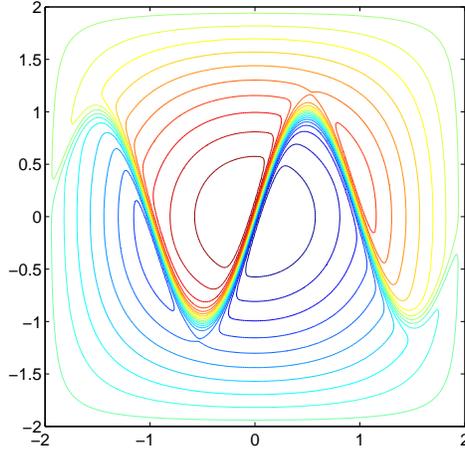
FIGURE 9. Contour plot of the initial solution u_0.

the isotropic procedure, for about the same number of slabs and accuracy, is larger
with respect to the anisotropic case, due to the higher number of mesh elements
(approximately three times as much).

5.2. An application to heat flow. We choose the space-time cylinder $\Omega \times J =
(-2, 2)^2 \times (0, 4)$, and pick the data in (1) as $\alpha = 1$, $D = I$, $\Gamma_N = \emptyset$, and f and u_0
ensuring that $u = \cos(\pi x_1/4) \cos(\pi x_2/4) \tanh \left[10 \big(x_2 - \cos(\pi t/2) \sin(\pi x_1) \big) \right]$ is the
exact solution to (1). The function u exhibits a steep gradient across the curve $x_2 =
\cos(\pi t/2) \sin(\pi x_1)$, i.e., a sine function harmonically varying in time. This test case
might model, roughly, the conduction of thermal energy through a matter, due to a
serpentine-shaped heat source which moves with time periodically. Figure 9 shows
a contour plot of the initial condition $u_0(\mathbf{x}) = \cos(\pi x_1/4) \cos(\pi x_2/4) \tanh \left[10 \big(x_2 -
\sin(\pi x_1) \big) \right]$. Notice the crowding of the contour lines around the three points of
inflexion, (-1,0), (0,0) and (1,0). During the time window of the simulation a full
period of the oscillation occurs, so that the sine function vanishes at $t = 1$, reverses
at $t = 2$, vanishes again at $t = 3$ and comes back to the original configuration at
$t = 4$.

We assess the performance of the adaptive procedure in Section 4.2 with the
choice $\tau^S = 1$ and $\tau^T = 0.1$ for the spatial and temporal tolerance, respectively.
The adapted meshes corresponding to the four levels $t = 1, 2, 3, 4$ are displayed in
Figure 10. The features of the exact solution are correctly detected throughout
the simulation. Two details of the adapted mesh at the final time are shown in
Figure 11 to highlight the anisotropic nature of the grid. Observe that the triangles
are mostly stretched in correspondence with the three points of inflexion.
The time histories of the time step and of the number of mesh elements, gathered
in Figure 12, show that: the time step is small around the time levels $t = 0, 2, 4$,
i.e., when the phenomenon to be described achieves its maximum spatial exten-
sion; it is larger around the time levels $t = 1, 3$, i.e., when the phenomenon
is negligible in practice. This behaviour can be explained by the contribution
$\|\rho_{K_n} - \overline{\rho}_{K_n}\|_{L^2(S_{K_n})}$ in the time residual $R_{K_n}^{T1}$ in (25). In this case this norm amounts
to $\|f - T_n f\|_{L^2(S_{K_n})}$, for which relation $\|f - T_n f\|_{L^2(S_{K_n})} \leq k_n \|\partial f/\partial t\|_{L^2(S_{K_n})}$
holds. Thus, k_n is expected to be inversely proportional to the size of $\partial f/\partial t$, which
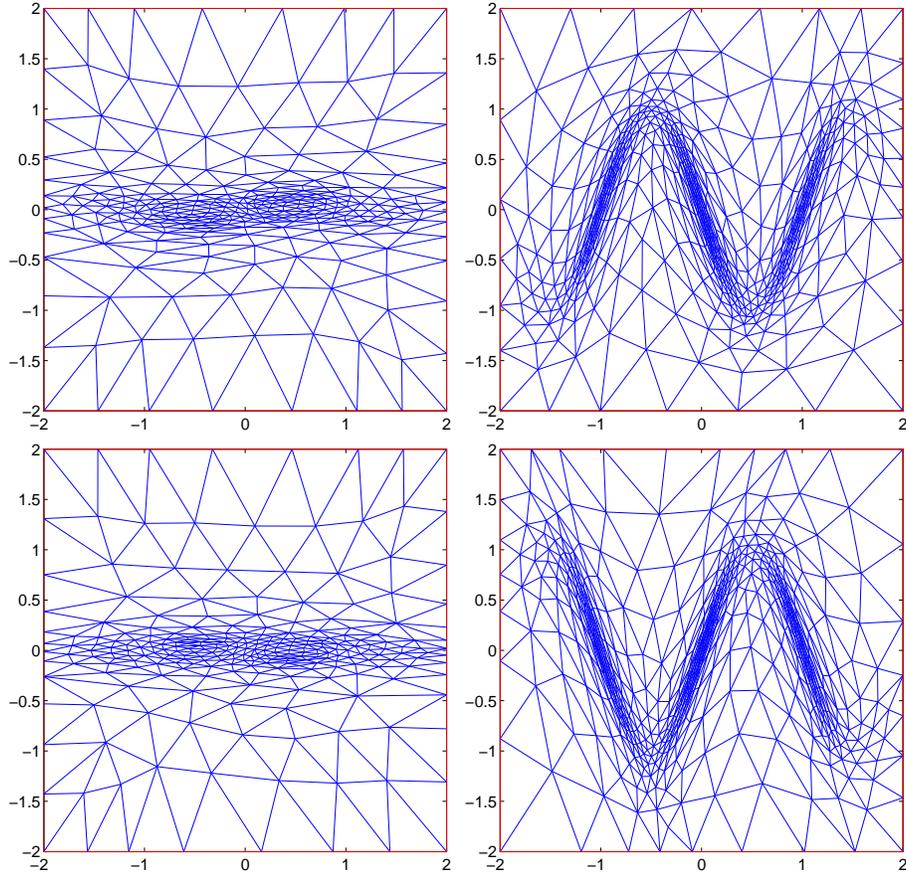achieves its maxima at $t = 0, 2, 4$.

FIGURE 10. Adapted mesh at the time level $t = 1$ (top-left), $t = 2$ (top-right), $t = 3$ (bottom-left) and $t = 4$ (bottom-right).
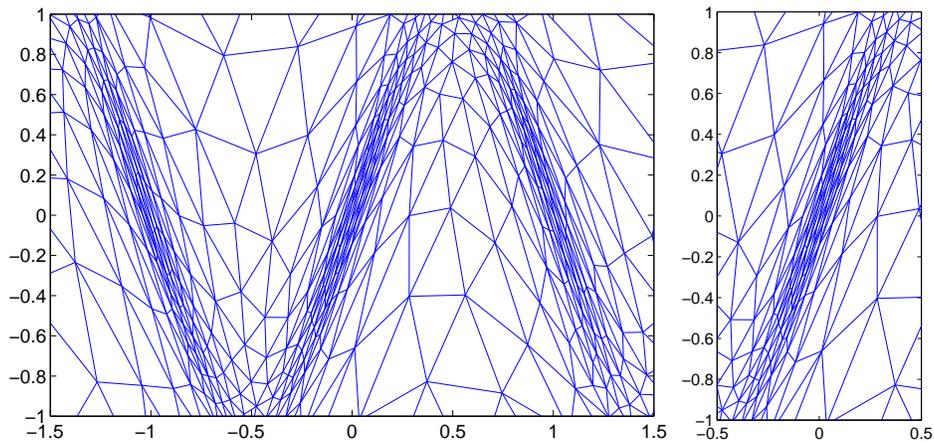

FIGURE 11. Details of the adapted mesh at the time level $t = 4$.

On the other hand, the number of mesh elements is small when the solution vanishes, and reaches its maximum at $t \simeq 1.5, 3.5$, that is, midway between the vanishing and the maximum extension phases. As far as the spatial error is concerned,

it seems, reasonably, more problematic to describe the transition from the flat status to the full vertical configuration than to capture this latter phase when the phenomenon stagnates.

As for the accuracy check, we obtain $|||e_h|||^2_{\mathrm{DG}} = 25.5276$, $\sum_{n=1}^{N} \eta_n^T = 3.21921$, $\sum_{n=1}^{N} \eta_n^S = 104.952$, i.e., E.I. $= 2.05849$. As in the previous test case, most of the error is due to the space approximation. Again the whole procedure goes to prove the efficiency and reliability.
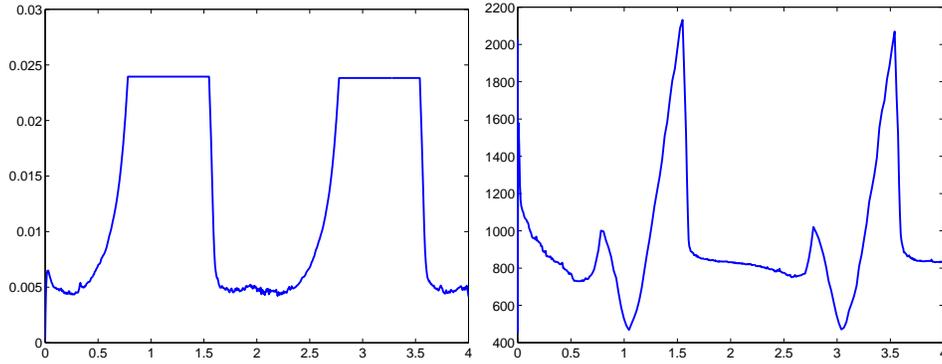
FIGURE 12. Time evolution of the time step (left) and of the number of mesh elements (right).

5.3. An application to hydrogeology. We simulate a so-called pumping test in the ambit of flow to wells in confined aquifers: water is extracted at a given rate from a well and the drawdown, that is the deviation from a reference level of the hydraulic head, is measured in a nearby observation well (see Figure 13). In practice
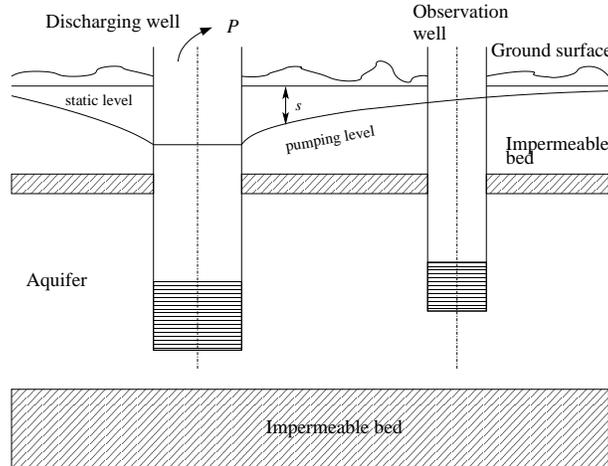

FIGURE 13. Cross section through a discharging well screened in a part of a nonleaky aquifer.

the actual purpose of this type of experiment is to solve an inverse problem: to infer, from these measurements, the hydraulic properties of water-bearing and associated rocks (J.E. Reed [42]). As far as we are concerned, we assume that such properties

are known and we compute instead the drawdown throughout the aquifer. This represents an interesting application of (1) in an axisymmetric context under the hypothesis of a partially penetrating well in a nonleaky aquifer. In more detail, we choose as computational domain a hollow tube of inner radius r_w (corresponding to the radius of the well) and outer radius R; vertically it is divided into three beds: a central aquifer for $z_1 < z < z_2$, and two aquitards, the lower one for $0 < z < z_1$ and the upper one for $z_2 < z < z_3$ (see Figure 14 (left)). The well is screened only
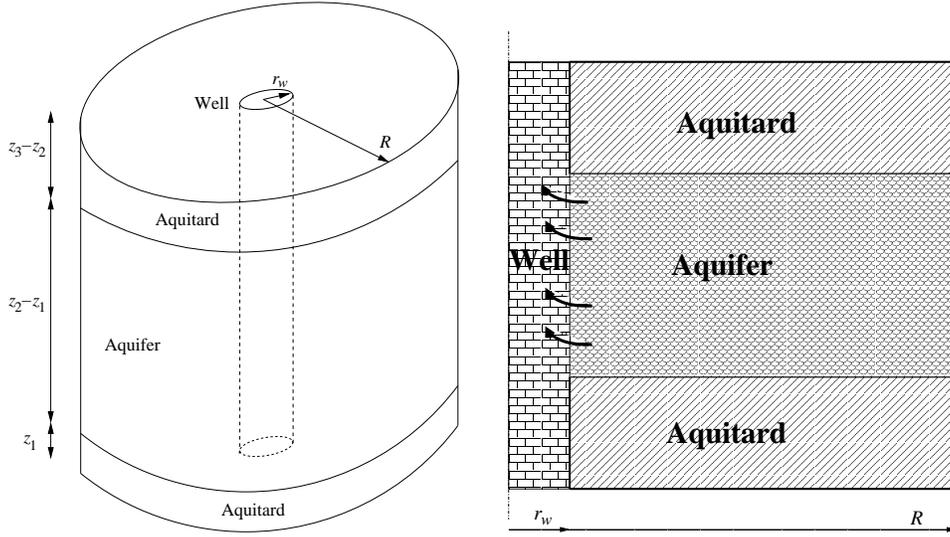

FIGURE 14. Geometry of the discharging well in 3D (left) and axisymmetric approximation (right).

in the aquifer, i.e., water is extracted from the well near the aquifer only. Under axisymmetric conditions, the real model is given by

$$
(38) \quad
\begin{cases}
\phi \dfrac{\partial s}{\partial t} - \nabla \cdot (T_r \nabla s) = 0 & (\mathbf{y}, t) \in \mathcal{Q} = \Omega \times J, \\[2mm]
s(\mathbf{y}, t) = 0 & (\mathbf{y}, t) \in \partial \mathcal{Q}_D = \Gamma_D \times J, \\[2mm]
r T_r \nabla s \cdot \mathbf{n} = P(2\pi)^{-1} & (\mathbf{y}, t) \in \partial \mathcal{Q}_{N_0} = \Gamma_{N_0} \times J, \\[2mm]
r T_r \nabla s \cdot \mathbf{n} = 0 & (\mathbf{y}, t) \in \partial \mathcal{Q}_{N_i} = \Gamma_{N_i} \times J, \ i = 1, \dots, 4, \\[2mm]
s(\mathbf{y}, 0) = 0 & \mathbf{y} \in \Omega,
\end{cases}
$$

where $\mathbf{y} = (r, z)$ collects the radial and the axial coordinate, respectively, the unknown s [m] models the drawdown, ϕ (dimensionless) is the storage coefficient, T_r [m^{2}s^{-1}] is the transmissivity, P [m^{3}s^{-1}] is the pumping rate (negative if water is extracted from the well), and

$$
\nabla \cdot (T_r \nabla s) = \frac{1}{r} \frac{\partial}{\partial r} \left(r \, T_r \frac{\partial s}{\partial r} \right) + \frac{\partial}{\partial z} \left(T_r \frac{\partial s}{\partial z} \right)
$$

is the divergence operator in polar coordinates. Concerning the computational domain, Ω coincides with the rectangular region $\{r_w < r < R\} \times \{0 < z < z_3\}$ (see Figure 14 (right)), the Dirichlet boundary is $\Gamma_D = \{r = R\} \times \{0 < z < z_3\}$, while

the five Neumann boundaries are

$$\Gamma_{N_0} = \{r = r_w\} \times \{z_1 < z < z_2\}, \quad \Gamma_{N_1} = \{r = r_w\} \times \{0 < z < z_1\},$$

$$\Gamma_{N_2} = \{r = r_w\} \times \{z_2 < z < z_3\}, \quad \Gamma_{N_3} = \{r_w < r < R\} \times \{z = 0\},$$

$$\Gamma_{N_4} = \{r_w < r < R\} \times \{z = z_3\}.$$

Observe that the Γ_{N_i}'s, for $i = 1, \ldots, 4$ are impermeable. The axisymmetric formulation of (38) is almost identical to the two-dimensional Cartesian case considered in (1) (T.J.R. Hughes [28]). The only difference is that a factor of $2\pi r$ needs to be included in each integrand of the variational equation to account for the correct volumetric weighting (i.e., $d\mathbf{x} = 2\pi r \, drdz$ replaces $d\mathbf{x} = dx_1 dx_2$). The common term 2π can then be canceled throughout. Thus the weak form of (38) can be cast in the setting (2), where $x_1 = r$, $x_2 = z$, $\alpha = x_1\phi$, $D = x_1 T_r$, $f = 0$, and g is equal to $P(2\pi)^{-1}$ on the lateral surface pertaining to the aquifer and zero in correspondence with the aquitards.

The actual values for the data are: $z_1 = 25$, $z_2 = 75$, $z_3 = 100$, $r_w = 1$, $R = 200$, $\phi = 1$, $T_r = 1$ in the aquitards and $T_r = 100$ in the aquifer, $P = -0.4 \exp(-(t-1)^2)$, and $T = 20$. In the following, for computational ease, all the spatial variables are scaled with respect to 100. The pumping rate shows a Gaussian variation profile in time with mean 1, variance 1 and peak value 0.4. This represents the driving force of the problem. The peak value of the pumping rate is reached at $t = 1$: shortly afterward, due to the increase draw of water leaving the aquifer, the gradient of the drawdown in correspondence with the lateral surface of the well becomes higher and higher. When the effect of the pumping fades away, the drawdown extinguishes towards a steady configuration where $s = 0$ everywhere.

We assess the space-time adaptive procedure above setting the tolerances as $\tau^S = 10^{-10}$ and $\tau^T = 10^{-11}$. This choice is motivated by the expected small size of the drawdown. The time histories of the time step (in log scale) and of the
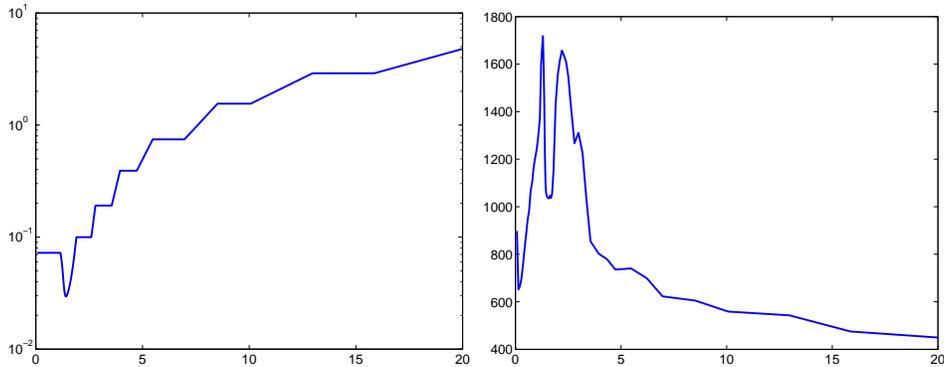

FIGURE 15. Time evolution of the time step (left) and of the number of mesh elements (right).

cardinality of the meshes are gathered in Figure 15. Essentially, the number of mesh elements reaches a local minimum (1035) at $t = 1.55$ and two local maxima at $t = 1.30$ (1718) and $t = 2.2$ (1657). Afterwards, this number gets smaller and smaller while the pumping is extinguishing, until the final time is reached (only 438 triangles suffice in this case). While the minimum can be ascribed to the presence of three really steep layers (for which an efficient anisotropic grid consists of few appropriately stretched triangles), the second maximum can be justified by the

delayed contribution of the two aquitards (compare also the zooms in Figure 17). On the other hand the behaviour of the time step shows a dip in the time interval $1.16 \div 1.9$ with a global minimum occurring at $t = 1.4$ (with $k_n = 0.0295$). Then it gradually increases as the phenomenon is reaching the steady-state configuration. Notice that at $t = T$ a time step equal to 6.0026 is employed.
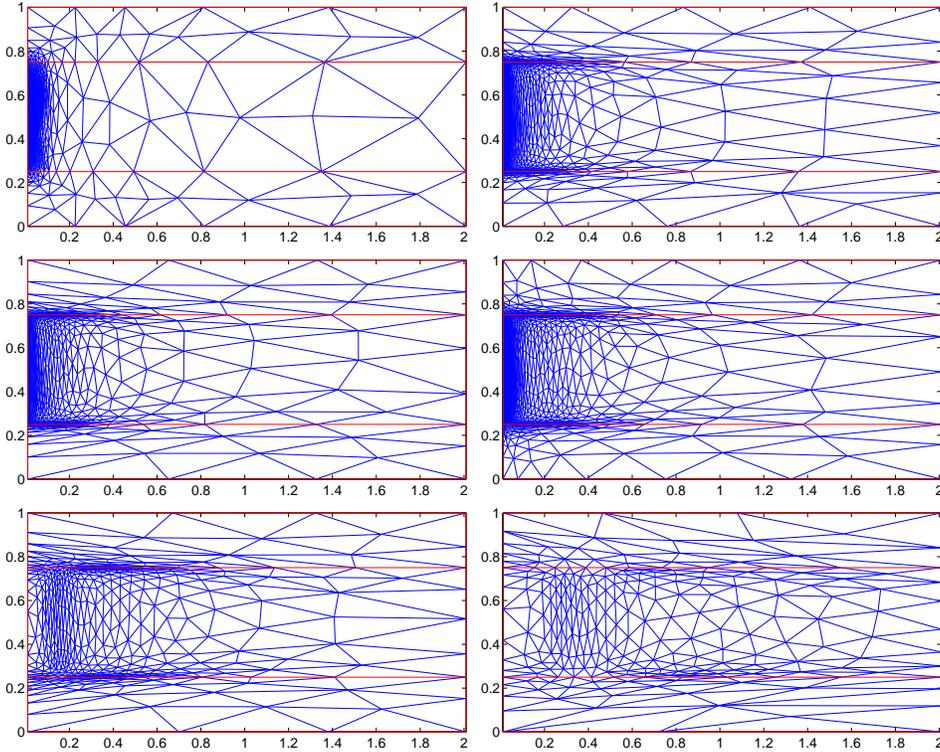
FIGURE 16. Adapted mesh at the time level $t = k_1$ (top-left), $t = 1.30$ (top-right), $t = 1.55$ (center-left), $t = 2.22$ (center-right), $t = 4.73$ (bottom-left) and $t = 20$ (bottom-right).
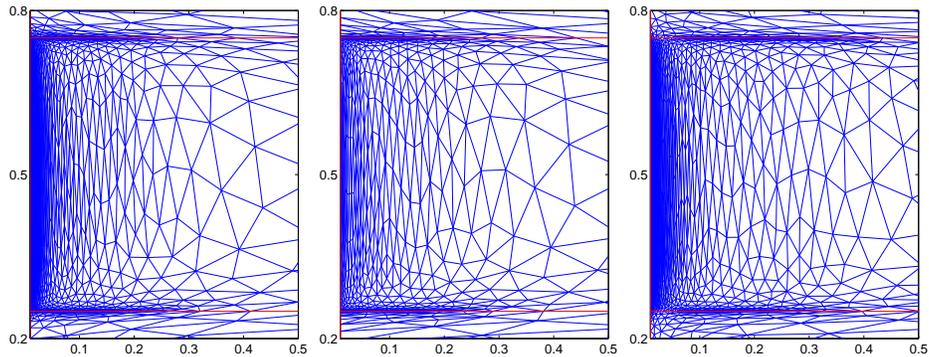
FIGURE 17. Details of the adapted mesh at the time level $t = 1.30$ (left), $t = 1.55$ (center) and $t = 2.22$ (right).

Figure 16 collects the adapted meshes corresponding to $t = k_1$ (top-left), $t = 1.30$ (top-right), $t = 1.55$ (center-left), $t = 2.22$ (center-right), $t = 4.73$ (bottom-left), $t = 20$ (bottom-right). The anisotropic features emphasize an orthogonal pattern: on the one hand, due to the pumping, the main flow is radial close to the well and takes place mostly in the aquifer, as the aquitards are much less conductive; on the other hand, the jump of the transmissivity across the beds is responsible for the horizontal layers.

Finally the details of the meshes associated with the two local maxima and with the local minimum of the mesh elements are displayed in Figure 17. Notice the different extent of crowding near the inner lateral surface of the aquifer and the different extent of refinement across the three beds.

6. What's ahead

We have proposed a space-time adaptive procedure for a cG(1)dG(0) approximation of the model problem (1). It is based on an anisotropic a posteriori error estimator for the cGdG energy norm of the space-time discretization error. A weak point of the presented procedure is the "empirical" management of the spatial and temporal tolerances τ^S and τ^T. Indeed in the test cases of Section 5 neither the total number of time intervals nor of the slabwise triangles is fixed a priori, nor even any threshold on the total error is demanded. In such a case it is the same adaptive algorithm which is steering these choices.

However, a simple modification to the algorithm sketched in Figure 4 allows us to set the total accuracy. Let ε^T and ε^S be the desired total time and space tolerances, respectively. It suffices to replace in the presented algorithm the slabwise constant time and space tolerances τ^T and τ^S with $\tau_n^T = (\varepsilon^T/T) k_n$ and $\tau_n^S = (\varepsilon^S/T) k_n$, respectively for any S_n with $n = 1, \ldots, N$. Notice that now τ_n^T and τ_n^S depend on the unknown time step and vary across the space-time slabs.

On the other hand a procedure aiming at minimizing the total error, for a fixed number of time intervals and a maximum number of mesh elements would imply solving several times the full problem on the whole time window in order to meet the constraint $\sum_{n=1}^N k_n = T$.

Other important aspects deserve to be investigated in the future. For instance it could be interesting to provide a rigorous analysis of the reliability and of the efficiency of the error estimator (34). The reliability of η is not necessarily guaranteed a priori, due to the recovery procedures (31) and (32) and to the approximation (33). Heuristically, we expect that to be the case though, as long as the recovered space-time derivatives of the error enjoy superconvergent properties. As far as the efficiency is concerned, we believe that it holds in practice due to the particular optimized adaptive procedure employed. A possible track for both a reliability and an efficiency analysis of η can be found in [34], where the case of a reaction-diffusion elliptic problem is dealt with.

The numerical assessment in Section 5 can at the moment fill in for the lack of a rigorous theoretical argument. The effectivity index E.I. evaluated in the first two test cases where an exact solution is available, is of the order of $2 \div 3$. This strengthens the robustness of the estimator η. Most importantly, in all the test cases the E.I. turns out to be independent of both the variable time step and of the stretching factor of the different slabwise meshes.

Our interest for problems related to practical applications will unavoidably lead to generalize the proposed analysis to more complex problems. We refer to S.

Micheletti & S. Perotto [36] for some preliminary results), as well as to a goal-oriented philosophy (see, e.g., W. Bangerth and R. Rannacher [4], P. Houston & E. Süli [27]), also in an optimal control framework ([33]). Yet the most crucial issue to be faced in such a framework turns out to be the efficient management of the reverse-in-time dual problem. This aspect represents another interesting issue to be investigated.

References

[1] G. Akrivis, C. Makridakis and R. Nochetto. A posteriori error estimates for the Crank-Nicolson method for parabolic equations. *Math. Comp.*, 75:511–531, 2006.

[2] T. Apel. Anisotropic Finite Elements: Local Estimates and Applications. Advances in Numerical Mathematics, Teubner, Stuttgart, 1999.

[3] A.K. Aziz and P. Monk. Continuous finite elements in space and time for the heat equation. *Math. Comp.*, 52:255–274, 1989.

[4] W. Bangerth and R. Rannacher. Adaptive Finite Element Methods for Differential Equations. Birkhauser Verlag, Basel, 2003.

[5] R. Becker and R. Rannacher. An optimal control approach to a posteriori error estimation in finite element methods. *Acta Numer.*, 10:1–102, 2001.

[6] J.C. Bruch. Free surface seepage problems solved using: a) the Zienkiewicz-Zhu error estimation procedure; and b) a parallel computer. *Pitman Res. Notes Math. Ser.*, 282:98–104, 1993.

[7] W. Cao. On the error of linear interpolation and the orientation, aspect ratio and internal angles of a triangle. *SIAM J. Numer. Anal.*, 43(1):19–40, 2005.

[8] W. Cao. An interpolation error estimate in $\mathcal{R}^2$ based on the anisotropic measures of higher order derivatives. *Math. Comp.*, 77(261):265–286, 2007.

[9] J.M. Cascón, L. Ferragut and M.I. Asensio. Space-time adaptive algorithm for the mixed parabolic problem. *Numer. Math.*, 103:367–392, 2006.

[10] M.J. Castro–Díaz, F. Hecht, B. Mohammadi and O. Pironneau. Anisotropic unstructured mesh adaption for flow simulations. *Internat. J. Numer. Methods Fluids*, 25(4):475–491, 1997.

[11] Z. Chen and J. Feng. An adaptive finite element algorithm with reliable and efficient error control for linear parabolic problems. *Math. Comp.*, 73(247):1167–1193, 2004.

[12] Ph. Ciarlet. The Finite Element Method for Elliptic Problems. North-Holland, Amsterdam, 1978.

[13] Ph. Clément. Approximation by finite element functions using local regularization. *RAIRO Anal. Numér.*, 2:77–84, 1975.

[14] B. Cockburn and C.-W. Shu. Runge-Kutta discontinuous Galerkin methods for convection-dominated problems. *J. Sci. Comput.*, 16(3):173–261, 2001.

[15] D.L. Darmofal and D.A. Venditti. Grid adaption for functional outputs: application to two–dimensional inviscid fluids. *J. Comput. Phys.*, 176(1):40–69, 2002.

[16] R. Dautray and J. -L. Lions. Mathematical Analysis and Numerical Methods for Science and Technology: Evolution Problems I. Vol. 5, Springer-Verlag, Berlin, 1992.

[17] L. Dedè, S. Micheletti and S. Perotto. Anisotropic error control for environmental applications. *Appl. Numer. Math.*, 58:1320–1339, 2008.

[18] J. Dompierre, M.-G. Vallet, Y. Bourgault, M. Fortin and W.G. Habashi. Anisotropic mesh adaptation: towards user-independent, mesh-independent and solver-independent CFD. Part III. Unstructured meshes. *Internat. J. Numer. Methods Fluids*, 39:675–702, 2002.

[19] K. Eriksson, D. Estep, P. Hansbo and C. Johnson. Computational Differential Equations. Cambridge University Press, Cambridge, 1996.

[20] K. Eriksson and C. Johnson. Adaptive finite element methods for parabolic problems. I: a linear model problem. *SIAM J. Numer. Anal.*, 28:43–77, 1991.

[21] K. Eriksson, C. Johnson and V. Thomée. Time discretization of parabolic problems by the discontinuous Galerkin method. *RAIRO Modelisation Math. Anal. Numer.*, 19:611–643, 1985.

[22] L. Formaggia and S. Perotto. New anisotropic a priori error estimates. *Numer. Math.*, 89:641–667, 2001.

[23] L. Formaggia and S. Perotto. Anisotropic error estimates for elliptic problems. *Numer. Math.*, 94:67–92, 2003.

[24] E.H. Georgoulis, E. Hall and P. Houston. Discontinuous Galerkin methods for advection-diffusion-reaction problems on anisotropically refined meshes. *SIAM J. Sci. Comput.*, 30(1):246–271, 2007.

[25] R. Hartmann and P. Houston. Adaptive discontinuous Galerkin finite element methods for nonlinear hyperbolic conservation laws. *SIAM J. Sci. Comput.*, 24(3):979–1004, 2002.

[26] F. Hecht. BAMG: bidimensional anisotropic mesh generator.
http://www.ann.jussieu.fr/~hecht/ftp/bamg, 2006.

[27] P. Houston and E. Süli. Adaptive Lagrange-Galerkin methods for unsteady convection-diffusion problems. *Math. Comp.*, 70(233):77–106, 2001.

[28] T.J.R. Hughes. The Finite Element Method. Linear Static and Dynamic Finite Element Analysis. Dover Publications Inc., Mineola, New-York, 2000.

[29] P. Jamet. Galerkin-type approximations which are discontinuous in time for parabolic equations in a variable domain. *SIAM J. Numer. Anal.*, 15:912–928, 1978.

[30] D. Kröner and M. Ohlberger. A posteriori error estimate for upwind finite volume schemes for nonlinear conservation laws in multidimensions. *Math. Comp.*, 69(229):25–39, 2000.

[31] K.L. Lawrence and R.V. Nambiar. The Zienkiewicz-Zhu error estimator for multiple material problems. *Commun. Appl. Numer. Methods*, 8:273–277, 1992.

[32] G. Maisano, S. Micheletti, S. Perotto and C.L. Bottasso. On some new recovery based a posteriori error estimators. *Comput. Methods Appl. Mech. Engrg.*, 195(37–40):4794–4815, 2006.

[33] D. Meidner and B. Vexler. Adaptive space-time finite element methods for parabolic optimization problems. *SIAM J. Control Optim.*, 46(1):116–142, 2007.

[34] S. Micheletti and S. Perotto. Reliability and efficiency of an anisotropic Zienkiewicz–Zhu error estimator. *Comput. Methods Appl. Mech. Engrg.*, 195(9–12):799–835, 2006.

[35] S. Micheletti and S. Perotto. Anisotropic mesh adaption for time-dependent problems. *Int. J. Numer. Meth. Fluids*, 58:1009–1015, 2008.

[36] S. Micheletti and S. Perotto. Space–Time Adaption for Advection-Diffusion-Reaction Problems on Anisotropic Meshes. In Numerical Mathematics and Advanced Applications (Eds. K. Kunisch, G. Of, O. Steinbach), p. 49–56, Springer-Verlag, 2008.

[37] S. Micheletti, S. Perotto and M. Picasso. Stabilized finite elements on anisotropic meshes: a priori error estimates for the advection-diffusion and the Stokes problems. *SIAM J. Numer. Anal.*, 41(3):1131–1162, 2003.

[38] R. Nochetto, A. Schmidt and C. Verdi. A posteriori error estimation and adaptivity for degenerate parabolic problems. *Math. Comp.*, 69:1–24, 2000.

[39] T.P. Pawlak, M.J. Wheeler and S.M. Yunus. Application of the Zienkiewicz-Zhu error estimator for plate and shell analysis. *Int. J. Numer. Methods Eng.*, 29:1281–1298, 1990.

[40] M. Picasso. Adaptive finite elements for a linear parabolic problem. *Comput. Methods Appl. Mech. Engrg.*, 167:223–237, 1998.

[41] M. Picasso. Anisotropic a posteriori error estimate for an optimal control problem governed by the heat equation. *Numer. Methods Partial Differential Equations*, 22(6):1314–1336, 2006.

[42] J.E. Reed. Type Curves for Selected Problems of Flow to Wells in Confined Aquifers. U.S. Geological Survey, Techniques of Water-Resources Investigations, Book 3, Chapter B3. http://pubs.usgs.gov/twri/twri3-b3/

[43] R. Rodriguez. Some remarks on Zienkiewicz–Zhu estimator. *Numer. Methods Partial Differential Equations*, 10(5):625–635, 1994.

[44] M. Schmich and B. Vexler. Adaptivity with dynamic meshes for space-time finite element discretizations of parabolic equations. *SIAM J. Control Optim.*, 30(1):369–393, 2008.

[45] A. Schmidt and K.G. Siebert. ALBERT: an adaptive hierarchical finite element toolbox. IAM, University of Freiburg, 2000.
http://www.mathematik.uni-freiburg.de/IAM/Research/projectsdz/albert

[46] L.R. Scott and S. Zhang. Finite element interpolation of non-smooth functions satisfying boundary conditions. *Math. Comp.*, 54:483–493, 1990.

[47] K.G. Siebert. An a posteriori error estimator for anisotropic refinement. *Numer. Math.*, 73:373–398, 1996.

[48] J.J. Sudirham, J.J.W. van der Vegt and R.M.J. van Damme. Space-time discontinuous Galerkin method for advection-diffusion problems on time-dependent domains. *Appl. Numer. Math.*, 56: 1491–1518, 2006.

[49] V. Thomée. Galerkin Finite Element Methods for Parabolic Problems. Springer Series in Computational Mathematics vol. 25., 2nd edn, Springer-Verlag, Berlin, 2006.

[50] R. Verfürth. A Review of A Posteriori Error Estimation and Adaptive Mesh–Refinement Techniques. Wiley–Teubner, New York, 1996.

[51] R. Verfürth. A posteriori error estimate for finite element discretizations of the heat equation. *Calcolo*, 40:195–212, 2003.

[52] O. C. Zienkiewicz and J. Z. Zhu. A simple error estimator and adaptive procedure for practical engineering analysis. *Int. J. Numer. Methods Engrg.*, 24(2):337–357, 1987.

[53] O. C. Zienkiewicz and J. Z. Zhu. The superconvergent patch recovery and a posteriori error estimates. Part I: the recovery technique. *Int. J. Numer. Methods Engrg.*, 33(7):1331–1364, 1992.

[54] O. C. Zienkiewicz and J. Z. Zhu. The superconvergent patch recovery and a posteriori error estimates. Part 2: error estimates and adaptivity. *Int. J. Numer. Methods Engrg.*, 33(7):1365–1382, 1992.

MOX–Dipartimento di Matematica "F. Brioschi", Politecnico di Milano, Via Bonardi 9, I-20133 Milano, Italy

E-mail: `stefano.micheletti@polimi.it and simona.perotto@polimi.it`

URL: `http://www1.mate.polimi.it/∼mike/ and http://www1.mate.polimi.it/∼simona/`