




# Space-Time Finite Element Methods for Parabolic Initial-Boundary Value Problems with Non-smooth Solutions

Ulrich Langer<sup>1</sup>  and Andreas Schafelner<sup>2</sup>

<sup>1</sup> Institute for Computational Mathematics, Johannes Kepler University Linz, Altenbergerstr. 69, 4040 Linz, Austria

[ulanger@numa.uni-linz.ac.at](mailto:ulanger@numa.uni-linz.ac.at)

<sup>2</sup> Doctoral Program “Computational Mathematics”, Johannes Kepler University Linz, Altenbergerstr. 69, 4040 Linz, Austria

[andreas.schafelner@dk-compmath.jku.at](mailto:andreas.schafelner@dk-compmath.jku.at)

<http://www.numa.uni-linz.ac.at>, <https://www.dk-compmath.jku.at>

**Abstract.** We propose consistent, locally stabilized, conforming finite element schemes on completely unstructured simplicial space-time meshes for the numerical solution of parabolic initial-boundary value problems under the assumption of maximal parabolic regularity. We present new a priori discretization error estimates for low-regularity solutions, and some numerical results including results for an adaptive version of the scheme and strong scaling results.

**Keywords:** Parabolic initial-boundary-value problems · Space-time finite element methods · Unstructured meshes · Adaptivity

## 1 Introduction

Parabolic initial-boundary value problems of the form

$$\partial_t u - \operatorname{div}_x(\nu \nabla_x u) = f \text{ in } Q, \quad u = 0 \text{ on } \Sigma, \quad u = u_0 \text{ on } \Sigma_0 \quad (1)$$

describe not only heat conduction and diffusion processes but also 2D eddy current problems in electromagnetics and many other evolution processes, where  $Q = \Omega \times (0, T)$ ,  $\Sigma = \partial\Omega \times (0, T)$ , and  $\Sigma_0 = \Omega \times \{0\}$  denote the space-time cylinder, its lateral boundary, and the bottom face, respectively. The spatial computational domain  $\Omega \subset \mathcal{R}^d$ ,  $d = 1, 2, 3$ , is supposed to be bounded and Lipschitz. The final time is denoted by  $T$ . The right-hand side  $f$  is a given source function from  $L_2(Q)$ . The given coefficient  $\nu$  may depend on the spatial variable  $x$  as well as the time variable  $t$ . In the latter case, the problem is called non-autonomous. We suppose at least that  $\nu$  is uniformly positive and bounded almost everywhere. We here consider homogeneous Dirichlet boundary

---

Supported by the Austrian Science Fund under the grant W1214, project DK4.

© Springer Nature Switzerland AG 2020

I. Lirkov and S. Margenov (Eds.): LSSC 2019, LNCS 11958, pp. 593–600, 2020.

[https://doi.org/10.1007/978-3-030-41032-2\\_68](https://doi.org/10.1007/978-3-030-41032-2_68)

conditions for the sake of simplicity. In practice, we often meet mixed boundary conditions. Discontinuous coefficients, non-smooth boundaries, changing boundary conditions, non-smooth or incompatible initial conditions, and non-smooth right-hand sides can lead to non-smooth solutions.

In contrast to the conventional time-stepping methods in combination with some spatial discretization method, or the more advanced, but closely related discontinuous Galerkin (dG) methods based on time slices, we here consider space-time finite element discretizations treating time as just another variable, and the term  $\partial_t u$  in (1) as a convection term in time. Following [7], we derive consistent, locally stabilized, conforming finite element schemes on completely unstructured simplicial space-time meshes under the assumption of maximal parabolic regularity; see, e.g., [5]. Unstructured space-time schemes have some advantages with respect to adaptivity, parallelization, and the numerical treatment of moving interfaces or special domains. However, the combination of adaptivity and parallelization is still a challenge. We refer the reader to the survey paper [9] that provides an excellent overview of completely unstructured space-time methods and simultaneous space-time adaptivity. In particular, we would like to mention the papers [8] that is based on an inf-sup-condition, [4] that uses mesh-grading in time, and [1] that also uses stabilization techniques. All three papers treat the autonomous case.

We here present new a priori estimates for low-regularity solutions. In order to avoid reduced convergence rates appearing in the case of uniform mesh refinement, we also consider adaptive refinement procedures in the numerical experiments presented in Sect. 5. The adaptive refinement procedures are based on residual a posteriori error indicators. The huge system of space-time finite element equations is then solved by means of the Generalized Minimal Residual Method (GMRES) preconditioned by an algebraic multigrid cycle. In particular, in the 4D space-time case that is 3D in space, simultaneous space-time adaptivity and parallelization can considerably reduce the computational time. The space-time finite element solver was implemented in the framework of MFEM. The numerical results nicely confirm our theoretical findings. The parallel version of the code shows an excellent parallel performance.

## 2 Weak Formulation and Maximal Parabolic Regularity

The weak formulation of the model problem (1) reads as follows: find  $u \in H_0^{1,0}(Q) := \{u \in L_2(Q) : \nabla_x u \in [L_2(Q)]^d, u = 0 \text{ on } \Sigma\}$  such that (s.t.)

$$\int_Q (-u \partial_t v + \nu \nabla_x u \cdot \nabla_x v) \, d(x, t) = \int_Q f v \, d(x, t) + \int_\Omega u_0 v|_{t=0} \, dx \quad (2)$$

for all  $v \in \hat{H}_0^1(Q) = \{v \in H^1(Q) : v = 0 \text{ on } \Sigma \cup \Sigma_T\}$ , where  $\Sigma_T := \Omega \times \{T\}$ . The existence and uniqueness of weak solutions is well understood; see, e.g., [6]. It was already shown in [6] that  $\partial_t u \in L^2(Q)$  and  $\Delta_x u \in L^2(Q)$  provided that  $\nu = 1$ ,  $f \in L^2(Q)$ , and  $u_0 = 0$ . This case is called maximal parabolic regularity. Similar results can be obtained under more general assumptions imposed on the data; see, e.g., [5] for some very recent results on the non-autonomous case.

### 3 Locally Stabilized Space-Time Finite Element Method

In order to derive the space-time finite element scheme, we need an admissible, shape regular decomposition  $\mathcal{T}_h = \{K\}$  of the space-time cylinder  $\overline{Q} = \bigcup_{K \in \mathcal{T}_h} \overline{K}$  into finite elements  $K$ . On  $\mathcal{T}_h$ , we define a  $H^1$  - conforming finite element space  $V_h$  by means of polynomial simplicial finite elements of the degree  $p$  in the usual way; see, e.g., [2]. Let us assume that the solution  $u$  of (2) belongs to the space  $V_0 = H_{0,0}^{\mathcal{L},1}(\mathcal{T}_h) := \{u \in L_2(Q) : \partial_t u \in L_2(K), \mathcal{L}u := \operatorname{div}_x(\nu \nabla_x u) \in L_2(K) \ \forall K \in \mathcal{T}_h, \text{ and } u|_{\Sigma \cup \Sigma_0} = 0\}$ , i.e., we only need some local version of maximal parabolic regularity, and, for simplicity, we assume homogeneous initial conditions, i.e.,  $u_0 = 0$ . Multiplying the PDE (1) on  $K$  by a local time-upwind test function  $v_h + \theta_K h_K \partial_t v_h$ , with  $v_h \in V_{0h} = \{v_h \in V_h : v_h = 0 \text{ on } \Sigma \cup \Sigma_0\}$ ,  $h_K = \operatorname{diam}(K)$ , and a parameter  $\theta_K > 0$  which we will specify later, integrating over  $K$ , integrating by parts, and summing up over all elements  $K \in \mathcal{T}_h$ , we arrive at the following consistent space-time finite element scheme: find  $u_h \in V_{0h}$  s.t.

$$a_h(u_h, v_h) = l_h(v_h), \quad \forall v_h \in V_{0h}, \tag{3}$$

with  $l_h(v_h) = \sum_{K \in \mathcal{T}_h} \int_K f v_h + \theta_K h_K f \partial_t v_h d(x, t)$  and the bilinear form

$$a_h(u_h, v_h) = \sum_{K \in \mathcal{T}_h} \int_K [\partial_t u_h v_h + \theta_K h_K \partial_t u_h \partial_t v_h + \nu \nabla_x u_h \cdot \nabla_x v_h - \theta_K h_K \operatorname{div}_x(\nu \nabla_x u_h) \partial_t v_h] d(x, t).$$

The bilinear form  $a_h(\cdot, \cdot)$  is coercive on  $V_{0h} \times V_{0h}$  w.r.t. to the norm

$$\|v\|_h^2 = \frac{1}{2} \|v\|_{L_2(\Sigma_T)}^2 + \sum_{K \in \mathcal{T}_h} \left[ \theta_K h_K \|\partial_t v\|_{L_2(K)}^2 + \|\nabla_x v\|_{L_2^\nu(K)}^2 \right], \tag{4}$$

i.e.,  $a_h(v_h, v_h) \geq \mu_c \|v_h\|_h^2$ ,  $\forall v_h \in V_{0h}$ , and bounded on  $V_{0h,*} \times V_{0h}$  w.r.t. to the norm

$$\|v\|_{h,*}^2 = \|v\|_h^2 + \sum_{K \in \mathcal{T}_h} \left[ (\theta_K h_K)^{-1} \|v\|_{L_2(K)}^2 + \theta_K h_K \|\operatorname{div}_x(\nu \nabla_x v)\|_{L_2(K)}^2 \right], \tag{5}$$

i.e.,  $a_h(u_h, v_h) \leq \mu_b \|u_h\|_{h,*} \|v_h\|_h$ ,  $\forall u_h \in V_{0h,*}, \forall v_h \in V_{0h}$ , where  $V_{0h,*} := H_{0,0}^{\mathcal{L},1}(\mathcal{T}_h) + V_{0h}$ ; see [7, Lemma 3.8] and [7, Remark 3.13], respectively. The coercivity constant  $\mu_c$  is robust in  $h_K$  provided that we choose  $\theta_K = \mathcal{O}(h_K)$ ; see Sect. 5 or [7, Lemma 3.8] for the explicit choice. From the above derivation of the scheme, we get consistency  $a_h(u, v_h) = l_h(v_h)$ ,  $\forall v_h \in V_{0h}$ , provided that the solution  $u$  belongs to  $H_{0,0}^{\mathcal{L},1}(\mathcal{T}_h)$  that is ensured in the case of maximal parabolic regularity. The space-time finite element scheme (3) and the consistency relation immediately yield Galerkin orthogonality

$$a_h(u - u_h, v_h) = 0, \quad \forall v_h \in V_{0h}. \tag{6}$$

We deduce that (3) is nothing but a huge linear system of algebraic equations. Indeed, let  $V_{0h} = \operatorname{span}\{p^{(j)}, j = 1, \dots, N_h\}$ , where  $\{p^{(j)}, j = 1, \dots, N_h\}$  is the

nodal finite element basis and  $N_h$  is the total number of space-time degrees of freedom (dofs). Then we can express each function in  $V_{0h}$  in terms of this basis, i.e., we can identify each finite element function  $v_h \in V_{0h}$  with its coefficient vector  $\mathbf{v}_h \in \mathbb{R}^{N_h}$ . Moreover, each basis function  $p^{(j)}$  is also a valid test function. Hence, we obtain  $N_h$  equations from (3), which we rewrite as a system of linear algebraic equations, i.e.  $\mathbf{K}_h \mathbf{u}_h = \mathbf{f}_h$ , with the solution vector  $\mathbf{u}_h = (u_j)_{j=1, \dots, N_h}$ , the vector  $\mathbf{f}_h = (l_h(p^{(i)}))_{i=1, \dots, N_h}$ , and system matrix  $\mathbf{K}_h = (a_h(p^{(j)}, p^{(i)}))_{i,j=1, \dots, N_h}$  that is non-symmetric, but positive definite.

### 4 A Priori Discretization Error Estimates

Galerkin orthogonality (6), together with coercivity and boundedness, enables us to prove a Céa-like estimate, where we bound the discretization error in the  $\|\cdot\|_h$ -norm by the best-approximation error in the  $\|\cdot\|_{h,*}$ -norm.

**Lemma 1.** *Let the bilinear form  $a_h(\cdot, \cdot)$  be coercive [7, Lemma 3.8] with constant  $\mu_c$  and bounded [7, Lemma 3.11, Remark 3.13] with constant  $\mu_b$ , and let  $u \in H_{0,\underline{0}}^{\mathcal{L},1}(\mathcal{T}_h)$  be the solution of the space-time variational problem (2). Then*

$$\|u - u_h\|_h \leq \left(1 + \frac{\mu_b}{\mu_c}\right) \inf_{v_h \in V_{0h}} \|u - v_h\|_{h,*}, \tag{7}$$

where  $u_h \in V_{0h}$  is the solution to the space-time finite element scheme (3).

*Proof.* Estimate (7) easily follows from triangle inequality and Galerkin-orthogonality; see [7, Lemma 3.15, Remark 3.16] for details.

Next, we estimate the best approximation error by the interpolation error, where we have to choose a proper interpolation operator  $\mathcal{I}_*$ . For smooth solutions, i.e.,  $u \in H^l(Q)$  with  $l > (d + 1)/2$ , we obtained a localized a priori error estimate, see [7, Theorem 3.17], where we used the standard Lagrange interpolation operator  $\mathcal{I}_h$ ; see e.g. [2]. In this paper, we are interested in non-smooth solutions, which means that we only require  $u \in H^l(Q)$ , with  $(d + 1)/2 \geq l > 1$ . Hence, we cannot use the Lagrange interpolator. We can, however, use so-called quasi-interpolators, e.g. Clément [3] or Scott-Zhang [2]. For this kind of operators, we need a neighborhood  $S_K$  of an element  $K \in \mathcal{T}_h$  which is defined as  $S_K := \{K' \in \mathcal{T}_h : \bar{K} \cap \bar{K}' \neq \emptyset\}$ . Let  $v \in H^l(Q)$ , with some real  $l > 1$ . Then, for the Scott-Zhang quasi-interpolation operator  $\mathcal{I}_{SZ} : L_2(Q) \rightarrow V_{0h}$ , we have the local estimate (see e.g. [2, (4.8.10)])

$$\|v - \mathcal{I}_{SZ}v\|_{H^k(K)} \leq C_{\mathcal{I}_{SZ}} h_K^{l-k} |v|_{H^l(S_K)}, \quad k = 0, 1. \tag{8}$$

For details on how to construct such a quasi-interpolator, we refer to [2] and the references therein. For simplicity, we now assume that  $\nu$  is piecewise constant, i.e.,  $\nu|_K = \nu_K$ , for all  $K \in \mathcal{T}_h$ . Then we can show the following lemma.

**Lemma 2.** *Let  $v \in V_0 \cap H^l(Q)$  with some  $l > 1$ , and let  $p \geq 1$  denote the polynomial degree of the finite element shape functions on the reference element. Then the following interpolation error estimates are valid:*

$$\|v - \mathfrak{I}_{SZ}v\|_{L_2(\Sigma_T)} \leq c_1 \left( \sum_{\substack{K \in \mathcal{T}_h \\ \partial K \cap \Sigma_T \neq \emptyset}} h_K^{2s-1} |v|_{H^s(S_K)}^2 \right)^{1/2}, \tag{9}$$

$$\|v - \mathfrak{I}_{SZ}v\|_h \leq c_2 \left( \sum_{K \in \mathcal{T}_h} h_K^{2(s-1)} |v|_{H^s(S_K)}^2 \right)^{1/2}, \tag{10}$$

$$\|v - \mathfrak{I}_{SZ}v\|_{h,*} \leq c_3 \left( \sum_{K \in \mathcal{T}_h} h_K^{2(s-1)} (|v|_{H^s(S_K)}^2 + \|\operatorname{div}_x(\nu \nabla_x v)\|_{L_2(K)}^2) \right)^{1/2}, \tag{11}$$

with  $s = \min\{l, p + 1\}$  and positive constants  $c_1, c_2$  and  $c_3$ , that do not depend on  $v$  or  $h_K$  provided that  $\theta_K = \mathcal{O}(h_K)$  for all  $K \in \mathcal{T}_h$ .

*Proof.* Estimates (9) and (10) are easy to proof by using the scaled trace inequality and quasi-interpolation estimates (8). Estimate (11) contains the term  $\theta_K h_K \|\operatorname{div}_x(\nu \nabla_x (v - \mathfrak{I}_{SZ}v))\|_{L_2(K)}^2$  that needs a special treatment. The case  $p = 1$  is straightforward since  $\operatorname{div}_x \mathfrak{I}_{SZ}(\nu \nabla_x v) = 0$ . Otherwise, adding and subtracting the linear quasi-interpolant  $\mathcal{I}_{SZ}^1 v$  of  $v$ , then using triangle and inverse inequalities, and finally the quasi-interpolation estimate (8) for  $k = 1$ , we arrive at the desired estimate for  $p > 1$ .

Now we are in the position to prove our main theorem.

**Theorem 1.** *Let  $l > 1$  and  $p \geq 1$ . Let  $u \in V_0 \cap H^l(Q)$  be the exact solution, and  $u_h \in V_{0h}$  be the approximate solution of the finite element scheme (3). Furthermore, let the assumptions of Lemma 1 (Céa-like estimate) and 2 (quasi-interpolation estimates) hold. Then the a priori error estimate*

$$\|u - u_h\|_h \leq C \left( \sum_{K \in \mathcal{T}_h} h_K^{2(s-1)} (|u|_{H^s(S_K)}^2 + \|\operatorname{div}_x(\nu \nabla_x u)\|_{L_2(K)}^2) \right)^{1/2} \tag{12}$$

holds, with  $s = \min\{l, p + 1\}$  and a positive generic constant  $C$ .

*Proof.* Choosing  $v_h = \mathfrak{I}_{SZ}u$  as some approximation in (7), we can apply the quasi-interpolation estimate (11) to obtain

$$\begin{aligned} \|u - u_h\|_h &\leq \left(1 + \frac{\mu_b}{\mu_c}\right) \|u - \mathfrak{I}_{SZ}u\|_{h,*} \\ &\leq c_3 \left(1 + \frac{\mu_b}{\mu_c}\right) \left( \sum_{K \in \mathcal{T}_h} h_K^{2(s-1)} (|u|_{H^s(S_K)}^2 + \|\operatorname{div}_x(\nu \nabla_x u)\|_{L_2(K)}^2) \right)^{1/2}. \end{aligned}$$

We set  $C = c_3(1 + \mu_b/\mu_c)$  to obtain (12), which closes the proof.

We again mention that, for  $l > (d + 1)/2$ , we can use the Lagrange interpolation that leads to a completely local estimate [7].

## 5 Numerical Results

We implemented the space-time finite element scheme (3) in C++, where we used the finite element library MFEM<sup>1</sup> and the solver library *hypr*<sup>2</sup>. The linear system was solved by means of the GMRES method, preconditioned by the algebraic multigrid *BoomerAMG*. We stopped the iterative procedure if the initial residual was reduced by a factor of  $10^{-8}$ . Both libraries are already fully parallelized with the Message Passing Interface (MPI). Therefore, we performed all numerical tests on the distributed memory cluster RADON1<sup>3</sup> in Linz. For each element  $K \in \mathcal{T}_h$ , we choose  $\theta_K = h_K/(\tilde{c}^2 \bar{\nu}_K)$ , where  $\tilde{c}$  is computed by solving a local generalized eigenvalue problem which comes from an inverse inequality [7].

### 5.1 Example: Highly Oscillatory Solution

We first consider the unit hypercube  $Q = (0, 1)^{d+1}$ , with  $d = 3$ , as the space-time cylinder, and set  $\nu \equiv 1$ . The manufactured function

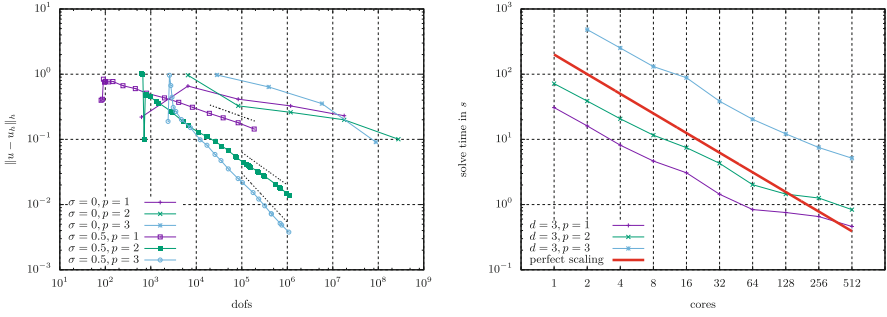
$$u(x, t) = \sin(1/(1/(10\pi) + (x_1^2 + x_2^2 + x_3^2 + t^2)^{0.5}))$$

serves as the exact solution, where we compute the right-hand side  $f$  accordingly. This solution is highly oscillatory. Hence, we do not expect optimal rates for uniform refinement in the pre-asymptotic range. However, using adaptive refinement, we may be able to recover the optimal rates. We use the residual based error indicator proposed by Steinbach and Yang in [9]. For each element  $K \in \mathcal{T}_h$ , we compute  $\eta_K^2 := h_K^2 \|R_h(u_h)\|_{L_2(K)}^2 + h_K \|J_h(u_h)\|_{L_2(\partial K)}^2$ , where  $u_h$  is the solution of (3),  $R_h(u_h) := f + \operatorname{div}_x(\nu \nabla_x u_h) - \partial_t u_h$  in  $K$  and  $J_h(u_h) := [\nu \nabla_x u_h]_e$  on  $e \subset \partial K$ , with  $[\cdot]_e$  denoting the jump across one face  $e \subset \partial K$ . Then we mark each element where the condition  $\eta_K \geq \sigma \max_{K \in \mathcal{T}_h} \eta_K$  is fulfilled, with an a priori chosen threshold  $\sigma$ , e.g.,  $\sigma = 0.5$ . Note that  $\sigma = 0$  results in uniform refinement. In Fig. 1, we observe indeed reduced convergence rates for all polynomial degrees tested. However, using an adaptive procedure, we are able to recover the optimal rates. Moreover, we significantly reduce the number of dofs required to reach a certain error threshold. For instance, in the case  $d = 3$  and  $p = 2$ , we need 276 922 881 dofs to get an error in the  $\|\cdot\|_h$ -norm of  $\sim 10^{-1}$ , whereas we only need 26 359 dofs with adaptive refinement. In terms of runtime, the uniform refinement needed 478.57 s for assembling and solving, while the complete adaptive procedure took 156.5 s only. The parallel performance of the code is also shown in Fig. 1, where we obtain a nice strong scaling up to 64 cores. Then the local problems are too small (only  $\sim 10\,000$  dofs for 128 cores) and the communication overhead becomes too large. Numerical results for  $d = 2$  can be found in [7].

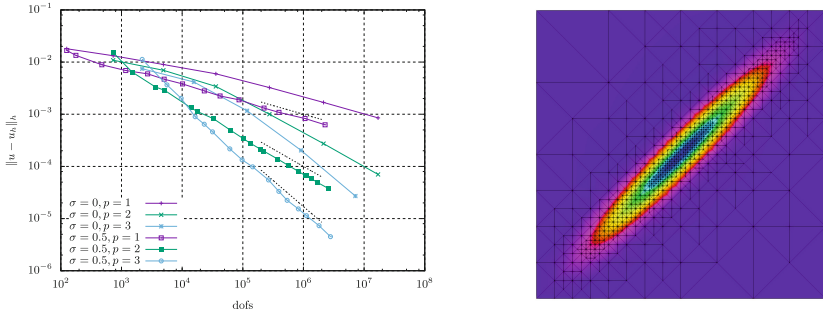
<sup>1</sup> <http://mfem.org/>.

<sup>2</sup> <https://www.llnl.gov/casc/hypr/>.

<sup>3</sup> <https://www.ricam.oew.ac.at/hpc/>.



**Fig. 1.** Example 5.1: Error rates in the  $\|\cdot\|_h$ -norm (left), the dotted lines indicate the optimal rate; Strong scaling on a mesh with 1 185 921 dofs for  $p = 1, 2$  and 5 764 801 dofs for  $p = 3$  (right).



**Fig. 2.** Example 5.2: Error rates in the  $\|\cdot\|_h$ -norm for  $d = 2$ , the dotted lines indicate the optimal rates (left); Diagonal cut through the space-time mesh  $\mathcal{T}_h$  along the line from  $(0, 0, 0)$  to  $(1, 1, 1)$  after 8 adaptive refinements (right).

**5.2 Example: Moving Peak**

For the second example, we consider the unit-cube  $Q = (0, 1)^3$ , i.e.  $d = 2$ . As diffusion coefficient, the choice  $\nu \equiv 1$  is made. We choose the function

$$u(x, t) = (x_1^2 - x_1)(x_2^2 - x_2)e^{-100((x_1-t)^2+(x_2-t)^2)},$$

as exact solution and compute all data accordingly. This function is smooth, and almost zero everywhere, except in a small region around the line from the origin  $(0, 0, 0)$  to  $(1, 1, 1)$ . This motivates again the use of an adaptive method. We use the residual based indicator  $\eta_K$  introduced in Example 5.1. In Fig. 2, we can observe that we indeed obtain optimal rates for both uniform and adaptive refinement. However, using the a posteriori error indicator, we reduce the number of dofs needed to reach a certain threshold by one order of magnitude. For instance, in the case  $p = 2$ , we need 16 974 593 dofs to obtain an error of  $\sim 7 \times 10^{-5}$  with uniform refinement. Using adaptive refinement, we need 1 066 777 dofs only.

## 6 Conclusions

Following [7], we introduced a space-time finite element solver for non-autonomous parabolic evolution problems on completely unstructured simplicial meshes. We only assumed that we have so-called maximal parabolic regularity, i.e., the PDE is well posed in  $L_2$ . We note that this property is only required locally in order to derive a consistent space-time finite element scheme. We extended the a priori error estimate in the mesh-dependent energy norm to the case of non-smooth solutions, i.e.  $u \in H^{1+\epsilon}(Q)$ , with  $0 < \epsilon \leq 1$ .

We performed two numerical examples with known solutions. The first example had a highly oscillatory solution, and the second one was almost zero everywhere except along a line through the space-time cylinder. Using a high-performance cluster, we solved both problems on a sequence of uniformly refined meshes, where we also obtained good strong scaling rates. In order to reduce the computational cost, we also applied an adaptive procedure, using a residual based error indicator. Moreover, we could observe that the AMG preconditioned GMRES method solves the problem quite efficiently.

## References

1. Bank, R.E., Vassilevski, P., Zikatanov, L.: Arbitrary dimension convection-diffusion scheme for space-time discretizations. *J. Comput. Appl. Math.* **310**, 19–31 (2017)
2. Brenner, S.C., Scott, L.R.: *The Mathematical Theory of Finite Element Methods*. Texts in Applied Mathematics, vol. 15, 3rd edn. Springer, New York (2008). <https://doi.org/10.1007/978-0-387-75934-0>
3. Clément, P.: Approximation by finite element functions using local regularization. *Rev. Française Automat. Informat. Recherche Opérationnelle Sér* **9**(R-2), 77–84 (1975)
4. Devaud, D., Schwab, C.: Space-time hp-approximation of parabolic equations. *Calcolo* **55**(3), 35 (2018)
5. Fackler, S.: Non-autonomous maximal regularity for forms given by elliptic operators of bounded variation. *J. Differ. Equ.* **263**, 3533–3549 (2017)
6. Ladyzhenskaya, O.A.: *The Boundary Value Problems of Mathematical Physics*. Applied Mathematical Sciences, vol. 49. Springer-Verlag, New York (1985). <https://doi.org/10.1007/978-1-4757-4317-3>
7. Langer, U., Neumüller, M., Schafelner, A.: Space-time finite element methods for parabolic evolution problems with variable coefficients. In: Apel, T., Langer, U., Meyer, A., Steinbach, O. (eds.) *FEM 2017*. LNCSE, vol. 128, pp. 247–275. Springer, Cham (2019). [https://doi.org/10.1007/978-3-030-14244-5\\_13](https://doi.org/10.1007/978-3-030-14244-5_13)
8. Steinbach, O.: Space-time finite element methods for parabolic problems. *Comput. Methods Appl. Math.* **15**(4), 551–566 (2015)
9. Steinbach, O., Yang, H.: Space-time finite element methods for parabolic evolution equations: discretization, a posteriori error estimation, adaptivity and solution. In: Langer, U., Steinbach, O. (eds.) *Space-Time Methods: Application to Partial Differential Equations*, Radon Series on Computational and Applied Mathematics, vol. 25, pp. 207–248. de Gruyter, Berlin (2019)