Nečas Center for Mathematical Modeling

A design of residual error estimates for a high order BDF-DGFEM applied to the compressible flows

V. Dolejší

Preprint no. 2013-0011

http://ncmm.karlin.mff.cuni.cz/

A design of residual error estimates for a high order BDF-DGFEM applied to the compressible flows *

Vít Dolejší

Charles University Prague, Faculty of Mathematics and Physics, Sokolovská 83, 186 75 Praha, Czech Republic, dolejsi@karlin.mff.cuni.cz

Abstract

We deal with the numerical solution of the non-stationary compressible Navier-Stokes equations with the aid of the backward difference formula – discontinuous Galerkin finite element (BDF-DGFE) method. This scheme is sufficiently stable, efficient and accurate with respect to the space as well as time coordinates. The nonlinear algebraic systems arising from the BDF-DGFE discretization are solved by an iterative Newton-like method. The main benefit of this paper are residual error estimates which are able to identify the computational errors following from the space and time discretizations and from the inexact solution of the nonlinear algebraic systems. Thus we propose an efficient algorithm where the algebraic, spatial and temporal errors are balanced. The computational performance of the proposed method is demonstrated by a list of numerical experiments.

Keywords: discontinuous Galerkin finite element method; compressible Navier-Stokes equations; nonlinear algebraic problems; residual error estimates

1 Introduction

Our aim is to develop a numerical scheme for the simulation of steady as well as unsteady viscous compressible flows, which is sufficiently efficient and accurate for a wide range of flow regimes. The *discontinuous Galerkin finite element* (DGFE) methods have become a very popular numerical technique for the solution of the compressible Navier-Stokes equations. DGFE space discretization uses a (higher order) piecewise polynomial discontinuous approximation on arbitrary meshes. DGFE methods were employed in many papers for the discretization of compressible fluid flow problems, let us cite the pioneering works [7, 8, 38, 9] and some recent papers [48, 34, 24, 6, 28, 12, 23] and the references cited therein. Recent progress of the use of the DGFE method for compressible flow simulations can be found in [35].

The space DGFE discretization leads to a system of stiff ordinary differential equations (ODEs) which should be solved by a suitable ODEs solver. Due to the stiffness of the ODEs, the implicit time discretizations are more efficient than the explicit ones. However, implicit schemes lead to the necessity to solve a nonlinear system of algebraic equations at each time level. These nonlinear systems are usually solved by the Newton method, which requires an evaluation of the Jacobi matrix, see [8, 32]. In [17, 22], we developed the *semi-implicit method* which is based on a formal linearization of the inviscid and viscous fluxes. Then we solved only a linear algebraic system at each time level. Although the semi-implicit method yields satisfactory results, the choice of the time step was a little problematic. In some situations, it should be chosen carefully in order to avoid a failure of the computational process.

In order to guarantee an accuracy with respect to the time, a high order time discretization approach should be employed. Among several possible main approaches, we prefer the *backward difference formula* (BDF) methods since they have lower computational costs (in comparison to the implicit Runge-Kutta methods and the time discontinuous Galerkin discretization). The size of the resulting nonlinear algebraic systems does not depend on the order of the BDF method.

*This work was supported by the Grant No. 201/08/0012 of the Czech Science Foundation.

[2] Dedicated to Professor Miloslav Feistauer on the occasion of his 70th birthday.

In this paper, we develop a fully implicit n-step *backward difference formula – discontinuous Galerkin finite element* (BDF-DGFE) method for the solution of the Navier-Stokes equations. The nonlinear algebraic systems are solved by an inexact Newton-time method where the Jacobi matrix is replaced by the "flux-matrix" arising in the semi-implicit scheme in [22]. The resulting scheme is practically unconditionally stable, it has a high order of accuracy with respect to space as well as time coordinates and it has low computational costs.

However, in order to achieve the accuracy as well as the efficiency of the numerical method, an ultimate goal is an *estimate* of the *computational error* (= difference between the exact and the approximate solution). This is a complex problem since the numerical solution is influenced by three types of errors:

- *space* (or spatial) *error* resulting from the space semi-discretization of the Navier-Stokes equations by the DGFE method when the exact solution is approximated piecewise polynomially,
- *time* (or temporal) *error* resulting from the solution of the resulting ODEs system with the aid of the BDF scheme,
- *algebraic error* (including rounding errors) resulting from the inexact solution of the corresponding nonlinear algebraic systems at each time step.

It is clear that in order to ensure the accuracy as well as the efficiency of the numerical method, these errors should be balanced.

There exist several theoretical papers dealing with a posteriori error estimates for model nonlinear time-dependent problems, let us mention [1, 37, 41, 46, 45, 19] and the references therein. For the linear time-dependent problems, the list of citations is much longer, see, e.g., [49] for a review. Concerning the compressible flow problem, a rigorous a posteriori numerical analysis is open. There exist only few papers dealing with error estimates for the compressible flow problems. Let us mention [33] using the so-called goal-oriented a posteriori error estimation for stationary compressible Navier-Stokes equations based on the approach [10], see also [29]. A similar idea was developed in [4] for the space-time discontinuous Galerkin method applied to the Navier-Stokes equations.

The difference between the exact and the approximate solutions is usually estimated by an error estimator reflecting the space as well as the time discretizations. Moreover, this estimator is split into its spatial and temporal parts which reflect the space and time discretization separately (in some sense).

The approach presented in this paper is different. The spatial error is considered as a difference between the approximate (=space-time discrete) solution and the time semi-discrete solution (which is formally exact with respect to the space). Similarly, the temporal error is considered as a difference between the approximate solution and the space semi-discrete solution (which is formally exact with respect to the time). Thereafter, we derive (rather heuristic) residual error estimators which are able to identify the *space*, *time* and *algebraic errors*. These estimates are based on an approximation of the errors in a dual norm similarly as in [18] where we dealt with steady nonlinear convection-diffusion problems. This approach is very fast and simple to implement since neither an additional problem is solved nor a finite element reconstruction is constructed.

The aim of this paper is to develop an algorithm which gives the numerical solution with the *smallest possible computational error* in the *shortest possible computational time* for the *given mesh*, the *given degree of polynomial approximation* in space and the *given order of BDF*. Therefore, based on the mentioned residual error estimators, we define an algorithm which gives a numerical solution where the time and algebraic errors do not essentially contribute to the total computational error, the time partition is not too fine and the nonlinear algebraic systems are not over-solved.

The content of the rest of the paper is the following. In Sections 2 and 3, we recall the system of the compressible Navier-Stokes equations and its discretization by the BDF-DGFE method from [17], respectively. In Section 4, we introduce the efficient solution strategy of the arising nonlinear algebraic systems. Moreover, Section 5 contains the derivation of the space, time and algebraic residual error estimates and the definition of the final algorithm. Section 6 contains a set of numerical experiments demonstrating the efficiency, accuracy and robustness of the proposed strategy. Concluding remarks are given in Section 7.

2 Compressible flow problem

Let $\Omega \subset \mathbb{R}^2$ be a bounded domain with a piecewise polynomial Lipschitz boundary and $T > 0$. We set $Q_T = \Omega \times (0, T)$ and by $\partial\Omega$ denote the boundary of Ω which consists of several disjoint parts. We distinguish inlet $\partial\Omega_i$, outlet $\partial\Omega_o$ and impermeable walls $\partial\Omega_W$, i.e. $\partial\Omega = \partial\Omega_i \cup \partial\Omega_o \cup \partial\Omega_W$. The system

of the Navier-Stokes equations describing the motion of a non-stationary viscous compressible flow can be written in the dimensionless form

$$\frac{\partial \boldsymbol{w}}{\partial t} + \sum_{i=1}^{2} \frac{\partial \boldsymbol{f}_i(\boldsymbol{w})}{\partial x_i} = \sum_{i=1}^{2} \frac{\partial}{\partial x_i} \left(\sum_{j=1}^{2} \boldsymbol{K}_{i,j}(\boldsymbol{w}) \frac{\partial \boldsymbol{w}}{\partial x_j} \right) \quad \text{in } Q_T, \tag{1}$$

where $\boldsymbol{w} = \boldsymbol{w}(x,t) : Q_T \to \mathbb{R}^4$, $\boldsymbol{w} = (\rho, \rho v_1, \rho v_2, e)^{\mathrm{T}}$ is the unknown state vector (ρ is the density, $\boldsymbol{v} = (v_1, v_2)$ is the vector of velocity and e is the energy), $\boldsymbol{f}_i : \mathbb{R}^4 \to \mathbb{R}^4$, $i = 1,2$, represents the inviscid fluxes and $\boldsymbol{K}_{i,j} : \mathbb{R}^4 \to \mathbb{R}^{4\times 4}$, $i,j = 1,2$ are matrices defining the viscous terms. The forms of vectors $\boldsymbol{f}_i$, $i = 1,2$, and matrices $\boldsymbol{K}_{i,j}$, $i = 1,2$, can be found, e.g., in [17] or [25, Section 4.3]. We consider the Newtonian type of fluid accompanied by the state equation of a perfect gas and the definition of total energy.

The system (1) is of *hyperbolic-parabolic* type and it is equipped with the initial condition $\boldsymbol{w}(x,0) = \boldsymbol{w}^0(x)$, $x \in \Omega$ and suitable boundary conditions. On the inlet and outlet, we prescribe Dirichlet boundary conditions for some of the flow variables, while Neumann conditions are used for the remaining variables, see e.g., in [17, 23, 25, 26]. On the impermeable walls, we set

$$\boldsymbol{v} = 0, \qquad \partial\theta/\partial\boldsymbol{n} = 0 \quad \text{ on } \partial\Omega_W, \tag{2}$$

where $\boldsymbol{v}$ is the velocity vector and $\partial\theta/\partial\boldsymbol{n}$ denotes the normal derivative of the temperature, for details, see [16, 17].

Let us mention that the Euler fluxes $\boldsymbol{f}_i$, $i = 1,2$, satisfy (see [25, Lemma 3.1]) $\boldsymbol{f}_i(\boldsymbol{w}) = \boldsymbol{A}_i(\boldsymbol{w})\boldsymbol{w}$, $i = 1,2$, where $\boldsymbol{A}_i(\boldsymbol{w}) = \frac{D\boldsymbol{f}_i(\boldsymbol{w})}{D\boldsymbol{w}}$, $i = 1,2$, are the Jacobi matrices of $\boldsymbol{f}_i$. Moreover, we define the matrix

$$\boldsymbol{P}(\boldsymbol{w}, \boldsymbol{n}) = \sum_{i=1}^{2} \boldsymbol{A}_i(\boldsymbol{w}) n_i, \tag{3}$$

where $\boldsymbol{n} = (n_1, n_2) \in \mathbb{R}^2$, $|n|^2 = n_1^2 + n_2^2 = 1$, which plays a role in the definition of the numerical flux. Finally, if $\boldsymbol{w}$ is the state vector satisfying the wall boundary condition (2), then

$$\sum_{j=1}^{2} \boldsymbol{K}_{i,j}(\boldsymbol{w}) \frac{\partial \boldsymbol{w}}{\partial x_j}\bigg|_{\partial\Omega_W} = (0, \boldsymbol{t}_{i,1}, \boldsymbol{t}_{i,2}, 0)^{\mathrm{T}} =: \sum_{j=1}^{2} \boldsymbol{K}^W_{i,j}(\boldsymbol{w}) \frac{\partial \boldsymbol{w}}{\partial x_j}\bigg|_{\partial\Omega_W}, \quad i = 1,2, \tag{4}$$

where $\boldsymbol{t}_{i,j}$ are the components of the stress tensor and $\boldsymbol{K}^W_{i,j}$, $i,j = 1,2$ are matrices which have the first and the last rows equal to zeros and the other rows are identical with the rows of $\boldsymbol{K}_{i,j}$, $i,j = 1,2$.

In the case of inviscid flow (i.e., $\boldsymbol{K}_{i,j} = 0$, $i,j = 1,2$), we obtain the reduced problem of the Euler equations. Thus the boundary conditions should be replaced by the appropriate "inviscid conditions" which are chosen in such a way that the system of the Euler equations is linearly well-posed. The boundary condition (2) is replaced by $\boldsymbol{v} \cdot \boldsymbol{n} = 0$ where $\boldsymbol{n}$ is a unit outer derivative to $\partial\Omega$. For more details see, e.g., [25, Section 3.3.6].

3 BDF-DGFE method for the Navier-Stokes equations

In this section, we recall the *backward difference formula – discontinuous Galerkin finite element* (BDF-DGFE) method for the solution of the Navier-Stokes equations (1) presented in [17]. However, in contrast to [17], we employ the treatment of the boundary conditions from [23] which has better convergence properties. Moreover, in contrast to [17, 23], we do not employ the semi-implicit time discretization but the fully implicit one.

We use the standard notation for function spaces with usual norms and semi-norms (see, e. g., [36], [40]): $L^2(M)$ denotes the Lebesgue space of square integrable functions over a set M, $H^k(M)$, $k = 0,1,\ldots$ are the Sobolev spaces of functions with square integrable weak derivatives of order k over M. The bolted symbols $\boldsymbol{H}^k(M)$, $k = 0,1,\ldots$ denote Sobolev spaces of vector-values functions from M to $\mathbb{R}^4$. By $(\cdot,\cdot)_{0,M}$ we denote the L^2-scalar product over M.

Furthermore, $L^2(I;X)$ ($H^1(I;X)$) is the Bochner space of functions square integrable (square integrable first time derivative) over an interval $I \subset \mathbb{R}$ with values in a Banach space X and $C(I;X)$ ($C^1(I;X)$) is the space of continuous (continuously differentiable) mappings of an interval $I \subset \mathbb{R}$ into X.

3.1 Triangulations

Let $\mathscr{T}_h$ ($h > 0$) be a partition of the closure $\overline{\Omega}$ of the domain Ω into a finite number of closed two-dimensional elements K with mutually disjoint interiors. Moreover, let $F_K : \hat{K} \to \mathbb{R}^2$ be a polynomial mapping such that $F_K(\hat{K}) = K$ where $\hat{K} = \{(\hat{x}_1, \hat{x}_2);\ \hat{x}_i \geq 0,\ i = 1, 2,\ \hat{x}_1 + \hat{x}_2 \leq 1\}$ is the reference triangle. If $K \cap \partial\Omega = \emptyset$ or $K \cap \partial\Omega$ is a straight line then F_K is an affine mapping and K is a triangle. Otherwise, F_K is a polynomial mapping giving an approximation of the segment $K \cap \partial\Omega$ and hence K is a curved triangle. For simplicity, we do not distinguish between Ω and $\Omega_h := \cup_{K\in\mathscr{T}_h} K$ and simply write Ω. We call $\mathscr{T}_h = \{K\}_{K\in\mathscr{T}_h}$ a *triangulation* of Ω.

By $\mathscr{F}_h$ we denote the set of all open (one-dimensional) edges of all elements $K \in \mathscr{T}_h$. Further, the symbol $\mathscr{F}_h^I$ stands for the set of all $\Gamma \in \mathscr{F}_h$ that are contained in Ω (inner faces). Moreover, we define $\mathscr{F}_h^W$, $\mathscr{F}_h^i$ and $\mathscr{F}_h^o$ as the sets of all $\Gamma \in \mathscr{F}_h$ such that $\Gamma \subset \partial\Omega_W$, $\Gamma \subset \partial\Omega_i$ and $\Gamma \subset \partial\Omega_o$, respectively. In order to simplify the notation, we set $\mathscr{F}_h^{io} = \mathscr{F}_h^i \cup \mathscr{F}_h^o$ and $\mathscr{F}_h^B = \mathscr{F}_h^W \cup \mathscr{F}_h^i \cup \mathscr{F}_h^o$. Finally, for each $\Gamma \in \mathscr{F}_h$ we define a unit normal vector $\boldsymbol{n}_\Gamma$. We assume that for $\Gamma \in \mathscr{F}_h^B$ the vector $\boldsymbol{n}_\Gamma$ has the same orientation as the outer normal of $\partial\Omega$. For each $\Gamma \in \mathscr{F}_h^I$, the orientation of $\boldsymbol{n}_\Gamma$ is arbitrary but fixed.

3.2 Discontinuous finite element spaces

Let $\mathscr{T}_h$ be the triangulation then we define the so-called *broken Sobolev spaces*

$$H^2(\mathscr{T}_h) := \{v : \Omega \to \mathbb{R};\ v|_K \in H^2(K)\ \forall\, K \in \mathscr{T}_h\} \quad \text{and} \quad \boldsymbol{H}^2(\mathscr{T}_h) := [H^2(\mathscr{T}_h)]^4 \tag{5}$$

of scalar and vector-valued functions, respectively, with the standard norm and seminorm satisfying

$$\begin{aligned} \|v\|^2_{H^2(\mathscr{T}_h)} &= \sum_{K\in\mathscr{T}_h} \|v\|^2_{H^2(K)}, \quad |v|^2_{H^2(\mathscr{T}_h)} = \sum_{K\in\mathscr{T}_h} |v|^2_{H^2(K)}, \quad v \in H^2(\mathscr{T}_h), \\ \|\boldsymbol{v}\|^2_{H^2(\mathscr{T}_h)} &= \sum_{K\in\mathscr{T}_h} \|\boldsymbol{v}\|^2_{\boldsymbol{H}^2(K)}, \quad |\boldsymbol{v}|^2_{H^2(\mathscr{T}_h)} = \sum_{K\in\mathscr{T}_h} |\boldsymbol{v}|^2_{\boldsymbol{H}^2(K)}, \quad \boldsymbol{v} \in \boldsymbol{H}^2(\mathscr{T}_h). \end{aligned} \tag{6}$$

The discontinuous Galerkin finite element (DGFE) solution of (1) is sought in a finite dimensional subspace of $\boldsymbol{H}^2(\mathscr{T}_h)$ which consists of piecewise polynomial functions. The DGFE method allows the use of different polynomial degrees over elements but for simplicity we consider here the fixed degree p of polynomial approximation for all $K \in \mathscr{T}_h$. Over the triangulation $\mathscr{T}_h$ we define the finite dimensional space of discontinuous piecewise polynomial functions associated with p by

$$S_h^p = \{v;\ v \in L^2(\Omega),\ v|_K \circ F_K \in P_p(\hat{K})\ \forall K \in \mathscr{T}_h\}, \tag{7}$$

where $P_p(\hat{K})$ denotes the space of all polynomials on $\hat{K}$ of degree $\leq p$.We seek the approximate solution in the space of vector-valued functions $\boldsymbol{S}_h^p = [S_h^p]^4$, whose dimension is $N_h = 2\#\mathscr{T}_h(p+1)(p+2)$, where $\#\mathscr{T}_h$ is the number of elements of $\mathscr{T}_h$. Obviously, $S_h^p \subset H^2(\mathscr{T}_h)$ and $\boldsymbol{S}_h^p \subset \boldsymbol{H}^2(\mathscr{T}_h)$.

For each $\Gamma \in \mathscr{F}_h^I$ there exist two elements $K^{(+)}, K^{(-)} \in \mathscr{T}_h$ such that $\Gamma \subset K^{(+)} \cap K^{(-)}$. We use the convention that $K^{(-)}$ lies in the direction of $\boldsymbol{n}_\Gamma$ and $K^{(+)}$ in the opposite direction of $\boldsymbol{n}_\Gamma$. Then for $v \in H^2(\mathscr{T}_h)$, we introduce the notation: $v|_\Gamma^{(+)}$ is the trace of $v|_{K^{(+)}}$ on Γ, $v|_\Gamma^{(-)}$ is the trace of $v|_{K^{(-)}}$ on Γ, $\langle v\rangle_\Gamma := \left(v|_\Gamma^{(+)} + v|_\Gamma^{(-)}\right)/2$ is the mean value of v on Γ and $[\![v]\!]_\Gamma := v|_\Gamma^{(+)} - v|_\Gamma^{(-)}$ is the jump of v on Γ. Similarly, we define the jumps and the mean values for vector-valued functions. In case that $\boldsymbol{n}_\Gamma$, $[\![\cdot]\!]_\Gamma$ and $\langle\cdot\rangle_\Gamma$ are arguments of $\int_\Gamma \ldots \mathrm{d}S$, $\Gamma \in \mathscr{F}_h$ we omit the subscript Γ and write simply $\boldsymbol{n}$, $[\![\cdot]\!]$ and $\langle\cdot\rangle$, respectively. Finally, for $\Gamma \in \mathscr{F}_h^B$ we denote by $v|_\Gamma^{(+)}$ the trace of $v|_{K^{(+)}}$ on Γ, where $K^{(+)} \in \mathscr{T}_h$ such that $\Gamma \subset K^{(+)} \cap \partial\Omega$.

3.3 Discretization forms

In this section we recall the definitions of forms arising from the interior penalty Galerkin (IPG) discretization of the Navier-Stokes equations (1). Their derivation can be found, e.g., in [16, 17, 32].

3.3.1 Inviscid terms

For $\boldsymbol{w}_h, \boldsymbol{\varphi}_h \in \boldsymbol{H}^2(\mathscr{T}_h)$, we define the form

$$\begin{aligned} \boldsymbol{b}_h(\boldsymbol{w}_h, \boldsymbol{\varphi}_h) &:= \sum_{\Gamma \in \mathscr{F}_h} \int_\Gamma \left(\boldsymbol{P}^+ \left(\langle \boldsymbol{w}_h \rangle, \boldsymbol{n} \right) \boldsymbol{w}_h|_\Gamma^{(+)} + \boldsymbol{P}^- \left(\langle \boldsymbol{w}_h \rangle, \boldsymbol{n} \right) \boldsymbol{w}_h|_\Gamma^{(-)} \right) \cdot [\![\boldsymbol{\varphi}_h]\!] \, \mathrm{d}S \\ &\quad - \sum_{K \in \mathscr{T}_h} \int_K \sum_{i=1}^{2} \boldsymbol{A}_i(\boldsymbol{w}_h) \boldsymbol{w}_h \cdot \frac{\partial \boldsymbol{\varphi}_h}{\partial x_i} \, \mathrm{d}x, \end{aligned} \tag{8}$$

where $\boldsymbol{P}^\pm$ are the positive and negative parts of the matrix $\boldsymbol{P}$ given by (3) which define the Vijayasundaram numerical flux [47] used for the approximation of inviscid fluxes through $\Gamma \in \mathscr{F}_h$. For $\Gamma \in \mathscr{F}_h^B$, we have to specify the meaning of $\boldsymbol{w}_h|_\Gamma^{(-)}$. For $\Gamma \in \mathscr{F}_h^W$, we set

$$\boldsymbol{w}_h|_\Gamma^{(-)} := \mathcal{M} \left(\boldsymbol{w}_h|_\Gamma^{(+)} \right), \tag{9}$$

where $\mathcal{M} : \mathbb{R}^4 \to \mathbb{R}^4$ is the *"mirror operator"* defined on $\Gamma \in \partial\Omega_W$ such that

$$\boldsymbol{w} = (\rho, \boldsymbol{v}, e)^{\mathrm{T}} \;\Rightarrow\; \begin{cases} \mathcal{M}(\boldsymbol{w}) = (\rho, \rho\boldsymbol{v} - 2\rho(\boldsymbol{v} \cdot \boldsymbol{n})\boldsymbol{n}, e)^{\mathrm{T}} & \text{for inviscid flow,} \\ \mathcal{M}(\boldsymbol{w}) = (\rho, -\rho\boldsymbol{v}, e)^{\mathrm{T}} & \text{for viscous flow,} \end{cases} \tag{10}$$

where $\boldsymbol{n}$ is the unit outer normal to $\partial\Omega_W$. Therefore, for inviscid flow, the normal components of the velocities of $\boldsymbol{w}$ and $\mathcal{M}(\boldsymbol{w})$ have the same magnitude and the opposite direction. For viscous flow, the velocities of $\boldsymbol{w}$ and $\mathcal{M}(\boldsymbol{w})$ have the same magnitude and the opposite directions. The remaining components of $\boldsymbol{w}$ and $\mathcal{M}(\boldsymbol{w})$ (density, energy and tangential component of the velocity) are the same.

Finally, for $\Gamma \in \mathscr{F}_h^{io}$, we set

$$\boldsymbol{w}_h|_\Gamma^{(-)} := LRP(\boldsymbol{w}_h|_\Gamma^{(+)}, \boldsymbol{w}_D, \boldsymbol{n}_\Gamma), \; \Gamma \in \mathscr{F}_h^{io}, \tag{11}$$

where $LRP(\cdot,\cdot,\cdot)$ represents the solution of the *local Riemann problem* considered on edge $\Gamma \in \mathscr{F}_h^{io}$ and $\boldsymbol{w}_D$ is a given state vector (e.g. from far-field boundary conditions), see [21]. For more details, we refer to [22] or [23].

3.3.2 Viscous terms

In [17], we presented and studied three variants of the DGFE discretization with interior penalty, namely SIPG, NIPG and IIPG. Based on our computational experiences, we prefer the IIPG (incomplete interior penalty Galerkin variant) namely due to its easier formulation. Then, for $\boldsymbol{w}_h, \boldsymbol{\varphi}_h \in \boldsymbol{H}^2(\mathscr{T}_h)$, we define the form

$$\begin{aligned} \boldsymbol{a}_h(\boldsymbol{w}_h, \boldsymbol{\varphi}_h) &:= \sum_{K \in \mathscr{T}_h} \int_K \sum_{i,j=1}^{2} \left(\boldsymbol{K}_{i,j}(\boldsymbol{w}_h) \frac{\partial \boldsymbol{w}_h}{\partial x_j} \right) \cdot \frac{\partial \boldsymbol{\varphi}_h}{\partial x_i} \, \mathrm{d}x \\ &\quad - \sum_{\Gamma \in \mathscr{F}_h^I} \int_\Gamma \sum_{i=1}^{2} \left\langle \sum_{j=1}^{2} \boldsymbol{K}_{i,j}(\boldsymbol{w}_h) \frac{\partial \boldsymbol{w}_h}{\partial x_j} \right\rangle n_i \cdot [\![\boldsymbol{\varphi}_h]\!] \, \mathrm{d}S \\ &\quad - \sum_{\Gamma \in \mathscr{F}_h^{io}} \int_\Gamma \sum_{i,j=1}^{2} \boldsymbol{K}_{i,j}(\boldsymbol{w}_h|_\Gamma^{(+)}) \frac{\partial \boldsymbol{w}_h}{\partial x_j} \, n_i \cdot \boldsymbol{\varphi}_h \, \mathrm{d}S \\ &\quad - \sum_{\Gamma \in \mathscr{F}_h^W} \int_\Gamma \sum_{i,j=1}^{2} \boldsymbol{K}_{i,j}^W(\boldsymbol{w}_h|_\Gamma^{(+)}) \frac{\partial \boldsymbol{w}_h}{\partial x_j} \, n_i \cdot \boldsymbol{\varphi}_h \, \mathrm{d}S \end{aligned} \tag{12}$$

where $\boldsymbol{K}_{i,j}^W$, $i,j = 1,2$ are defined by (4).

3.3.3 Interior and boundary penalties

For $\boldsymbol{w}_h, \boldsymbol{\varphi}_h \in \boldsymbol{H}^2(\mathscr{T}_h)$, we define the form

$$\begin{aligned} \boldsymbol{J}_h^\sigma(\boldsymbol{w}_h, \boldsymbol{\varphi}_h) &= \sum_{\Gamma \in \mathscr{F}_h^I} \int_\Gamma \sigma [\![\boldsymbol{w}_h]\!] \cdot [\![\boldsymbol{\varphi}_h]\!] \, \mathrm{d}S + \sum_{\Gamma \in \mathscr{F}_h^{io}} \int_\Gamma \sigma (\boldsymbol{w}_h - \mathcal{B}(\boldsymbol{w}_h)) \cdot \boldsymbol{\varphi}_h \, \mathrm{d}S \\ &\quad + \sum_{\Gamma \in \mathscr{F}_h^W} \int_\Gamma \sigma (\boldsymbol{w}_h - \mathcal{B}(\boldsymbol{w}_h)) \cdot \mathcal{V}(\boldsymbol{\varphi}_h) \, \mathrm{d}S, \end{aligned} \tag{13}$$

where $\mathcal{B}(\cdot) : \mathbb{R}^4 \to \mathbb{R}^4$ is the operator of the boundary condition given by

$$\mathcal{B}(\boldsymbol{w}_h) := (\rho|_\Gamma, 0, 0, \rho|_\Gamma c_V \theta|_\Gamma) \quad \text{for} \quad \Gamma \in \mathscr{F}_h^W, \tag{14}$$

$$\mathcal{B}(\boldsymbol{w}_h) := LRP(\boldsymbol{w}_h|_\Gamma^{(+)}, \boldsymbol{w}_D, \boldsymbol{n}_\Gamma) \quad \text{for} \quad \Gamma \in \mathscr{F}_h^{io}, \tag{15}$$

where c_V is the specific heat at constant volume and LRP is given by (11). Moreover, the operator $\mathcal{V} : \mathbb{R}^4 \to \mathbb{R}^4$ is given by $\mathcal{V}(\boldsymbol{\varphi}) = (0, \varphi_2, \varphi_3, 0)$ for $\boldsymbol{\varphi} = (\varphi_1, \varphi_2, \varphi_3, \varphi_4)$. The role of $\mathcal{V}$ is to penalize only the components of $\boldsymbol{w}$, for which the Dirichlet boundary conditions are prescribed on fixed walls. Moreover, the penalty parameter σ is chosen by

$$\sigma|_\Gamma = C_W p^2 / (\mathrm{diam}(\Gamma)\,\mathrm{Re}), \quad \Gamma \in \mathscr{F}_h, \tag{16}$$

where Re is the Reynolds number of the flow, p is the degree of the polynomial approximation and $C_W > 0$ is a suitable constant which guarantees the convergence of the method. In numerical experiments presented here, we set $C_W = 10$. For inviscid flows, we set $\sigma|_\Gamma = 0,\ \Gamma \in \mathscr{F}_h$.

3.4 Space semi-discretization

In order to simplify the notation, for $\boldsymbol{w}_h, \boldsymbol{\varphi}_h \in \boldsymbol{H}^2(\mathscr{T}_h)$, we set

$$\boldsymbol{c}_h(\boldsymbol{w}_h, \boldsymbol{\varphi}_h) := \boldsymbol{a}_h(\boldsymbol{w}_h, \boldsymbol{\varphi}_h) + \boldsymbol{b}_h(\boldsymbol{w}_h, \boldsymbol{\varphi}_h) + \boldsymbol{J}_h^\sigma(\boldsymbol{w}_h, \boldsymbol{\varphi}_h). \tag{17}$$

Due to the consistency of the DGFE method, we find that if $\boldsymbol{w} \in C^1((0,T); \boldsymbol{H}^2(\Omega))$ is the regular solution of the Navier-Stokes equations (1) with the corresponding initial and boundary conditions, then

$$\left(\frac{\partial \boldsymbol{w}(x,t)}{\partial t}, \boldsymbol{\varphi}(x)\right)_{0,\Omega} + \boldsymbol{c}_h(\boldsymbol{w}(x,t),\ \boldsymbol{\varphi}(x)) = 0 \qquad \forall \boldsymbol{\varphi} \in \boldsymbol{H}^2(\mathscr{T}_h)\ \forall t \in (0,T). \tag{18}$$

Now, we introduce the space semi-discretization of (1). Let $C^1([0,T]; \boldsymbol{S}_h^p)$ denote the space of continuously differentiable mappings of the interval $[0,T]$ into $\boldsymbol{S}_h^p$.

Definition 3.1. *A function $\boldsymbol{w}_h \in C^1([0,T]; \boldsymbol{S}_h^p)$ is called the* space semi-discrete solution *of* (1)*, if*

$$\left(\frac{\partial \boldsymbol{w}_h(t)}{\partial t}, \boldsymbol{\varphi}_h\right)_{0,\Omega} + \boldsymbol{c}_h(\boldsymbol{w}_h(t), \boldsymbol{\varphi}_h) = 0 \qquad \forall \boldsymbol{\varphi}_h \in \boldsymbol{S}_h^p\ \forall t \in (0,T), \tag{19a}$$

$$\boldsymbol{w}_h(0) = \boldsymbol{w}_h^0, \tag{19b}$$

where $\boldsymbol{w}_h^0 \in \boldsymbol{S}_h^p$ denotes the L^2-projection of the initial condition $\boldsymbol{w}^0$ in $\boldsymbol{S}_h^p$.

The problem (19) represents a system of ordinary differential equations (ODEs) for $\boldsymbol{w}_h(t)$, which has to be discretized in time by a suitable method. Since this system is stiff, it is advantageous to use an implicit time discretization. In order to obtain a sufficiently stable and accurate approximation with respect to the time coordinate, we use the *backward difference formula* (BDF), see, e.g., [30], for the solution the ODE problem (19).

3.5 Full space-time discretization

In order to finalise the discretization of (1), we introduce a partition of the time interval $(0,T)$ with $0 = t_0 < t_1 < t_2 < \cdots < t_r = T$. The subscript r denotes the number of time steps necessary to achieve final time T. Obviously, r depends on the size of the time steps $\tau_k := t_k - t_{k-1},\ k = 1, \ldots, r$ and since we use an adaptive choice of the time step, its value is not a priori known. In the following we do not denote the explicit dependence of r on the computational parameters. Moreover, we set $I_k := (t_{k-1}, t_k),\ k = 1, \ldots, r$ and $\mathcal{I}_\tau := \{I_k\}_{k=1}^r$. Let $\boldsymbol{w}_h^k \in \boldsymbol{S}_h^p$ denote a piecewise polynomial approximation of $\boldsymbol{w}_h(t_k),\ k = 0, 1, \ldots, r$. We define the following scheme.

Definition 3.2. *The* approximate solution *of* (1) *by the* implicit n-step BDF-DGFE method *is defined as functions $\boldsymbol{w}_h^k \in \boldsymbol{S}_h^p,\ k = 0, \ldots, r$, satisfying the conditions*

$$\frac{1}{\tau_k}\left(\sum_{l=0}^{n} \alpha_{n,l} \boldsymbol{w}_h^{k-l}, \boldsymbol{\varphi}_h\right)_{0,\Omega} + \boldsymbol{c}_h\left(\boldsymbol{w}_h^k, \boldsymbol{\varphi}_h\right) = 0 \quad \forall \boldsymbol{\varphi}_h \in \boldsymbol{S}_h^p,\ k = n, \ldots, r, \tag{20a}$$

$$\boldsymbol{w}_h^0 \ \textit{ is the } L^2\textit{-projection of the initial condition } \boldsymbol{w}^0 \textit{ in } \boldsymbol{S}_h^p, \tag{20b}$$

$$\boldsymbol{w}_h^l \in \boldsymbol{S}_h^p,\ l = 1, \ldots, n-1 \textit{ are given by a suitable "less-step" method,} \tag{20c}$$

	constant time step			variable time step		
	$n=1$	$n=2$	$n=3$	$n=1$	$n=2$	$n=3$
$\alpha_{n,0}$	1	$\frac{3}{2}$	$\frac{11}{6}$	1	$\frac{2\,\theta_k+1}{\theta_k+1}$	$\frac{\theta_k\theta_{k-1}}{\theta_k\theta_{k-1}+\theta_{k-1}+1}+\frac{2\theta_k+1}{\theta_k+1}$
$\alpha_{n,1}$	-1	-2	-3	-1	$-(\theta_k+1)$	$-\frac{(\theta_k+1)(\theta_k\theta_{k-1}+\theta_{k-1}+1)}{\theta_{k-1}+1}$
$\alpha_{n,2}$		$\frac{1}{2}$	$\frac{3}{2}$		$\frac{\theta_k^2}{\theta_k+1}$	$\frac{\theta_k^2(\theta_k\theta_{k-1}+\theta_{k-1}+1)}{\theta_k+1}$
$\alpha_{n,3}$			$-\frac{1}{3}$			$-\frac{(\theta_k+1)\theta_k^2\theta_{k-1}^3}{(\theta_{k-1}+1)(\theta_k\theta_{k-1}+\theta_{k-1}+1)}$

Tab. 1: Values of $\alpha_{n,l},\ l=0,\dots,n$ for $n=2,3$, $\theta_k := \tau_k/\tau_{k-1},\ k=1,2,\dots,r$.

where $n \geq 1$ is the degree of the BDF scheme and the BDF coefficients $\alpha_{n,l},\ l=0,\dots,n$ depend on time steps $\tau_{k-l},\ l=0,\dots,n$.

The n-step BDF-DGFE method (shortly n-BDF-DGFE method) has formally the order of convergence $O(h^p+\tau^n)$ in the $L^2(0,T;H^1(\Omega))$-norm. For $n=1$ and $n=2$ these methods are unconditionally stable, and for increasing n they lose more and more stability, for $n>7$ these methods are definitely unstable, see [30, Section III.5]. In practice, we employ the n-BDF-DGFE scheme for $n=1,2,3$, the values of the corresponding coefficients $\alpha_{n,l},\ l=0,\dots,n,\ n=1,2,3$ are given in Table 1 for a constant and a variable time steps.

The problem (20) represents a nonlinear algebraic system for each $k=1,\dots,r$ which should be solved by a suitable solver, see Section 4.

Remark 3.3. *In practice, we realise relation* (20c) *in such a way, that for $l=1,\dots,n-1$ we employ the l-step BDF-DGFE scheme, i.e., the one step BDF (= backward Euler) at first time step, the two steps BDF at second time step, etc. For simplicity, we formally replace* (20a) *and* (20c) *by*

$$\frac{1}{\tau_k}\left(\sum_{l=0}^{n}\alpha_{n,l}\boldsymbol{w}_h^{k-l},\boldsymbol{\varphi}_h\right)_{0,\Omega}+\boldsymbol{c}_h\left(\boldsymbol{w}_h^k,\boldsymbol{\varphi}_h\right)=0\quad\forall\,\boldsymbol{\varphi}_h\in\boldsymbol{S}_h^p,\ k=1,\dots,r, \tag{21}$$

and we do not emphasise this fact in the following.

4 Solution strategy

In this section, we develop a solution strategy of the nonlinear algebraic systems (20).

4.1 Algebraic representation

Let N_h denote the dimension of the piecewise polynomial space $\boldsymbol{S}_h^p$ and $B_h := \{\boldsymbol{\varphi}_i(x)\}_{i=1}^{N_h}$ denote a set of linearly independent functions forming a basis of $\boldsymbol{S}_h^p$. It is possible to construct a basis B_h as a composition of local bases constructed separately for each $K\in\mathscr{T}_h$ and each component of the vector-values functions. We construct the (almost) L^2-orthonormal basis $\hat{B}$ of polynomials of degree $\leq p$ on the reference triangle $\hat{K}$ by the Gram-Schmidt orthogonalization process applied to the set of monomials of degree $\leq p$ on $\hat{K}$. Finally, with the aid of the mappings $F_K,\ K\in\mathscr{T}_h$, we define the local basis for each $K\in\mathscr{T}_h$. Obviously, if F_K is linear, then the L^2-orthogonality is preserved. Otherwise, the orthogonality is violated but it does not cause any problems in practical applications since the a curved face $K\cap\partial\Omega$ is close to a straight one and hence the corresponding mapping F_K is close to a linear one. More details can be found in [23].

Let $\boldsymbol{w}_h^k\in\boldsymbol{S}_h^p$ be a piecewise polynomial function. It can be expressed as

$$\boldsymbol{w}_h^k(x)=\sum_{j=1}^{N_h}\xi^{k,j}\boldsymbol{\varphi}_j(x)\in\boldsymbol{S}_h^p\quad\longleftrightarrow\quad\boldsymbol{\xi}_k:=\{\xi^{k,j}\}_{j=1}^{N_h}\in\mathbb{R}^{N_h},\ k=1,\dots,r, \tag{22}$$

where $\xi^{k,j} \in \mathbb{R},\ j = 1, \dots, N_h,\ k = 1, \dots, r$ are the basis coefficients. Obviously, (22) defines an isomorphism between $\boldsymbol{w}_h^k \in \boldsymbol{S}_h^p$ and $\boldsymbol{\xi}_k \in \mathbb{R}^{N_h}$.

In order to rewrite the nonlinear algebraic systems (21), we define the vector-valued function $\boldsymbol{F}_h : [\mathbb{R}^{N_h}]^n \times \mathbb{R}^{N_h} \to \mathbb{R}^{N_h}$ by

$$\boldsymbol{F}_h\left(\left\{\boldsymbol{\xi}_{k-l}\right\}_{l=1}^n ; \boldsymbol{\xi}_k\right) := \left\{ \frac{1}{\tau_k} \left(\sum_{l=0}^{n} \alpha_{n,l} \boldsymbol{w}_h^{k-l}, \boldsymbol{\varphi}_i \right) + \boldsymbol{c}_h(\boldsymbol{w}_h^k, \boldsymbol{\varphi}_i) \right\}_{i=1}^{N_h}, \ k = 1, \dots, r, \tag{23}$$

where $\boldsymbol{\xi}_{k-l} \in \mathbb{R}^{N_h}$ is the algebraic representation of $\boldsymbol{w}_h^{k-l} \in \boldsymbol{S}_h^p$ for $l = 1, \dots, n$. We do not emphasise that $\boldsymbol{F}_h$ depends explicitly on τ_k. Therefore, the algebraic representation of the systems (21) reads: for the given vectors $\boldsymbol{\xi}_{k-1}, \boldsymbol{\xi}_{k-2}, \dots, \boldsymbol{\xi}_{k-n} \in \mathbb{R}^{N_h}$

$$\text{find } \boldsymbol{\xi}_k \in \mathbb{R}^{N_h} \text{ such that } \boldsymbol{F}_h(\left\{\boldsymbol{\xi}_{k-l}\right\}_{l=1}^n ; \boldsymbol{\xi}_k) = \boldsymbol{0}, \quad k = 1, \dots, r. \tag{24}$$

The system (24) is strongly nonlinear and we solve it by a Newton-like iterative method where the Jacobi matrix in the Newton method is replaced by the so-called *flux matrix* developed in the context of the semi-implicit DGFE method in [17, 22, 23].

4.2 Flux matrix

In virtue of (8), (12), (13) and (17), for $\bar{\boldsymbol{w}}_h, \boldsymbol{w}_h, \boldsymbol{\varphi}_h \in \boldsymbol{H}^2(\mathscr{T}_h)$, we define the form $\boldsymbol{c}_h^L : \boldsymbol{H}^2(\mathscr{T}_h) \times \boldsymbol{H}^2(\mathscr{T}_h) \times \boldsymbol{H}^2(\mathscr{T}_h) \to \mathbb{R}$ by

$$\begin{aligned}
&\boldsymbol{c}_h^L(\bar{\boldsymbol{w}}_h, \boldsymbol{w}_h, \boldsymbol{\varphi}_h) \qquad\qquad (25)\\
&:= \sum_{\Gamma \in \mathscr{F}_h} \int_\Gamma \boldsymbol{P}^+ \left(\langle \bar{\boldsymbol{w}}_h \rangle, \boldsymbol{n} \right) \boldsymbol{w}_h|_\Gamma^{(+)} \cdot [\![\boldsymbol{\varphi}_h]\!] \, \mathrm{d}S + \sum_{\Gamma \in \mathscr{F}_h \setminus \mathscr{F}_h^{io}} \int_\Gamma \boldsymbol{P}^- \left(\langle \bar{\boldsymbol{w}}_h \rangle, \boldsymbol{n} \right) \boldsymbol{w}_h|_\Gamma^{(-)} \cdot [\![\boldsymbol{\varphi}_h]\!] \, \mathrm{d}S \\
&\quad - \sum_{K \in \mathscr{T}_h} \int_K \sum_{i=1}^{2} \boldsymbol{A}_i(\bar{\boldsymbol{w}}_h) \boldsymbol{w}_h \cdot \frac{\partial \boldsymbol{\varphi}_h}{\partial x_i} \, \mathrm{d}x + \sum_{K \in \mathscr{T}_h} \int_K \sum_{i,j=1}^{2} \left(\boldsymbol{K}_{i,j}(\bar{\boldsymbol{w}}_h) \frac{\partial \boldsymbol{w}_h}{\partial x_j} \right) \cdot \frac{\partial \boldsymbol{\varphi}_h}{\partial x_i} \, \mathrm{d}x \\
&\quad - \sum_{\Gamma \in \mathscr{F}_h^I} \int_\Gamma \sum_{i=1}^{2} \left\langle \sum_{j=1}^{2} \boldsymbol{K}_{i,j}(\bar{\boldsymbol{w}}_h) \frac{\partial \boldsymbol{w}_h}{\partial x_j} \right\rangle n_i \cdot [\![\boldsymbol{\varphi}_h]\!] \, \mathrm{d}S \\
&\quad - \sum_{\Gamma \in \mathscr{F}_h^{io}} \int_\Gamma \sum_{i=1}^{2} \sum_{j=1}^{2} \boldsymbol{K}_{i,j}(\bar{\boldsymbol{w}}_h|_\Gamma^{(+)}) \frac{\partial \boldsymbol{w}_h}{\partial x_j} \, n_i \cdot [\![\boldsymbol{\varphi}_h]\!] \, \mathrm{d}S \\
&\quad - \sum_{\Gamma \in \mathscr{F}_h^W} \int_\Gamma \sum_{i,j=1}^{2} \boldsymbol{K}_{i,j}^W(\bar{\boldsymbol{w}}_h|_\Gamma^{(+)}) \frac{\partial \boldsymbol{w}_h}{\partial x_j} \, n_i \cdot \boldsymbol{\varphi}_h \, \mathrm{d}S \\
&\quad + \sum_{\Gamma \in \mathscr{F}_h^I} \int_\Gamma \sigma [\![\boldsymbol{w}_h]\!] \cdot [\![\boldsymbol{\varphi}_h]\!] \, \mathrm{d}S + \sum_{\Gamma \in \mathscr{F}_h^{io}} \int_\Gamma \sigma \boldsymbol{w}_h \cdot \boldsymbol{\varphi}_h \, \mathrm{d}S + \sum_{\Gamma \in \mathscr{F}_h^W} \int_\Gamma \sigma \boldsymbol{w}_h \cdot \mathcal{V}(\boldsymbol{\varphi}_h) \, \mathrm{d}S
\end{aligned}$$

and the form $\boldsymbol{d}_h : \boldsymbol{H}^2(\mathscr{T}_h) \times \boldsymbol{H}^2(\mathscr{T}_h) \to \mathbb{R}$ by

$$\begin{aligned}
\boldsymbol{d}_h(\bar{\boldsymbol{w}}_h, \boldsymbol{\varphi}_h) := &- \sum_{\Gamma \in \mathscr{F}_h^{io}} \int_\Gamma \boldsymbol{P}^- \left(\langle \bar{\boldsymbol{w}}_h \rangle, \boldsymbol{n} \right) \bar{\boldsymbol{w}}_h|_\Gamma^{(-)} \cdot [\![\boldsymbol{\varphi}_h]\!] \, \mathrm{d}S \qquad (26)\\
&+ \sum_{\Gamma \in \mathscr{F}_h^{io}} \int_\Gamma \sigma \mathcal{B}(\bar{\boldsymbol{w}}_h) \cdot \boldsymbol{\varphi}_h \, \mathrm{d}S + \sum_{\Gamma \in \mathscr{F}_h^W} \int_\Gamma \sigma \mathcal{B}(\bar{\boldsymbol{w}}_h) \cdot \mathcal{V}(\boldsymbol{\varphi}_h) \, \mathrm{d}S.
\end{aligned}$$

Obviously, due to relations (8), (12), (13), (17) and (25) – (26), we have the consistency between forms $\boldsymbol{c}_h$ and $\boldsymbol{c}_h^L$, namely

$$\boldsymbol{c}_h(\boldsymbol{w}_h, \boldsymbol{\varphi}_h) = \boldsymbol{c}_h^L(\boldsymbol{w}_h, \boldsymbol{w}_h, \boldsymbol{\varphi}_h) - \boldsymbol{d}_h(\boldsymbol{w}_h, \boldsymbol{\varphi}_h) \qquad \forall \boldsymbol{w}_h, \boldsymbol{\varphi}_h \in \boldsymbol{H}^2(\mathscr{T}_h). \tag{27}$$

Furthermore, the form $\boldsymbol{c}_h^L$ is linear with respect to its second and third arguments.

Using the notation from Section 4.1, we define the $N_h \times N_h$ *flux matrix*

$$\mathbb{C}_h\left(\bar{\boldsymbol{\xi}}\right) := \left\{ \frac{\alpha_{n,0}}{\tau_k} \left(\boldsymbol{\varphi}_j, \boldsymbol{\varphi}_i \right)_{0,\Omega} + \boldsymbol{c}_h^L(\bar{\boldsymbol{w}}_h, \boldsymbol{\varphi}_j, \boldsymbol{\varphi}_i) \right\}_{i,j=1}^{N_h} \tag{28}$$

and the vector

$$\boldsymbol{q}_h\left(\left\{\boldsymbol{\xi}_{k-l}\right\}_{l=1}^{n},\bar{\boldsymbol{\xi}}\right):=\left\{-\frac{1}{\tau_k}\left(\sum_{i=1}^{n}\alpha_{n,i}\boldsymbol{w}_h^{k-l},\boldsymbol{\varphi}_i\right)_{0,\Omega}+\boldsymbol{d}_h(\bar{\boldsymbol{w}}_h,\boldsymbol{\varphi}_i)\right\}_{i=1}^{N_h}, \tag{29}$$

where $\boldsymbol{\varphi}_i \in B_h,\ i=1,\dots,N_h$ are the basis functions, $\bar{\boldsymbol{\xi}} \in \mathbb{R}^{N_h}$ and $\boldsymbol{\xi}_{k-l} \in \mathbb{R}^{N_h},\ l=1,\dots,n$ are the algebraic representation of $\bar{\boldsymbol{w}}_h \in \boldsymbol{S}_h^p$ and $\boldsymbol{w}_h^{k-l} \in \boldsymbol{S}_h^p,\ l=1,\dots,n$, respectively. We do not emphasise that $\mathbb{C}_h$ and $\boldsymbol{q}_h$ depend explicitly on τ_k. Finally, using (23) and (27) – (29), we have

$$\boldsymbol{F}_h(\left\{\boldsymbol{\xi}_{k-l}\right\}_{l=1}^{n};\boldsymbol{\xi}_k)=\mathbb{C}_h(\boldsymbol{\xi}_k)\boldsymbol{\xi}_k-\boldsymbol{q}_h(\left\{\boldsymbol{\xi}_{k-l}\right\}_{l=1}^{n},\boldsymbol{\xi}_k),\quad k=1,\dots,r. \tag{30}$$

Let us note that the flux matrix $\mathbb{C}_h$ has a block structure and it is sparse. In virtue of (25), we easily find that each block-row of $\mathbb{C}_h$ corresponds to one $K \in \mathscr{T}_h$ and it contains a diagonal block and several off-diagonal blocks. Each off-diagonal block corresponds to one face $\Gamma \in \mathscr{F}_h$. Obviously, the sparsity of $\mathbb{C}_h$ is identical to the sparsity of the Jacobi matrix $D\boldsymbol{F}_h(\left\{\boldsymbol{\xi}_{k-l}\right\}_{l=1}^{n};\boldsymbol{\xi})/D\boldsymbol{\xi}$. Therefore, in the following Newton-like method, we use $\mathbb{C}_h$ as the approximation of $D\boldsymbol{F}_h(\left\{\boldsymbol{\xi}_{k-l}\right\}_{l=1}^{n};\boldsymbol{\xi})/D\boldsymbol{\xi}$ in the definition of our iterative Newton-like method. This approximation follows from relation (30), when we fix the arguments of $\mathbb{C}_h$ and $\boldsymbol{q}_h$ and perform the differentiation with respect to $\boldsymbol{\xi}_k$.

Remark 4.1. *Let us mention computational costs of the evaluation of $\boldsymbol{F}_h$ and $\mathbb{C}_h$. For simplicity, let us consider conforming triangular grids. Then $\boldsymbol{F}_h$ has $N_h = \#\mathscr{T}_h(p+1)(p+1)/2$ entries and $\mathbb{C}_h$ has approximately $4\#\mathscr{T}_h((p+1)(p+1)/2)^2$ non-vanishing entries. Therefore, an evaluation of $\boldsymbol{F}_h$ is approximately $2(p+1)(p+2)$-times cheaper than an evaluation of $\mathbb{C}_h$.*

4.3 Iterative algorithm

In the following, we do not emphasise the dependence of $\boldsymbol{F}_h$ on $\boldsymbol{\xi}_{k-l},\ l=1,\dots,n$, hence we set $\boldsymbol{F}_h(\boldsymbol{\xi}) := \boldsymbol{F}_h(\left\{\boldsymbol{\xi}_{k-l}\right\}_{l=1}^{n};\boldsymbol{\xi})$. In order to determine solution $\boldsymbol{\xi}_k$ of the system (24), we employ a damped Newton-like method [15] which generates a sequence of approximations $\boldsymbol{\xi}_k^l,\ l=0,1,\dots$ to the actual numerical solution $\boldsymbol{\xi}_k$ using the following algorithm. Given an iterate $\boldsymbol{\xi}_k^l$, the update $\boldsymbol{d}^l$ of $\boldsymbol{\xi}_k^l$ to get to the next iterate

$$\boldsymbol{\xi}_k^{l+1} := \boldsymbol{\xi}_k^l + \lambda^l \boldsymbol{d}^l \tag{31}$$

is defined by: find $\boldsymbol{d}^l \in \mathbb{R}^{N_h}$ such that

$$\mathbb{C}_h(\boldsymbol{\xi}_k^l)\boldsymbol{d}^l = -\boldsymbol{F}_h(\boldsymbol{\xi}_k^l), \tag{32}$$

where $\mathbb{C}_h$ is the flux matrix given by (30) and $\lambda^l \in (0,1]$ is a damping parameter which ensures convergence of (31) – (32) in case when the initial guess $\boldsymbol{\xi}_k^0$ is far from the solution of (24).

4.3.1 Choice of the damping parameter

We start from the value $\lambda^l = 1$ and evaluate a monitoring function $\delta^l := \left\|\boldsymbol{F}_h(\boldsymbol{\xi}_k^{l+1})\right\| / \left\|\boldsymbol{F}_h(\boldsymbol{\xi}_k^l)\right\|$. If $\delta^l < 1$ we proceed to the next Newton iteration. Otherwise, we set $\lambda^l := \lambda^l/2$ and repeat the actual Newton iteration. Analysis of the convergence of the Newton method and the monitoring function can be found in [15].

4.3.2 Update of the flux matrix

Obviously, it is not necessary to update the flux matrix $\mathbb{C}_h(\boldsymbol{\xi}_k^l)$ at each Newton iteration $l=1,2,\dots$ and each time level $k=1,\dots,r$. In virtue of Remark 4.1, it is much cheaper to evaluate $\boldsymbol{F}_h$ than $\mathbb{C}_h$. Therefore, it is more efficient to perform more Newton iterations than to update $\mathbb{C}_h$. In practice, we update $\mathbb{C}_h$, either the damping parameter λ achieves a minimal prescribed value (using the algorithm described in Section 4.3.1) or the prescribed maximal number of Newton iteration was achieved.

4.3.3 Termination of iterative process

The iterative process (31) – (32) is terminated if a suitable *algebraic stopping criterion* is achieved. The standard approach is to set

$$\left\| \boldsymbol{F}_h(\boldsymbol{\xi}_k^l) \right\| \leq \mathrm{TOL}, \tag{33}$$

where $\|\cdot\|$ and TOL are a given norm and a given tolerance, respectively. However, it is difficult to choose TOL in order to guarantee the accuracy and in order to avoid an over-solution of the algebraic system. In Section 5.6, we present an algebraic stopping criterion following from the framework of residual error estimators.

4.3.4 Solution of the linear algebraic systems

The linear algebraic systems (32) are solved by the GMRES method ([42]) with the block ILU(0) preconditioner, its sparsity is the same as the sparsity of matrix $\mathbb{C}$, see [23] for detail. The solution from the previous Newton iteration is used as an initial condition for GMRES. In order not to oversolve the linear systems, we usually perform only few GMRES iterations, particularly we stop this iterative solver when the actual preconditioned residuum is two times smaller than the initial one. This criterion may seem to be too weak, however, numerical experiments (not presented here) indicate that it is sufficient.

4.3.5 Choice of the time step

We still have to specify the choice of the time steps $\tau_k,\ k = 1, \ldots, r$ in (20). Obviously, too large time steps cause a decrease of accuracy and on the other hand, too small time steps lead to a decrease in efficiency. Standard approaches (e.g., [30], [31]) estimate the local discretization error of the time discretization of the ODE system (19). Then the size of the time step is chosen based on this estimate in such a way that the local discretization error is under a given tolerance ω. We employed a similar approach in [20] for the n-BDF-DGFE scheme. However, the tolerance ω should be given (empirically) by a user. Our aim is to adapt the time step in such a way that the temporal error is controlled by the spatial one. In Section 5.6, we present an adaptive time step algorithm following from the framework of residual error estimators.

5 Error estimates

In this section we present the main novelty of this paper. We have defined the BDF-DGFE solution of the Navier-Stokes equations by (20), which was solved with the aid of the iterative algorithm (31) – (32). The total computational error (= the difference between the (unknown) exact solution and approximate solution resulting from (31) – (32)) depends on the following discretization parameters:

- space mesh $\mathscr{T}_h$,
- degrees of polynomial approximation p,
- size of time steps τ_k,
- degree of the BDF scheme n,
- an inexactness of the iterative solver represented by the violation of $\boldsymbol{F}_h(\,\cdot\,;\boldsymbol{\xi}_k) = \boldsymbol{0}$ introduced in (24).

Our goal is to identify the errors originating from the space discretization, from the time discretization and algebraic errors resulting from the inaccurate solution of (24). Particularly, we want to determine a stopping criterion for the iterative algorithm (31) – (32) and the strategy for adaptive choice of the time step is such a way that

- the resulting approximate BDF-DGFE solution is not essentially influenced by the time discretization and algebraic errors,
- the time partition $\mathcal{I}_\tau$ is not too fine and the nonlinear algebraic systems are not over-solved.

5.1 Functional spaces

We define several functional spaces, which are the base for the identification of the total, space, time and algebraic errors. Let $\mathcal{I}_\tau$ be the time partition of $(0,T)$ introduced in Section 3.4 and let $Q_{Kk} :=$

$K \times I_k,\ K \in \mathscr{T}_h, I_k \in \mathcal{I}_\tau$ denote a space-time element. We define the *broken space-time space* over $\mathscr{T}_h \times \mathcal{I}_\tau$ by

$$H^1(\mathcal{I}_\tau, \boldsymbol{H}^2(\mathscr{T}_h)) := \left\{\boldsymbol{\psi} : Q_T \to \mathbb{R}^4;\ \boldsymbol{\psi}|_{Q_{Kk}} \in H^1(I_k; \boldsymbol{H}^2(K)),\ K \in \mathscr{T}_h,\ I_k \in \mathcal{I}_\tau\right\}, \tag{34}$$

which consists of piecewise regular functions on space time elements $Q_{Kk},\ K \in \mathscr{T}_h, I_k \in \mathcal{I}_\tau$, which are in general discontinuous between two neighbouring elements $K, K' \in \mathscr{T}_h$ and between two time intervals $I_k, I_{k+1} \in \mathcal{I}_\tau$.

Let $\boldsymbol{S}_h^p$ be the space of vector-valued piecewise polynomial functions defined in Section 3.2 and let $P^q(\mathcal{I}_\tau) := \{\boldsymbol{v} : (0,T) \to \mathbb{R}^4,\ \boldsymbol{v}|_{I_k} \in [P^q(I_k)]^4,\ I_k \in \mathcal{I}_\tau\}$, where $q \geq 0$ is an integer and $P^q(I_k)$ is the space of vector-valued polynomials of order $\leq q$ on interval I_k, $k = 1, \ldots, r$. Now, we define three subspaces of $H^1(\mathcal{I}_\tau, \boldsymbol{H}^2(\mathscr{T}_h))$, namely

$$H^1(\mathcal{I}_\tau; \boldsymbol{S}_h^p) := \left\{\boldsymbol{\psi} \in H^1(\mathcal{I}_\tau, \boldsymbol{H}^2(\mathscr{T}_h));\ \boldsymbol{\psi}(\cdot, t) \in \boldsymbol{S}_h^p \text{ for a.e. } t \in (0,T)\right\}, \tag{35a}$$

$$S_\tau^q(\mathcal{I}_\tau; \boldsymbol{H}^2(\mathscr{T}_h)) := \left\{\boldsymbol{\psi} \in H^1(\mathcal{I}_\tau, \boldsymbol{H}^2(\mathscr{T}_h));\ \boldsymbol{\psi}(x, \cdot) \in P^q(\mathcal{I}_\tau) \text{ for a.e. } x \in \Omega\right\}, \tag{35b}$$

$$S_\tau^q(\mathcal{I}_\tau; \boldsymbol{S}_h^p) := \left\{\boldsymbol{\psi} \in H^1(\mathcal{I}_\tau, \boldsymbol{H}^2(\mathscr{T}_h));\ \boldsymbol{\psi}|_{Q_{Kk}} \in [P^p(K) \times P^q(I_k)]^4,\ K \in \mathscr{T}_h,\ I_k \in \mathcal{I}_\tau\right\}, \tag{35c}$$

where $P^p(K) \times P^q(I_k)$ is the space of polynomials on Q_{Kk} of the degree $\leq p$ with respect to $x \in K$ and the degree $\leq q$ with respect to $t \in I_k$ for $K \in \mathscr{T}_h$ and $I_k \in \mathcal{I}_\tau$. Therefore, all three spaces from (35) are piecewise regular on space time elements $Q_{Kk},\ K \in \mathscr{T}_h,\ I_k \in \mathcal{I}_\tau$ but generally discontinuous on Q_T. Moreover, $H^1(\mathcal{I}_\tau; \boldsymbol{S}_h^p)$ consists of functions piecewise polynomial with respect to the space coordinates, $S_\tau^q(\mathcal{I}_\tau; \boldsymbol{H}^2(\mathscr{T}_h))$ consists of functions piecewise polynomial with respect to the time coordinate and $S_\tau^q(\mathcal{I}_\tau; \boldsymbol{S}_h^p)$ consists of functions piecewise polynomial with respect to the space as well as the time coordinates. Obviously, each function $\boldsymbol{\psi}$ from $S_\tau^q(\mathcal{I}_\tau; \boldsymbol{H}^2(\mathscr{T}_h))$ on $\Omega \times I_k$ can be expressed as

$$\boldsymbol{\psi}|_{\Omega \times I_k} = \sum_{i=0}^{q} \boldsymbol{\varphi}_i^k(x)\phi_i^k(t), \quad \phi_i^k \in P^q(I_k), \qquad k = 0, 1, \ldots, r, \tag{36}$$

where $\boldsymbol{\varphi}_i^k(x) \in \boldsymbol{H}^2(\mathscr{T}_h)$ for $i = 0, \ldots, q,\ k = 0, 1, \ldots, r$. Similarly, each function $\boldsymbol{\psi}$ from $S_\tau^q(\mathcal{I}_\tau; \boldsymbol{S}_h^p)$ on $\Omega \times I_k$ can be expressed by (36) where $\boldsymbol{\varphi}_i^k(x) \in \boldsymbol{S}_h^p$ for $i = 0, \ldots, q,\ k = 0, 1, \ldots, r$.

Finally, we define a subspace of $S_\tau^q(\mathcal{I}_\tau; \boldsymbol{S}_h^p)$ by

$$P_\tau^q(\mathcal{I}_\tau; \boldsymbol{S}_h^p) := C(0, T; \boldsymbol{S}_h^p) \cap S_\tau^q(\mathcal{I}_\tau; \boldsymbol{S}_h^p), \tag{37}$$

which consists of functions continuous with respect to time on $(0,T)$, generally discontinuous in space and their restrictions on each space-time element $Q_{Kk},\ K \in \mathscr{T}_h,\ k = 1, \ldots, r$ are polynomial functions of degree q and p with respect to time and space, respectively. Obviously, we have inclusions

$$\begin{aligned} P_\tau^q(\mathcal{I}_\tau; \boldsymbol{S}_h^p) \subset S_\tau^q(\mathcal{I}_\tau; \boldsymbol{S}_h^p) \subset H^1(\mathcal{I}_\tau; \boldsymbol{S}_h^p) \subset H^1(\mathcal{I}_\tau, \boldsymbol{H}^2(\mathscr{T}_h)), \\ C(0, T; \boldsymbol{H}^2(\Omega)) \subset S_\tau^q(\mathcal{I}_\tau; \boldsymbol{H}^2(\mathscr{T}_h)) \subset H^1(\mathcal{I}_\tau, \boldsymbol{H}^2(\mathscr{T}_h)). \end{aligned} \tag{38}$$

Definition 5.1. *Let $\{\boldsymbol{v}^l\}_{l=0}^r$ be a finite sequence of functions $\boldsymbol{v}^l \in \boldsymbol{H}^2(\mathscr{T}_h),\ l = 0, \ldots, r$. We identify this sequence with a function $\bar{\boldsymbol{v}} \in S_\tau^0(\mathcal{I}_\tau; \boldsymbol{H}^2(\mathscr{T}_h))$ (piecewise constant with respect to time) such that*

$$\bar{\boldsymbol{v}} \in S_\tau^0(\mathcal{I}_\tau; \boldsymbol{H}^2(\mathscr{T}_h)),\ \bar{\boldsymbol{v}}|_{\Omega \times I_l} := \boldsymbol{v}^l,\ l = 1, \ldots, r. \tag{39}$$

For simplicity, we write $\bar{\boldsymbol{v}}|_{I_l} := \bar{\boldsymbol{v}}|_{\Omega \times I_l} = \boldsymbol{v}^l,\ l = 1, \ldots, r$ in the following. Similarly, we identify a finite sequence of functions $\{\boldsymbol{v}_h^l\}_{l=0}^r,\ \boldsymbol{v}_h \in \boldsymbol{S}_h^p,\ l = 0, \ldots, r$ with $\bar{\boldsymbol{v}}_h \in S_\tau^0(\mathcal{I}_\tau; \boldsymbol{S}_h^p)$ by (39) and we write $\bar{\boldsymbol{v}}_h|_{I_l} := \bar{\boldsymbol{v}}_h|_{\Omega \times I_l} = \boldsymbol{v}_h^l,\ l = 1, \ldots, r$.

5.2 Reformulation of the (semi-)discretizations

In this section, we introduce a reformulation of the semi-discrete and discrete problems from Section 3. We define the form $\boldsymbol{A}_{h\tau}(\boldsymbol{z}, \boldsymbol{\psi}) : H^1(\mathcal{I}_\tau, \boldsymbol{H}^2(\mathscr{T}_h)) \times H^1(\mathcal{I}_\tau, \boldsymbol{H}^2(\mathscr{T}_h)) \to \mathbb{R}$ by

$$\boldsymbol{A}_{h\tau}(\boldsymbol{z}, \boldsymbol{\psi}) := \int_0^T \left[\left(\frac{\partial \boldsymbol{z}}{\partial t}, \boldsymbol{\psi}\right)_{0,\Omega} + \boldsymbol{c}_h(\boldsymbol{z},\ \boldsymbol{\psi})\right] \mathrm{d}t, \qquad \boldsymbol{z}, \boldsymbol{\psi} \in H^1(\mathcal{I}_\tau, \boldsymbol{H}^2(\mathscr{T}_h)), \tag{40}$$

where $\boldsymbol{c}_h$ is given by (17), and the forms $\boldsymbol{A}_{h\tau}^k : S_\tau^0(\mathcal{I}_\tau; \boldsymbol{H}^2(\mathscr{T}_h)) \times \boldsymbol{H}^2(\mathscr{T}_h) \to \mathbb{R},\ k = 1, \ldots, r$ by

$$\boldsymbol{A}_{h\tau}^k(\bar{\boldsymbol{z}}, \boldsymbol{\varphi}) := \frac{1}{\tau_k}\left(\sum_{l=0}^{n} \alpha_{n,l}\bar{\boldsymbol{z}}|_{I_{k-l}}, \boldsymbol{\varphi}\right)_{0,\Omega} + \boldsymbol{c}_h(\bar{\boldsymbol{z}}|_{I_k}, \boldsymbol{\varphi}), \quad \bar{\boldsymbol{z}} \in S_\tau^0(\mathcal{I}_\tau; \boldsymbol{H}^2(\mathscr{T}_h)),\ \boldsymbol{\varphi} \in \boldsymbol{H}^2(\mathscr{T}_h), \tag{41}$$

where $n \geq 1$ is the order of the BDF method, $\alpha_{n,l},\ l = 0, \ldots, n$ are the corresponding BDF coefficients and $\bar{\boldsymbol{z}}|_{I_l},\ l = 0, \ldots, r$ has the meaning from Definition 5.1. Let us mention that each form $\boldsymbol{A}^k_{h\tau},\ k = 1, \ldots, r$ does not depend explicitly on $\bar{\boldsymbol{z}}(t)$ for all $t \in (0, T)$ but only on $\bar{\boldsymbol{z}}(t)$ for $t \in I_{k-n} \cup \cdots \cup I_k$. The range of t, on which $\boldsymbol{A}^k_{h\tau}$ depends non-trivially, is given by the superscripts of $\boldsymbol{A}^k_{h\tau}$.

Finally, we define the form $\overline{\boldsymbol{A}}_{h\tau} : S^0_\tau(\mathcal{I}_\tau; \boldsymbol{H}^2(\mathscr{T}_h)) \times S^0_\tau(\mathcal{I}_\tau; \boldsymbol{H}^2(\mathscr{T}_h)) \to \mathbb{R}$ by

$$\overline{\boldsymbol{A}}_{h\tau}(\bar{\boldsymbol{z}}, \bar{\boldsymbol{\psi}}) := \sum_{k=1}^{r} \tau_k \boldsymbol{A}^k_{h\tau}(\bar{\boldsymbol{z}}, \bar{\boldsymbol{\psi}}|_{I_k}), \quad \bar{\boldsymbol{z}}, \bar{\boldsymbol{\psi}} \in S^0_\tau(\mathcal{I}_\tau; \boldsymbol{H}^2(\mathscr{T}_h)), \tag{42}$$

where $\boldsymbol{A}^k_{h\tau},\ k = 1, \ldots, r$ are given by (41).

Now, we are ready to proceed to the reformulation of the relation for the exact and (semi)discrete solutions. Let $\boldsymbol{w} \in C^1(0, T, \boldsymbol{H}^2(\Omega))$ formally denote the exact solution of the Navier-Stokes equations (1) satisfying the corresponding initial and boundary conditions. We already mentioned that due to the consistency of the DGFE method, $\boldsymbol{w}$ satisfies the identity (18). This together with (40) implies that the *exact solution* $\boldsymbol{w}$ satisfies identity

$$\boldsymbol{A}_{h\tau}(\boldsymbol{w}, \boldsymbol{\psi}) = 0 \qquad \forall \boldsymbol{\psi} \in H^1(\mathcal{I}_\tau, \boldsymbol{H}^2(\mathscr{T}_h)). \tag{43}$$

Moreover, let $\boldsymbol{w}^l_h \in \boldsymbol{S}^p_h,\ l = 1, \ldots, r$ be the *approximate solution* given by (20), we identify it with a function $\bar{\boldsymbol{w}}_{h\tau} \in S^0_\tau(\mathcal{I}_\tau; \boldsymbol{S}^p_h)$ using (39). Then (21), (41) and (42) imply

$$\boldsymbol{A}^k_{h\tau}\left(\bar{\boldsymbol{w}}_{h\tau}, \boldsymbol{\varphi}_h\right) = 0 \quad \forall \boldsymbol{\varphi}_h \in \boldsymbol{S}^p_h,\ k = 1, \ldots, r \implies \overline{\boldsymbol{A}}_{h\tau}(\bar{\boldsymbol{w}}_{h\tau}, \bar{\boldsymbol{\psi}}_h) = 0 \quad \forall \bar{\boldsymbol{\psi}}_h \in S^0_\tau(\mathcal{I}_\tau; \boldsymbol{S}^p_h). \tag{44}$$

In Section 3.4, we derived the approximate solution $\boldsymbol{w}^k_h \in \boldsymbol{S}^p_h,\ k = 1, \ldots, r$ from the *space semi-discrete solution* $\boldsymbol{w}_h \in C^1([0, T]; \boldsymbol{S}^p_h)$ satisfying (19a). Then, in virtue of (40), we have

$$\boldsymbol{A}_{h\tau}(\boldsymbol{w}_h, \boldsymbol{\psi}_h) = 0 \qquad \forall \boldsymbol{\psi}_h \in H^1(\mathcal{I}_\tau; \boldsymbol{S}^p_h). \tag{45}$$

The approximate solution $\boldsymbol{w}^k_h \in \boldsymbol{S}^p_h,\ k = 1, \ldots, r$ satisfying (20) can be alternatively derived by the application of the BDF discretization to the weak formulation (18), which leads to a *time semi-discrete solution* $\boldsymbol{w}^k_\tau \in \boldsymbol{H}^2(\mathscr{T}_h),\ k = 1, \ldots, r$ satisfying

$$\frac{1}{\tau_k}\left(\sum_{l=0}^{n} \alpha_{n,l} \boldsymbol{w}^{k-l}_\tau, \boldsymbol{\varphi}\right)_{0,\Omega} + \boldsymbol{c}_h\left(\boldsymbol{w}^k_\tau, \boldsymbol{\varphi}\right) = 0 \quad \forall \boldsymbol{\varphi} \in \boldsymbol{H}^2(\mathscr{T}_h),\ k = 1, \ldots, r, \tag{46a}$$

$$\boldsymbol{w}^0_\tau \text{ is the } L^2\text{-projection of the initial condition } \boldsymbol{w}^0 \text{ in } \boldsymbol{H}^2(\mathscr{T}_h), \tag{46b}$$

where $n \geq 1$ is the degree of the BDF scheme and $\alpha_{n,l},\ l = 0, \ldots, n$ are the BDF coefficients in the same way as in Definition 3.2 and with the formalism given by Remark 3.3. Again, we identify the time semi-discrete solution $\{\boldsymbol{w}^l_\tau\}_{l=0}^r$ with a function $\bar{\boldsymbol{w}}_\tau \in S^0_\tau(\mathcal{I}_\tau; \boldsymbol{H}^2(\mathscr{T}_h))$ with the aid of (39), i.e, $\bar{\boldsymbol{w}}_\tau|_{I_l} = \boldsymbol{w}^l_\tau,\ l = 0, \ldots, r$. Then (46a), due to (41) and (42), can be written in the equivalent form

$$\begin{aligned} &\boldsymbol{A}^k_{h\tau}\left(\bar{\boldsymbol{w}}_\tau, \boldsymbol{\varphi}\right) = 0 \quad \forall \boldsymbol{\varphi} \in \boldsymbol{H}^2(\mathscr{T}_h),\ k = 1, \ldots, r \\ &\qquad\qquad \implies \quad \overline{\boldsymbol{A}}_{h\tau}(\bar{\boldsymbol{w}}_\tau, \bar{\boldsymbol{\psi}}) = 0 \quad \forall \bar{\boldsymbol{\psi}} \in S^0_\tau(\mathcal{I}_\tau; \boldsymbol{H}^2(\mathscr{T}_h)). \end{aligned} \tag{47}$$

We summarise the "solutions" of the Navier-Stokes equations (1) introduced above:

- *exact solution* $\boldsymbol{w} \in C^1(0, T; \boldsymbol{H}^2(\Omega))$ satisfying (43),
- *space semi-discrete solution* $\boldsymbol{w}_h \in C^1(0, T, \boldsymbol{S}^p_h)$ satisfying (45),
- *time semi-discrete solution* $\bar{\boldsymbol{w}}_\tau \in S^0_\tau(\mathcal{I}_\tau; \boldsymbol{H}^2(\mathscr{T}_h))$ satisfying (47),
- *approximate (or space-time discrete) solution* $\bar{\boldsymbol{w}}_{h\tau} \in S^0_\tau(\mathcal{I}_\tau; \boldsymbol{S}^p_h)$ satisfying (44).

5.3 Error measures and error estimators

Let $\boldsymbol{w}^k_h \in \boldsymbol{S}^p_h,\ k = 0, \ldots, r$ be the approximate solution given by (20). With the aid of (39), we have already introduced the function $\bar{\boldsymbol{w}}_{h\tau} \in S^0_\tau(\mathcal{I}_\tau; \boldsymbol{S}^p_h)$ which is piecewise constant on $I_k \in \mathcal{I}_\tau$ with respect to time. Moreover, we define a function $\boldsymbol{w}_{h\tau} \in P^n_\tau(\mathcal{I}_\tau; \boldsymbol{S}^p_h)$ (continuous and piecewise polynomials of degree n in time) by

$$\begin{aligned} &\boldsymbol{w}_{h\tau}(x, t_k) = \boldsymbol{w}^k_h(x),\ x \in \Omega,\ k = 0, \ldots, r, \\ &\boldsymbol{w}_{h\tau}(x, t)|_{\Omega \times I_k} = \mathscr{L}^n(\boldsymbol{w}^{k+1}_h, \boldsymbol{w}^k_h, \ldots, \boldsymbol{w}^{k-n+1}_k)|_{\Omega \times I_k}, \end{aligned} \tag{48}$$

where $\mathscr{L}^n$ is the Lagrangian interpolation of degree n in the space $\mathbb{R} \times \boldsymbol{S}_h^p$ constructed over pairs

$$(t_{k-n+1}, \boldsymbol{w}_h^{k-n+1}),\ (t_{k-n+2}, \boldsymbol{w}_h^{k-n+2}),\ \ldots,\ (t_k, \boldsymbol{w}_h^k),\ (t_{k+1}, \boldsymbol{w}_h^{k+1}).$$

(The idea of piecewise polynomial reconstruction in time was developed and analysed in [39, 2] for linear evolution problems). Therefore, both functions

$$\bar{\boldsymbol{w}}_{h\tau} \in S_\tau^0(\mathcal{I}_\tau; \boldsymbol{S}_h^p) \qquad \text{and} \qquad \boldsymbol{w}_{h\tau} \in P_\tau^n(\mathcal{I}_\tau; \boldsymbol{S}_h^p)$$

can be considered as the approximate solution of (1) given by (20).

Similarly as in, e.g., [3, 11, 13, 19], we employ an error measure in the dual norm in the following way. Let V be a linear vector space with a norm $\|\cdot\|_V$, (the space V does not need to be complete with respect to $\|\cdot\|_V$) and let $a(\cdot,\cdot) : V \times V \to \mathbb{R}$ be a form linear with respect to its second argument and let V_h be a finite dimensional subspace of V. Moreover, let $u \in V$ and $u_h \in V_h$ be an exact and approximate solution of a fictitious problem defined by

$$a(u, \varphi) = 0 \quad \forall \varphi \in V \qquad \text{and} \qquad a(u_h, \varphi_h) = 0 \quad \forall \varphi_h \in V_h, \tag{49}$$

respectively. Then the *error measure in the dual norm* on the space V is given by

$$E(u_h) := \|Au_h - Au\|_{V'} := \sup_{\substack{\varphi \in V \\ \varphi \neq 0}} \frac{a(u_h, \varphi) - a(u, \varphi)}{\|\varphi\|_V} = \sup_{\substack{\varphi \in V \\ \varphi \neq 0}} \frac{a(u_h, \varphi)}{\|\varphi\|_V}, \tag{50}$$

where A is the operator from V to its dual space corresponding to $a(\cdot,\cdot)$ given by $\langle Au, \varphi\rangle := a(u, \varphi),\ u, \varphi \in V$, where $\langle\cdot,\cdot\rangle$ denotes the duality between V and V'. The last equality in (50) follows from (49).

Now, in virtue of (49) – (50), we introduce an error measure of the approximate solution, particularly the *space-time* (total) *error*, the *time error*, the *space error* and the (nonlinear) *algebraic error*.

Space-time error is defined as the difference between the exact solution $\boldsymbol{w}$ and the approximate (=space-time discrete) solution $\boldsymbol{w}_{h\tau}$ in the dual norm of the space $H^1(\mathcal{I}_\tau, \boldsymbol{H}^2(\mathscr{T}_h))$, namely

$$\mathscr{E}_{ST}(\boldsymbol{w}_{h\tau}) := \sup_{\substack{\boldsymbol{\psi} \in H^1(\mathcal{I}_\tau, \boldsymbol{H}^2(\mathscr{T}_h)) \\ \boldsymbol{\psi} \neq 0}} \frac{\boldsymbol{A}_{h\tau}(\boldsymbol{w}_{h\tau}, \boldsymbol{\psi}) - \boldsymbol{A}_{h\tau}(\boldsymbol{w}, \boldsymbol{\psi})}{\|\boldsymbol{\psi}\|_X} = \sup_{\substack{\boldsymbol{\psi} \in H^1(\mathcal{I}_\tau, \boldsymbol{H}^2(\mathscr{T}_h)) \\ \boldsymbol{\psi} \neq 0}} \frac{\boldsymbol{A}_{h\tau}(\boldsymbol{w}_{h\tau}, \boldsymbol{\psi})}{\|\boldsymbol{\psi}\|_X}, \tag{51}$$

where $\|\cdot\|_X$ is a norm defined on $H^1(\mathcal{I}_\tau, \boldsymbol{H}^2(\mathscr{T}_h))$ (and on all its subspaces of course) and it will be specified later. The equality in (51) follows from (43).

Time error is defined as the difference between the space semi-discrete solution $\boldsymbol{w}_h$ (which is formally exact in time) and the approximate (= space-time discrete) solution $\boldsymbol{w}_{h\tau}$ in the dual norm of the space $H^1(\mathcal{I}_\tau; \boldsymbol{S}_h^p)$, namely

$$\mathscr{E}_T(\boldsymbol{w}_{h\tau}) := \sup_{\substack{\boldsymbol{\psi} \in H^1(\mathcal{I}_\tau; \boldsymbol{S}_h^p) \\ \boldsymbol{\psi} \neq 0}} \frac{\boldsymbol{A}_{h\tau}(\boldsymbol{w}_{h\tau}, \boldsymbol{\psi}) - \boldsymbol{A}_{h\tau}(\boldsymbol{w}_h, \boldsymbol{\psi})}{\|\boldsymbol{\psi}\|_X} = \sup_{\substack{\boldsymbol{\psi} \in H^1(\mathcal{I}_\tau; \boldsymbol{S}_h^p) \\ \boldsymbol{\psi} \neq 0}} \frac{\boldsymbol{A}_{h\tau}(\boldsymbol{w}_{h\tau}, \boldsymbol{\psi})}{\|\boldsymbol{\psi}\|_X}, \tag{52}$$

where the equality follows from (45).

Space error is defined as the difference between the time semi-discrete solution $\bar{\boldsymbol{w}}_\tau$ (which is formally exact in space) and the approximate (= space-time discrete) solution $\bar{\boldsymbol{w}}_{h\tau}$ in the dual norm of the space $S_\tau^0(\mathcal{I}_\tau; \boldsymbol{H}^2(\mathscr{T}_h))$, namely

$$\mathscr{E}_S(\bar{\boldsymbol{w}}_{h\tau}) := \sup_{\substack{\bar{\boldsymbol{\psi}} \in S_\tau^0(\mathcal{I}_\tau; \boldsymbol{H}^2(\mathscr{T}_h)) \\ \bar{\boldsymbol{\psi}} \neq 0}} \frac{\overline{\boldsymbol{A}}_{h\tau}(\bar{\boldsymbol{w}}_{h\tau}, \bar{\boldsymbol{\psi}}) - \overline{\boldsymbol{A}}_{h\tau}(\bar{\boldsymbol{w}}_\tau, \bar{\boldsymbol{\psi}})}{\|\bar{\boldsymbol{\psi}}\|_X} = \sup_{\substack{\bar{\boldsymbol{\psi}} \in S_\tau^0(\mathcal{I}_\tau; \boldsymbol{H}^2(\mathscr{T}_h)) \\ \bar{\boldsymbol{\psi}} \neq 0}} \frac{\overline{\boldsymbol{A}}_{h\tau}(\bar{\boldsymbol{w}}_{h\tau}, \bar{\boldsymbol{\psi}})}{\|\bar{\boldsymbol{\psi}}\|_X}, \tag{53}$$

where the equality follows from (47).

Algebraic error represents the inaccuracy of the solution of the discrete problem (21) $\Leftrightarrow$ (44) due to the iterative Newton-like method presented in Section 4.3. Let $\tilde{\boldsymbol{w}}_h^k \in \boldsymbol{S}_h^p,\ k = 0, \ldots, r$ correspond to the output $\boldsymbol{\xi}_k^n$ of this iterative method through the isomorphism (22) and $\boldsymbol{w}_h^k \in \boldsymbol{S}_h^p,\ k = 0, \ldots, r$ be the approximate solution satisfying (exactly) (44). Let $\tilde{\bar{\boldsymbol{w}}}_{h\tau} \in S_\tau^0(\mathcal{I}_\tau; \boldsymbol{S}_h^p)$ and $\bar{\boldsymbol{w}}_{h\tau} \in S_\tau^0(\mathcal{I}_\tau; \boldsymbol{S}_h^p)$ be the functions corresponding to $\{\tilde{\boldsymbol{w}}_h^k\}_{k=0}^r$ and $\{\boldsymbol{w}_h^k\}_{k=0}^r$ by (39), respectively. Then we set

$$\mathscr{E}_A(\tilde{\bar{\boldsymbol{w}}}_{h\tau}) := \sup_{\substack{\bar{\boldsymbol{\psi}}_h \in S_\tau^0(\mathcal{I}_\tau; \boldsymbol{S}_h^p) \\ \bar{\boldsymbol{\psi}} \neq 0}} \frac{\overline{\boldsymbol{A}}_{h\tau}(\tilde{\bar{\boldsymbol{w}}}_{h\tau}, \bar{\boldsymbol{\psi}}_h) - \overline{\boldsymbol{A}}_{h\tau}(\bar{\boldsymbol{w}}_{h\tau}, \bar{\boldsymbol{\psi}}_h)}{\|\bar{\boldsymbol{\psi}}_h\|_X} = \sup_{\substack{\bar{\boldsymbol{\psi}}_h \in S_\tau^0(\mathcal{I}_\tau; \boldsymbol{S}_h^p) \\ \bar{\boldsymbol{\psi}} \neq 0}} \frac{\overline{\boldsymbol{A}}_{h\tau}(\tilde{\bar{\boldsymbol{w}}}_{h\tau}, \bar{\boldsymbol{\psi}}_h)}{\|\bar{\boldsymbol{\psi}}_h\|_X}, \tag{54}$$

where the equality follows from (44).

However, an evaluation of error measures $\mathscr{E}_{ST}(\boldsymbol{w}_{h\tau})$, $\mathscr{E}_T(\boldsymbol{w}_{h\tau})$ and $\mathscr{E}_S(\bar{\boldsymbol{w}}_{h\tau})$ is practically impossible since the supremum is taken over infinite dimensional spaces $H^1(\mathcal{I}_\tau, \boldsymbol{H}^2(\mathscr{T}_h))$, $H^1(\mathcal{I}_\tau; \boldsymbol{S}_h^p)$ and $S_\tau^0(\mathcal{I}_\tau; \boldsymbol{H}^2(\mathscr{T}_h))$, respectively. Therefore, in our approach, we seek the maximum over some sufficiently large but finite dimension subspaces of the spaces mentioned above. Particularly, for $q = 0, 1, 2, \dots$ we define spaces

$$S_\tau^q(\mathcal{I}_\tau; \boldsymbol{S}_h^{p+1}) := \left\{ \boldsymbol{\psi} \in H^1(\mathcal{I}_\tau, \boldsymbol{H}^2(\mathscr{T}_h));\ \boldsymbol{\psi}|_{Q_{Kk}} \in [P^{p+1}(K) \times P^q(I_k)]^4, K \in \mathscr{T}_h,\ I_k \in \mathcal{I}_\tau \right\}, \tag{55}$$

which consists of vector-values piecewise polynomial functions on Q_{Kk} of the degree $\le p+1$ with respect to $x \in K$ and the degree $\le q$ with respect to $t \in I_k$ for $K \in \mathscr{T}_h$ and $I_k \in \mathcal{I}_\tau$.

Let $\tilde{\boldsymbol{w}}_h^k \in \boldsymbol{S}_h^p,\ k = 0, \dots, r$ denote the resulting approximate solution of the n-BDF-DGFE method (20) computed by the Newton-like method and let $\tilde{\boldsymbol{w}}_{h\tau} \in P_\tau^n(\mathcal{I}_\tau; \boldsymbol{S}_h^p)$ and $\bar{\tilde{\boldsymbol{w}}}_{h\tau} \in S_\tau^0(\mathcal{I}_\tau; \boldsymbol{S}_h^p)$ be the corresponding piecewise polynomial and piecewise constant reconstruction introduced in Section 5.3. Then based on (51) – (54), we define the *residual error estimators*

$$\eta_{\mathrm{ST}}(\tilde{\boldsymbol{w}}_{h\tau}) := \sup_{\substack{\boldsymbol{\psi}_h \in S_\tau^n(\mathcal{I}_\tau; \boldsymbol{S}_h^{p+1}) \\ \boldsymbol{\psi}_h \neq 0}} \frac{\boldsymbol{A}_{h\tau}(\tilde{\boldsymbol{w}}_{h\tau}, \boldsymbol{\psi}_h)}{\|\boldsymbol{\psi}_h\|_X}, \tag{56a}$$

$$\eta_{\mathrm{T}}(\tilde{\boldsymbol{w}}_{h\tau}) := \sup_{\substack{\boldsymbol{\psi}_h \in S_\tau^n(\mathcal{I}_\tau; \boldsymbol{S}_h^{p}) \\ \boldsymbol{\psi}_h \neq 0}} \frac{\boldsymbol{A}_{h\tau}(\tilde{\boldsymbol{w}}_{h\tau}, \boldsymbol{\psi}_h)}{\|\boldsymbol{\psi}_h\|_X}, \tag{56b}$$

$$\eta_{\mathrm{S}}(\bar{\tilde{\boldsymbol{w}}}_{h\tau}) := \sup_{\substack{\bar{\boldsymbol{\psi}}_h \in S_\tau^0(\mathcal{I}_\tau; \boldsymbol{S}_h^{p+1}) \\ \bar{\boldsymbol{\psi}}_h \neq 0}} \frac{\overline{\boldsymbol{A}}_{h\tau}(\bar{\tilde{\boldsymbol{w}}}_{h\tau}, \bar{\boldsymbol{\psi}}_h)}{\|\bar{\boldsymbol{\psi}}_h\|_X}, \tag{56c}$$

$$\eta_{\mathrm{A}}(\bar{\tilde{\boldsymbol{w}}}_{h\tau}) := \sup_{\substack{\bar{\boldsymbol{\psi}}_h \in S_\tau^0(\mathcal{I}_\tau; \boldsymbol{S}_h^{p}) \\ \bar{\boldsymbol{\psi}}_h \neq 0}} \frac{\overline{\boldsymbol{A}}_{h\tau}(\bar{\tilde{\boldsymbol{w}}}_{h\tau}, \bar{\boldsymbol{\psi}}_h)}{\|\bar{\boldsymbol{\psi}}_h\|_X} = \mathscr{E}_A(\bar{\tilde{\boldsymbol{w}}}_{h\tau}), \tag{56d}$$

which we call *space-time*, *time*, *space* and *algebraic residual error estimators*, respectively. Here, n denotes the order of the employed BDF scheme.

5.4 Properties of the residual error estimators

Obviously, if $\boldsymbol{w} \in C^1(0, T; \boldsymbol{H}^2(\Omega)) \subset H^1(\mathcal{I}_\tau, \boldsymbol{H}^2(\mathscr{T}_h))$ is the exact solution of (43) then due to the consistency, we have $0 = \eta_{\mathrm{ST}}(\boldsymbol{w}) = \eta_{\mathrm{T}}(\boldsymbol{w})$. Moreover, let $\bar{\boldsymbol{w}} \in S_\tau^0(\mathcal{I}_\tau; \boldsymbol{H}^2(\mathscr{T}_h))$ be constructed from functions $\boldsymbol{w}(\cdot, t_k) \in \boldsymbol{H}^2(\Omega),\ k = 1, \dots, r$ by Definition 5.1 then $\eta_{\mathrm{S}}(\bar{\boldsymbol{w}}) = O(\tau^{n+1}) = \eta_{\mathrm{A}}(\bar{\boldsymbol{w}})$, where n is the degree of the BDF scheme.

Furthermore, we have immediately lower bounds

$$\eta_{\mathrm{ST}}(\tilde{\boldsymbol{w}}_{h\tau}) \le \mathscr{E}_{ST}(\tilde{\boldsymbol{w}}_{h\tau}),\ \eta_{\mathrm{T}}(\tilde{\boldsymbol{w}}_{h\tau}) \le \mathscr{E}_T(\tilde{\boldsymbol{w}}_{h\tau}),\ \eta_{\mathrm{S}}(\bar{\tilde{\boldsymbol{w}}}_{h\tau}) \le \mathscr{E}_S(\bar{\tilde{\boldsymbol{w}}}_{h\tau}),\ \eta_{\mathrm{A}}(\bar{\tilde{\boldsymbol{w}}}_{h\tau}) = \mathscr{E}_A(\bar{\tilde{\boldsymbol{w}}}_{h\tau}), \tag{57}$$

$\tilde{\boldsymbol{w}}_{h\tau} \in P_\tau^n(\mathcal{I}_\tau; \boldsymbol{S}_h^p),\ \bar{\tilde{\boldsymbol{w}}}_{h\tau} \in S_\tau^0(\mathcal{I}_\tau; \boldsymbol{S}_h^p)$, since the suprema in the estimates η_* are taken over smaller spaces than the suprema in the error estimates $\mathscr{E}_*$. However, it is open if there exists an upper bound, i.e., $\mathscr{E}_*(\cdot) \le C\eta_*(\cdot)$, where $C > 0$. This will be the subject of future research.

Finally, from (56), we simply found that

$$\eta_{\mathrm{T}}(\tilde{\boldsymbol{w}}_{h\tau}) \le \eta_{\mathrm{ST}}(\tilde{\boldsymbol{w}}_{h\tau}), \quad \eta_{\mathrm{A}}(\bar{\tilde{\boldsymbol{w}}}_{h\tau}) \le \eta_{\mathrm{S}}(\bar{\tilde{\boldsymbol{w}}}_{h\tau}), \quad \tilde{\boldsymbol{w}}_{h\tau} \in P_\tau^n(\mathcal{I}_\tau; \boldsymbol{S}_h^p),\ \bar{\tilde{\boldsymbol{w}}}_{h\tau} \in S_\tau^0(\mathcal{I}_\tau; \boldsymbol{S}_h^p). \tag{58}$$

On the other hand, although the numerical experiments show (see Section 6.1) that

$$\eta_{\mathrm{S}}(\bar{\tilde{\boldsymbol{w}}}_{h\tau}) \le \eta_{\mathrm{ST}}(\tilde{\boldsymbol{w}}_{h\tau}), \quad \eta_{\mathrm{A}}(\bar{\tilde{\boldsymbol{w}}}_{h\tau}) \le \eta_{\mathrm{T}}(\tilde{\boldsymbol{w}}_{h\tau}), \quad \tilde{\boldsymbol{w}}_{h\tau} \in P_\tau^n(\mathcal{I}_\tau; \boldsymbol{S}_h^p),\ \bar{\tilde{\boldsymbol{w}}}_{h\tau} \in S_\tau^0(\mathcal{I}_\tau; \boldsymbol{S}_h^p), \tag{59}$$

the validity of (59) is open.

Remark 5.2. *It would be possible to define space $\boldsymbol{S}_h^{p+1}$ in a different way, e.g., to enrich it by polynomials of even higher degree or introduce some interelement splitting of elements $K \in \mathscr{T}_h$. However, any further enrichment of $\boldsymbol{S}_h^{p+1}$ requires additional computational time and the numerical experiments show that the presented choice of $\boldsymbol{S}_h^{p+1}$ is sufficient.*

5.5 Evaluation of the residual error estimators

In order to simplify the notation in the following, we introduce a generic definition of the residual error estimators (56) by

$$\eta_\star(\tilde{\boldsymbol{w}}_{h\tau}) := \sup_{\substack{\boldsymbol{\psi}_h\in X_h\\ \boldsymbol{\psi}_h\neq 0}} \frac{\boldsymbol{A}_{h\tau}(\tilde{\boldsymbol{w}}_{h\tau},\boldsymbol{\psi}_h)}{\|\boldsymbol{\psi}_h\|_X}, \tag{60}$$

which formally represents any definition (56a) or (56b) or (56c) or (56d), where X_h denotes the corresponding functional space in (56). We define the *residual error estimates at time interval* I_k by replacing (60) by

$$\eta_\star^k(\tilde{\boldsymbol{w}}_{h\tau}) := \sup_{\substack{0\neq\boldsymbol{\psi}_h\in X_h\\ \mathrm{supp}(\boldsymbol{\psi}_h)\subset\Omega\times I_k}} \frac{\boldsymbol{A}_{h\tau}(\tilde{\boldsymbol{w}}_{h\tau},\boldsymbol{\psi}_h)}{\|\boldsymbol{\psi}_h\|_X}, \qquad k=1,\dots,r \tag{61}$$

and the *element residual error estimate* by replacing (60) by

$$\eta_\star^{k,K}(\tilde{\boldsymbol{w}}_{h\tau}) := \sup_{\substack{0\neq\boldsymbol{\psi}_h\in X_h\\ \mathrm{supp}(\boldsymbol{\psi}_h)\subset K\times I_k}} \frac{\boldsymbol{A}_{h\tau}(\tilde{\boldsymbol{w}}_{h\tau},\boldsymbol{\psi}_h)}{\|\boldsymbol{\psi}_h\|_X}, \qquad k=1,\dots,r, \quad K\in\mathscr{T}_h. \tag{62}$$

The generic definitions (61) and (62) define the (space-time, time, space and algebraic) residual error estimates η_{ST}^k, η_{T}^k, η_{S}^k and η_{A}^k on the time interval I_k, $k=1,\dots,r$ and the element (space-time, time, space and algebraic) residual error estimates $\eta_{\mathrm{ST}}^{k,K}$, $\eta_{\mathrm{T}}^{k,K}$, $\eta_{\mathrm{S}}^{k,K}$ and $\eta_{\mathrm{A}}^{k,K}$ on element $K\in\mathscr{T}_h$ and on the time interval I_k, $k=1,\dots,r$, respectively.

For simplicity, we set $X := H^1(\mathcal{I}_\tau,\boldsymbol{H}^2(\mathscr{T}_h))$. Obviously, $S_\tau^n(\mathcal{I}_\tau;\boldsymbol{S}_h^p)\subset S_\tau^n(\mathcal{I}_\tau;\boldsymbol{S}_h^{p+1})\subset X$ and $S_\tau^0(\mathcal{I}_\tau;\boldsymbol{S}_h^p)\subset S_\tau^0(\mathcal{I}_\tau;\boldsymbol{S}_h^{p+1})\subset X$. If the norm $\|\cdot\|_X$ is suitable chosen then the evaluation of $\eta_\star$, $\star\in\{ST,T,S,A\}$ is cheap. First, we present the following lemma.

Lemma 5.3. *Let $(\cdot,\cdot)_X : X\times X\to\mathbb{R}$ be a scalar product generating the norm $\|\cdot\|_X$. Let $(\cdot,\cdot)_X$ satisfy the* element-orthogonality *condition*

$$(\psi_h,\psi_h')_X = 0 \qquad \begin{array}{l}\forall\psi_h,\psi_h'\in X \text{ such that}\\ \mathrm{supp}(\psi_h)\subset K\times I_k,\ \mathrm{supp}(\psi_h')\subset K'\times I_{k'},\ K\neq K'\ k\neq k'.\end{array} \tag{63}$$

Then

$$\eta_\star(\boldsymbol{w}_{h\tau})^2 = \sum_{k=1}^r \eta_\star^k(\boldsymbol{w}_{h\tau})^2 = \sum_{k=1}^r\sum_{K\in\mathscr{T}_h} \eta_\star^{k,K}(\boldsymbol{w}_{h\tau})^2. \tag{64}$$

Proof. The proof is analogous to the proof [18, Lemma 4.3]. □

If the norm $\|\cdot\|_X$ satisfies (63) then the global estimators can be evaluated simply be summing of squares of the element estimators, due to (64). Then, it is sufficient to evaluate the element residual error estimators $\eta_\star^{k,K}$, for all $k=1,\dots,r$ and all $K\in\mathscr{T}_h$. This is a standard task of seeking a constrained extremum over the finite dimensional space

$$X_h^{K,k} := \{\boldsymbol{\psi}_h\in X_h;\ \mathrm{supp}(\boldsymbol{\psi}_h)\subset K\times I_k\}\subset[P^{p+1}(K)\times P^q(I_k)]^4, \quad K\in\mathscr{T}_h,\ k=1,\dots,r.$$

We evaluate the form $\boldsymbol{A}_{h\tau}(\tilde{\boldsymbol{w}}_{h\tau},\boldsymbol{\psi}_h)$ in (62) for functions $\boldsymbol{\psi}_h$ forming a basis of $X_h^{K,k}$ and then we seek the constrained extremum by the technique of the Lagrange multipliers, see [18, Appendix] for more details. Let us note that the seeking of the constrained extrema is relatively fast in comparison to the other parts of the computational process, see Tables 4 and 6.

In order to fulfil the favorable property (64), we employ (based on numerical experiments) the scalar product

$$(u,v)_X := \left\{\int_0^T\left[(u,v)_{0,\Omega} + \frac{1}{\mathrm{Re}}\sum_{K\in\mathscr{T}_h}(\nabla u,\nabla v)_{0,K}\right]\mathrm{d}t\right\}^{1/2}, \ u,v\in X, \tag{65}$$

where Re is the Reynolds number, for inviscid flow we put $1/\mathrm{Re} := 0$. This scalar product satisfies (63) with the corresponding norm

$$\|u\|_X := \left\{\int_0^T\left[\|u\|_{L^2(\Omega)}^2 + \frac{1}{\mathrm{Re}}|u|_{\boldsymbol{H}^1(\mathscr{T}_h)}^2\right]\mathrm{d}t\right\}^{1/2}, \quad u\in X. \tag{66}$$

5.6 An employment of the error estimates in the solution strategy

In Section 4, we introduced the solution strategy for the solution of the sequence of the nonlinear algebraic systems given by Definition 3.2 with the aid of the inexact Newton-like method. Two aspects stay open there: the termination of the iterative process (31) – (32) and the choice of the time step. We employ the residual error estimates introduced in previous sections. Again, let $\tilde{\boldsymbol{w}}_h^k \in \boldsymbol{S}_h^p,\ k = 0, \dots, r$ denote the resulting approximate solution of the n-BDF-DGFE method (20) computed by iterative process (31) – (32) and let $\tilde{\boldsymbol{w}}_{h\tau} \in P_\tau^n(\mathcal{I}_\tau; \boldsymbol{S}_h^p)$ and $\bar{\tilde{\boldsymbol{w}}}_{h\tau} \in S_\tau^0(\mathcal{I}_\tau; \boldsymbol{S}_h^p)$ be the corresponding piecewise polynomial and constant reconstructions introduced in Section 5.3.

Termination of the iterative process (31) – (32) We stop this iterative process at time level t_k if the algebraic residual error estimate at time interval I_k is sufficiently small in comparison to the space residual error estimate at time interval I_k, i.e.,

$$\eta_{\mathrm{A}}^k(\bar{\tilde{\boldsymbol{w}}}_{h\tau}) \le c_A \eta_{\mathrm{S}}^k(\bar{\tilde{\boldsymbol{w}}}_{h\tau}), \quad k = 1, \dots, r, \tag{67}$$

where $0 < c_A < 1$ is a suitable constant which checks a relative influence of the nonlinear algebraic error to the space discretization. Its is reasonable to set $c_A \in [10^{-2}, 10^{-1}]$.

Choice of the time step in (20) The aim is to choose the time step τ_k such that the time residual error estimate at time interval I_k is controlled by the space residual error estimate at time interval I_k, i.e.,

$$\eta_{\mathrm{T}}^k(\tilde{\boldsymbol{w}}_{h\tau}) \le c_T \eta_{\mathrm{S}}^k(\bar{\tilde{\boldsymbol{w}}}_{h\tau}), \quad k = 1, \dots, r, \tag{68}$$

where $c_T > 0$ is a suitable constant representing a desired ratio of the time and space error. Therefore, at each time level $k = 1, \dots, r$, we evaluate estimates $\eta_{\mathrm{T}}^k(\tilde{\boldsymbol{w}}_{h\tau})$ and $\eta_{\mathrm{S}}^k(\bar{\tilde{\boldsymbol{w}}}_{h\tau})$ and define the "optimal" time step

$$\tau_k^{\mathrm{opt}} := \tau_k \left(\frac{c_T \eta_{\mathrm{S}}^k(\bar{\tilde{\boldsymbol{w}}}_{h\tau})}{\eta_{\mathrm{T}}^k(\tilde{\boldsymbol{w}}_{h\tau})} \right)^{1/(n+1)}. \tag{69}$$

Now, if condition (68) is not satisfied we repeat the k^{th}-time step with τ_k^{opt} instead of τ_k otherwise we proceed to the $(k+1)^{\mathrm{th}}$-time step with $\tau_{k+1} := \tau_k^{\mathrm{opt}}$. This technique is standard, more details can be found in [30, 31, 20].

For non-stationary flow simulation, its is reasonable to set $c_T \in [10^{-2}, 10^{-1}]$ in order to suppress the influence of the time discretization with respect to the space discretization. On the other hand, when we use the scheme (20) for a steady state simulation with the aid of time stabilization, c_T can be large, e.g., $c_T := 1$, $c_T := 10$ or even higher values. Finally, let us note that numerical experiments showed that parameters c_A and c_T should satisfy $c_A < c_T$ otherwise some instability may appear in the computation due to an insufficient solution of the algebraic systems. This is a reasonable requirement that the algebraic residual error estimate is smaller than the time residual error estimate.

Remark 5.4. *Let us note that it is possible to consider the* local criteria*, when we replace* (67) *and* (68) *by*

$$\eta_{\mathrm{A}}^{k,K}(\bar{\tilde{\boldsymbol{w}}}_{h\tau}) \le c'_A \eta_{\mathrm{S}}^{k,K}(\bar{\tilde{\boldsymbol{w}}}_{h\tau}) \quad \textit{and} \quad \eta_{\mathrm{T}}^{k,K}(\tilde{\boldsymbol{w}}_{h\tau}) \le c'_T \eta_{\mathrm{S}}^{k,K}(\bar{\tilde{\boldsymbol{w}}}_{h\tau}) \quad \forall K \in \mathscr{T}_h,\ k = 1, \dots, r,$$

respectively, where $c'_A > 0$ and $c'_T > 0$ are given tolerances. However, based on our experiences, these local criteria are problematic in the situation when the approximate solution is equal (up to a computer precision) to the exact one on some $K \in \mathscr{T}_h$ (e.g., the flow is constant on elements K which are far from the airfoil). Then we have $\eta_{\mathrm{A}}^{k,K}(\bar{\tilde{\boldsymbol{w}}}_{h\tau}) \approx \eta_{\mathrm{S}}^{k,K}(\bar{\tilde{\boldsymbol{w}}}_{h\tau})$ for these elements and thus the local criterion above can not be satisfied.

5.7 Space-time adaptive BDF-DGFE method

It is challenging to develop a full space-time adaptive technique for non-stationary problems. Hence, using the residual error estimators described above, we propose an adaptive algorithm which adapts (locally) the mesh and (globally) the size of the time step. Therefore, the computational grid will not be further fixed but it may change at each time level. By $\mathscr{T}_h^k$ we denote the mesh used on the time level $t_k,\ k = 1, \dots, r$.

The aim is to adapt the mesh and the time step in such a way that the residual error estimator η_{ST} is under a given tolerance $\omega > 0$, i.e.,

$$\eta_{\mathrm{ST}}(\tilde{\boldsymbol{w}}_{h\tau}) \le \omega. \tag{70}$$

In the computational process, we prescribe the tolerance for the space-time residual error estimates $\eta_{\mathrm{ST}}^k,\ k=1,\dots,r$ on the time interval I_k, namely

$$\eta_{\mathrm{ST}}^k(\tilde{\boldsymbol{w}}_{h\tau}) \le \omega_k, \quad \omega_k := \omega\sqrt{\tau_k/T}, \quad k=1,\dots,r, \tag{71}$$

where ω_k is the tolerance for the time level $t_k,\ k=1,\dots,r$. The condition (71) implies (70) due to (64). Moreover, in order to guarantee condition the (71), we prescribe the tolerance for the element space-time residual error estimates $\eta_{\mathrm{ST}}^{k,K}$, namely

$$\eta_{\mathrm{ST}}^{k,K}(\tilde{\boldsymbol{w}}_{h\tau}) \le \omega_{K,k}, \quad \omega_{K,k} := \omega_k(\#\mathscr{T}_h^k)^{-1/2}, \quad K\in\mathscr{T}_h^k,\ k=1,\dots,r, \tag{72}$$

where $\#\mathscr{T}_h^k$ denotes the number of triangles of $\mathscr{T}_h^k$. The condition (72) implies (70) due to (64).

Then we define the following *space-time adaptive process*:

1) let $\omega>0$ be a given tolerance, $\mathscr{T}_h^0$ the initial mesh and τ_0 the initial time step,
2) let $k=1$,
3) we solve problem (24) by the iterative method (31) – (32) till of the stopping criterion $\eta_{\mathrm{A}}^k(\bar{\tilde{\boldsymbol{w}}}_{h\tau}) \le c_A\eta_{\mathrm{S}}^k(\bar{\tilde{\boldsymbol{w}}}_{h\tau})$ is satisfied,
4) if $\eta_{\mathrm{T}}^k(\tilde{\boldsymbol{w}}_{h\tau}) > c_T\eta_{\mathrm{S}}^k(\bar{\tilde{\boldsymbol{w}}}_{h\tau})$ we adapt the time step τ_k according (69) and go to step 3),
5) if $\eta_{\mathrm{ST}}^{k,K}(\tilde{\boldsymbol{w}}_{h\tau}) > \omega_{K,k}$ for some $K\in\mathscr{T}_h^k$ then we refine K and go to step 3),
6) if $t_k=T$ then the computation finishes,
7) we put with $\mathscr{T}_h^{k+1} := \mathscr{T}_h^k$, $\tau_{k+1} := \tau_k^{\mathrm{opt}}$, $k := k+1$ and go to step 3).

The elements K violating condition (72) are refined by splitting into four daughter elements by connecting the centres of its edges. Then the so-called *hanging nodes* arise. Let us note that the presented space-time adaptive process requires several technical aspects which are not presented here. In order to achieve a reasonable computational performance, we need to prescribe the maximal number of level of refinement, the maximal level of the hanging node, etc. Furthermore, in order to increase the efficiency, we have to remove elements where the residual error estimate is sufficiently small. Moreover, the use of n-step BDF method for $n\ge 2$ requires an interpolation of the solution given on the previous meshes to the new one. In Section 6.4, we present the application of this algorithm to the viscous shock-vortex interaction. However, this is only a preliminary result showing some potential of the presented residual error estimates. The main drawback of this approach is the isotropic mesh refinement which is not too much efficient for problems with strong shock waves.

6 Numerical experiments

In this section, we present several numerical experiments which demonstrate the computational performance of the implicit n-step BDF-DGFE scheme (20) and the residual error estimators (56). Our aim is to demonstrate the efficiency of the presented solution strategy and its robustness with respect to the flow regime.

6.1 Isentropic vortex propagation

We consider the propagation of an isentropic vortex in compressible inviscid flow, analysed numerically in [43]. This example is suitable for the demonstration of the performance of the proposed residual error estimators since the regular exact solution is known and thus we can simply evaluate the computational error. Then we are able to identify the influence of the space and time discretization parameters h and τ, respectively, to the total computational error.

The computational domain is taken as $[0,10]\times[0,10]$, extended periodically in both directions. The mean flow is $\rho=1$, $\boldsymbol{v}=(1,1)$ (diagonal flow) and $\mathtt{p}=1$ (symbol $\mathtt{p}$ denotes the pressure of the flow whereas p denotes the degree of polynomial approximation). To this mean flow we add an *isentropic vortex*, i.e. perturbation in $\boldsymbol{v}$ and the temperature $\theta=\mathtt{p}/\rho$, but no perturbation in the entropy $\eta=\mathtt{p}/\rho^\gamma$:

$$\delta\boldsymbol{v} = \frac{\varepsilon}{2\pi}\exp[(1-r^2)/2](-\bar{x}_2,\bar{x}_1), \quad \delta\theta = -\frac{(\gamma-1)\varepsilon^2}{8\gamma\pi^2}\exp[1-r^2], \quad \delta\eta=0, \tag{73}$$

where $(-\bar{x}_2,\bar{x}_1)=(x_1-5,x_2-5)$, $r^2=x_1^2+x_2^2$, and the vortex strength $\varepsilon=5$. The perturbations $\delta\rho$ and $\delta\mathtt{p}$ are obtained from the above relations.

It is clear that the exact solution of the Euler equations with the initial conditions

$$\rho(x,0) = \bar{\rho} + \delta\rho, \quad \boldsymbol{v}(x,0) = \bar{\boldsymbol{v}} + \delta\boldsymbol{v}, \quad \mathrm{p}(x,0) = \bar{\mathrm{p}} + \delta\mathrm{p}, \tag{74}$$

and periodic boundary conditions is just the passive convection of the vortex with the mean velocity. Therefore, we are able to evaluate the computational error $\|\boldsymbol{w} - \tilde{\boldsymbol{w}}_{h\tau}\|$ over the space-time domain $Q_T := \Omega \times (0,T)$, where $\boldsymbol{w}$ is the exact solution of (1) and $\tilde{\boldsymbol{w}}_{h\tau} \in P_\tau^n(\mathcal{I}_\tau; \boldsymbol{S}_h^p)$ is the piecewise polynomial reconstruction of the output of the Newton-like iterative process (31) – (32) $\tilde{\boldsymbol{w}}_h^k,\ k = 0, \dots, r$ with the aid of (48). We evaluate $\|\boldsymbol{w} - \tilde{\boldsymbol{w}}_{h\tau}\|$ in the $L^2(0,T;\boldsymbol{H}^1(\mathscr{T}_h))$-seminorm and in the $L^2(Q_T)$-norm.

Investigation of the temporal discretization error We investigate the behaviour of the residual error estimates for different τ on a (fixed) unstructured quasi-uniform triangular grid with $\#\mathscr{T}_h = 986$ triangles. The *average element size* of this mesh is given by $h := 1/\sqrt{\#\mathscr{T}_h/2} \approx 0.450$. The simulation was performed with the aid of the n-BDF-DGFE scheme (20) for P_1, P_2 and P_3 polynomial approximation with respect to space and for $n = 1, 2$ and 3 (order of the method with respect to the time). We carried out computation for the final time $T = 10$ (1 period in time). In order to suppress the influence of the nonlinear algebraic error we set $c_A := 10^{-3}$ in (67).

In the following, we investigate namely the relation between the ratio $\eta_{\mathrm{T}}(\tilde{\boldsymbol{w}}_{h\tau})/\eta_{\mathrm{S}}(\bar{\tilde{\boldsymbol{w}}}_{h\tau})$ and the error $\|\boldsymbol{w} - \tilde{\boldsymbol{w}}_{h\tau}\|$. Therefore, we carried out 8 computations with fixed time step $\tau^\ell,\ \ell = 1, \dots, 8$ for each pair $(P_p, n),\ p = 1,2,3,\ n = 1,2,3$. We set $\tau^1 := 0.05$ and $\tau^{\ell+1} := \tau^\ell/2,\ \ell = 1, \dots, 7$.

The first part of Table 2 shows the value of errors $\|e_{h\tau}\|_1 := \|e_{h\tau}\|_{L^2(0,T;\boldsymbol{H}^1(\mathscr{T}_h))}$, $\|e_{h\tau}\|_0 := \|e_{h\tau}\|_{L^2(Q_T)}$, residual error estimators η_{A}, η_{S}, η_{T}, η_{ST}, the ratio $\eta_{\mathrm{T}}/\eta_{\mathrm{S}}$ and the computational time (some non-interesting cases are omitted). The computations where $\eta_{\mathrm{T}}/\eta_{\mathrm{S}} \le 0.01$ are bolted.

From these results, we observe the following facts:

- Generally, higher degree of DGFE method p and higher order of BDF n give smaller computational errors.
- The estimator η_{S} is almost independent of τ and the estimator η_{T} decreases for a decreasing τ.
- For each pair $(P_p, n),\ p = 1,2,3,\ n = 1,2,3$ there exists a limit value $\tau^\star$ such that for any time step $\tau^\ell < \tau^\star$ we obtain (almost) the same computational error as for $\tau^\star$. It means that (from a practical point of view) the temporal error is negligible compared to the spatial error. Obviously, the limit value $\tau^\star$ is larger for large degree n of the BDF scheme.
- The condition $\eta_{\mathrm{T}}(\tilde{\boldsymbol{w}}_{h\tau})/\eta_{\mathrm{S}}(\bar{\tilde{\boldsymbol{w}}}_{h\tau}) \le 10^{-2}$ looks like as a reasonable detection if the corresponding time step τ^ℓ is bellow the limit value $\tau^\star$. We can easily observe that the computational errors almost do not change for the bolted lines in the first part of Table 2, an exception is the first case $p = 1$ and $n = 1$.

The relation between the ratio $\eta_{\mathrm{T}}(\tilde{\boldsymbol{w}}_{h\tau})/\eta_{\mathrm{S}}(\bar{\tilde{\boldsymbol{w}}}_{h\tau})$ and the computational error is also demonstrated by Figures 1 and 2 which show the isolines of the Mach number and the Mach number distribution along a diagonal at $t = 10$ for the P_3 polynomial approximation, the 3-step BDF and selected $\tau^\ell,\ \ell = 1,2,3,7$. We observe that for larger τ^ℓ the quality of the solution is violated. However, there is now difference for $\ell = 3$ and $\ell = 7$. For $\ell = 3$, we have $\eta_{\mathrm{T}}(\tilde{\boldsymbol{w}}_{h\tau})/\eta_{\mathrm{S}}(\bar{\tilde{\boldsymbol{w}}}_{h\tau}) \approx 0.155$ which is sufficient for the quality of the visualised results.

The second part of Table 2 shows the results obtained by the adaptive choice of the time step with $c_T = 0.01$ (here τ means the average time step) presented in Section 5.6. We find that for each pair $(P_p, n),\ p = 1,2,3,\ n = 1,2,3$ we are able to achieve almost the same level of the computational error as for fixed time step computations with $\tau^\ell < \tau^\star$ but the computational time is significantly reduced.

Investigation of the spatial discretization error We investigate the behaviour of the residual error estimates for different meshes. We perform the computation on unstructured quasi-uniform triangular grids having 580, 2484 and 10008 elements which corresponds to the average element size $h = 0.587$, $h = 0.284$ and $h = 0.141$, respectively. For each grid, we employ the $P_k,\ k = 1,2,3$ polynomial approximation and the 3-step BDF with the adaptive choice of the time step with $c_T = 0.01$. The achieved results are presented in the third part of Table 2. The penultimate column in this table shows the index $i_X := \eta_{\mathrm{ST}}/\|e_{h\tau}\|_{L^2(0,T;\boldsymbol{H}^1(\mathscr{T}_h))}$ reflecting the ratio between the space-time residual error estimate η_{ST} and the computational error in the norm $\|\cdot\|_{L^2(0,T;\boldsymbol{H}^1(\mathscr{T}_h))}$. The index i_X is not the usual effectivity index since η_{ST} is the approximation of the error in the dual norm.

We observe that due to the adaptive choice of the time step with $c_T = 0.01$ in (68), the spatial error dominates the temporal one for each case. Moreover, the computational error in the norm $\|\cdot\|_{L^2(0,T;\boldsymbol{H}^1(\mathscr{T}_h))}$

FIXED TIME STEP, FIXED MESH with h = 4.50E-01

τ	P_p	n	$\|e_{h\tau}\|_1$	$\|e_{h\tau}\|_0$	η_{A}	η_{S}	η_{T}	η_{ST}	$\eta_{\mathrm{T}}/\eta_{\mathrm{S}}$	CPU(s)
5.00E-02	1	1	6.98E+00	3.82E+00	4.41E-03	8.00E+00	1.21E+00	8.11E+00	1.52E-01	293
2.50E-02	1	1	4.81E+00	2.43E+00	3.96E-03	9.17E+00	7.41E-01	9.20E+00	8.09E-02	454
1.25E-02	1	1	3.47E+00	1.52E+00	5.78E-03	1.00E+01	4.17E-01	1.00E+01	4.16E-02	633
6.25E-03	1	1	2.81E+00	1.01E+00	8.23E-04	1.05E+01	2.22E-01	1.05E+01	2.11E-02	1167
3.13E-03	1	1	2.52E+00	7.46E-01	7.85E-05	1.08E+01	1.15E-01	1.08E+01	1.06E-02	2335
1.57E-03	1	1	**2.40E+00**	**6.23E-01**	**7.13E-03**	**1.10E+01**	**5.91E-02**	**1.10E+01**	**5.39E-03**	3685
7.85E-04	1	1	**2.35E+00**	**5.65E-01**	**3.83E-03**	**1.10E+01**	**2.98E-02**	**1.10E+01**	**2.70E-03**	5590
3.92E-04	1	1	**2.32E+00**	**5.38E-01**	**1.62E-03**	**1.11E+01**	**1.49E-02**	**1.11E+01**	**1.35E-03**	11174
5.00E-02	1	2	2.65E+00	6.97E-01	6.15E-03	1.10E+01	4.04E-01	1.10E+01	3.68E-02	299
2.50E-02	1	2	2.36E+00	5.35E-01	5.21E-03	1.11E+01	1.94E-01	1.11E+01	1.74E-02	447
1.25E-02	1	2	**2.31E+00**	**5.16E-01**	**5.14E-03**	**1.11E+01**	**9.31E-02**	**1.11E+01**	**8.37E-03**	616
6.25E-03	1	2	**2.30E+00**	**5.13E-01**	**1.11E-03**	**1.11E+01**	**4.12E-02**	**1.11E+01**	**3.71E-03**	1208
3.13E-03	1	2	**2.30E+00**	**5.13E-01**	**4.59E-04**	**1.11E+01**	**1.68E-02**	**1.11E+01**	**1.51E-03**	2414
1.57E-03	1	2	**2.30E+00**	**5.13E-01**	**8.93E-03**	**1.11E+01**	**1.10E-02**	**1.11E+01**	**9.91E-04**	2953
7.85E-04	1	2	**2.30E+00**	**5.13E-01**	**3.46E-03**	**1.11E+01**	**4.21E-03**	**1.11E+01**	**3.78E-04**	5893
5.00E-02	1	3	2.33E+00	5.56E-01	6.12E-03	1.10E+01	3.64E-01	1.11E+01	3.29E-02	310
2.50E-02	1	3	2.31E+00	5.24E-01	5.87E-03	1.11E+01	1.87E-01	1.11E+01	1.69E-02	433
1.25E-02	1	3	**2.30E+00**	**5.16E-01**	**4.28E-03**	**1.11E+01**	**9.18E-02**	**1.11E+01**	**8.26E-03**	630
6.25E-03	1	3	**2.30E+00**	**5.14E-01**	**4.27E-04**	**1.11E+01**	**4.09E-02**	**1.11E+01**	**3.68E-03**	1252
3.13E-03	1	3	**2.30E+00**	**5.13E-01**	**4.64E-04**	**1.11E+01**	**1.68E-02**	**1.11E+01**	**1.51E-03**	2482
1.57E-03	1	3	**2.30E+00**	**5.13E-01**	**8.22E-03**	**1.11E+01**	**1.04E-02**	**1.11E+01**	**9.39E-04**	3070
7.85E-04	1	3	**2.30E+00**	**5.13E-01**	**3.34E-03**	**1.11E+01**	**4.10E-03**	**1.11E+01**	**3.69E-04**	6137
5.00E-02	2	1	6.61E+00	3.68E+00	4.74E-04	8.75E-01	1.20E+00	1.52E+00	1.37E+00	664
2.50E-02	2	1	4.07E+00	2.21E+00	5.70E-04	1.05E+00	7.40E-01	1.30E+00	7.08E-01	1195
1.25E-02	2	1	2.32E+00	1.23E+00	4.70E-04	1.18E+00	4.17E-01	1.26E+00	3.53E-01	1698
6.25E-03	2	1	1.29E+00	6.51E-01	5.42E-04	1.27E+00	2.22E-01	1.30E+00	1.75E-01	2802
3.13E-03	2	1	7.59E-01	3.39E-01	1.95E-04	1.33E+00	1.15E-01	1.33E+00	8.67E-02	4367
1.57E-03	2	1	5.21E-01	1.77E-01	1.76E-05	1.36E+00	5.87E-02	1.36E+00	4.32E-02	8680
7.85E-04	2	1	4.32E-01	9.63E-02	1.35E-06	1.37E+00	2.96E-02	1.37E+00	2.15E-02	17482
3.92E-04	2	1	4.03E-01	5.89E-02	9.59E-04	1.38E+00	1.49E-02	1.38E+00	1.08E-02	23213
5.00E-02	2	2	1.08E+00	4.39E-01	6.89E-04	1.35E+00	2.71E-01	1.38E+00	2.01E-01	676
2.50E-02	2	2	4.85E-01	1.19E-01	7.27E-04	1.38E+00	8.97E-02	1.38E+00	6.49E-02	1157
1.25E-02	2	2	4.00E-01	4.64E-02	6.02E-04	1.39E+00	3.28E-02	1.39E+00	2.36E-02	1600
6.25E-03	2	2	**3.90E-01**	**3.66E-02**	**8.13E-04**	**1.39E+00**	**1.31E-02**	**1.39E+00**	**9.40E-03**	2370
3.13E-03	2	2	**3.89E-01**	**3.57E-02**	**8.88E-05**	**1.39E+00**	**5.26E-03**	**1.39E+00**	**3.78E-03**	4523
1.57E-03	2	2	**3.89E-01**	**3.56E-02**	**4.84E-05**	**1.39E+00**	**2.06E-03**	**1.39E+00**	**1.48E-03**	9014
7.85E-04	2	2	**3.89E-01**	**3.56E-02**	**1.80E-05**	**1.39E+00**	**7.74E-04**	**1.39E+00**	**5.56E-04**	17738
5.00E-02	2	3	1.24E+00	2.11E-01	8.76E-04	1.71E+00	5.97E-01	1.81E+00	3.50E-01	661
2.50E-02	2	3	3.95E-01	4.83E-02	7.12E-04	1.39E+00	7.60E-02	1.39E+00	5.48E-02	1122
1.25E-02	2	3	3.89E-01	3.70E-02	4.82E-04	1.39E+00	3.04E-02	1.39E+00	2.19E-02	1610
6.25E-03	2	3	**3.89E-01**	**3.58E-02**	**6.62E-04**	**1.39E+00**	**1.26E-02**	**1.39E+00**	**9.09E-03**	2379
3.13E-03	2	3	**3.89E-01**	**3.56E-02**	**7.04E-05**	**1.39E+00**	**5.18E-03**	**1.39E+00**	**3.73E-03**	4665
1.57E-03	2	3	**3.89E-01**	**3.56E-02**	**6.40E-06**	**1.39E+00**	**2.04E-03**	**1.39E+00**	**1.47E-03**	9189
7.85E-04	2	3	**3.88E-01**	**3.56E-02**	**3.94E-05**	**1.39E+00**	**7.72E-04**	**1.39E+00**	**5.55E-04**	17999
1.25E-02	3	1	2.24E+00	1.23E+00	5.97E-05	1.28E-01	4.17E-01	4.39E-01	3.27E+00	4601
6.25E-03	3	1	1.20E+00	6.47E-01	5.47E-05	1.41E-01	2.22E-01	2.67E-01	1.58E+00	6990
3.13E-03	3	1	6.21E-01	3.33E-01	2.06E-05	1.49E-01	1.15E-01	1.91E-01	7.72E-01	12622
1.57E-03	3	1	3.20E-01	1.70E-01	4.35E-05	1.54E-01	5.87E-02	1.66E-01	3.82E-01	17855
7.85E-04	3	1	1.68E-01	8.55E-02	3.69E-06	1.56E-01	2.96E-02	1.60E-01	1.89E-01	35690
2.50E-02	3	2	2.52E-01	1.11E-01	7.66E-05	1.57E-01	8.40E-02	1.79E-01	5.35E-01	2726
1.25E-02	3	2	8.10E-02	2.80E-02	6.90E-05	1.59E-01	2.75E-02	1.61E-01	1.73E-01	4273
6.25E-03	3	2	5.67E-02	7.66E-03	6.06E-05	1.59E-01	9.36E-03	1.59E-01	5.89E-02	6636
3.13E-03	3	2	5.54E-02	3.67E-03	7.69E-05	1.59E-01	3.30E-03	1.59E-01	2.07E-02	11393
1.57E-03	3	2	**5.54E-02**	**3.27E-03**	**1.81E-05**	**1.59E-01**	**1.18E-03**	**1.59E-01**	**7.42E-03**	18216
7.85E-04	3	2	**5.55E-02**	**3.24E-03**	**1.36E-06**	**1.59E-01**	**4.21E-04**	**1.59E-01**	**2.65E-03**	36108
2.50E-02	3	3	1.22E+00	9.63E-02	5.69E-04	1.07E+00	5.07E-01	1.18E+00	4.73E-01	2665
1.25E-02	3	3	5.76E-02	8.87E-03	6.82E-05	1.59E-01	2.47E-02	1.61E-01	1.55E-01	4216
6.25E-03	3	3	5.56E-02	3.91E-03	4.60E-05	1.59E-01	8.84E-03	1.59E-01	5.56E-02	6736
3.13E-03	3	3	5.55E-02	3.30E-03	9.15E-05	1.59E-01	3.20E-03	1.59E-01	2.01E-02	10502
1.57E-03	3	3	**5.55E-02**	**3.25E-03**	**1.34E-05**	**1.59E-01**	**1.16E-03**	**1.59E-01**	**7.30E-03**	19118
7.85E-04	3	3	**5.55E-02**	**3.24E-03**	**9.62E-07**	**1.59E-01**	**4.17E-04**	**1.59E-01**	**2.62E-03**	37156

ADAPT TIME STEP, FIXED MESH with h = 4.50E-01

τ	P_p	n	$\|e_{h\tau}\|_1$	$\|e_{h\tau}\|_0$	η_{A}	η_{S}	η_{T}	η_{ST}	$\eta_{\mathrm{T}}/\eta_{\mathrm{S}}$	CPU(s)
2.37E-03	1	1	2.46E+00	6.89E-01	2.21E-04	1.09E+01	8.75E-02	1.09E+01	8.04E-03	3006
3.17E-02	1	2	2.40E+00	5.35E-01	5.99E-03	1.11E+01	8.03E-02	1.11E+01	7.21E-03	409
5.78E-02	1	3	2.26E+00	4.77E-01	6.26E-03	1.12E+01	7.33E-02	1.12E+01	6.54E-03	286
2.93E-04	2	1	3.98E-01	5.09E-02	6.63E-04	1.38E+00	1.13E-02	1.38E+00	8.16E-03	27515
1.14E-02	2	2	3.96E-01	4.20E-02	4.90E-04	1.39E+00	1.01E-02	1.39E+00	7.28E-03	1734
2.52E-02	2	3	3.87E-01	3.44E-02	7.51E-04	1.39E+00	9.10E-03	1.40E+00	6.52E-03	1141
3.81E-03	3	2	5.54E-02	4.06E-03	3.37E-05	1.59E-01	1.15E-03	1.59E-01	7.26E-03	10165
1.40E-02	3	3	5.62E-02	3.51E-03	7.29E-05	1.59E-01	1.05E-03	1.59E-01	6.61E-03	4001

SPACE CONVERGENCE with adapt time step

h	P_p	n	$\|e_{h\tau}\|_1$	$\|e_{h\tau}\|_0$	η_{A}	η_{S}	η_{T}	η_{ST}	$\eta_{\mathrm{T}}/\eta_{\mathrm{S}}$	i_X	CPU(s)
5.87E-01	1	3	3.07E+00	9.41E-01	7.83E-03	1.37E+01	8.93E-02	1.37E+01	6.51E-03	4.47	160
2.84E-01	1	3	1.36E+00	1.63E-01	4.12E-03	7.55E+00	4.99E-02	7.55E+00	6.62E-03	5.54	993
1.41E-01	1	3	6.66E-01	3.65E-02	2.04E-03	3.85E+00	2.57E-02	3.85E+00	6.68E-03	5.78	7212
5.87E-01	2	3	6.94E-01	9.02E-02	1.28E-03	2.34E+00	1.52E-02	2.34E+00	6.51E-03	3.36	539
2.84E-01	2	3	1.61E-01	9.31E-03	2.79E-04	5.71E-01	3.82E-03	5.71E-01	6.69E-03	3.55	4326
1.41E-01	2	3	3.99E-02	1.12E-03	5.71E-05	1.43E-01	9.51E-04	1.43E-01	6.66E-03	3.58	35330
5.87E-01	3	3	1.31E-01	1.07E-02	1.70E-04	3.53E-01	2.30E-03	3.53E-01	6.52E-03	2.70	1869
2.84E-01	3	3	1.57E-02	6.87E-04	1.74E-05	4.21E-02	2.77E-04	4.21E-02	6.59E-03	2.68	16247
1.41E-01	3	3	2.83E-03	6.80E-05	1.76E-06	5.30E-03	3.54E-05	5.30E-03	6.69E-03	1.87	134013

Tab. 2: Isentropic vortex propagation: the value of errors $\|e_{h\tau}\|_1$, $\|e_{h\tau}\|_0$, residual error estimators η_{A}, η_{S}, η_{T}, η_{ST} and the ratio $\eta_{\mathrm{T}}/\eta_{\mathrm{S}}$, (results with $\eta_{\mathrm{T}}/\eta_{\mathrm{S}} \leq 0.01$ are bolted) and $i_X := \eta_{\mathrm{ST}}/\|e_{h\tau}\|_1$.

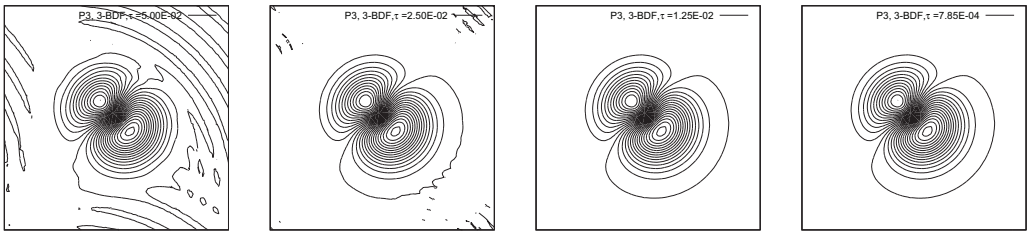

Fig. 1: Isentropic vortex propagation: isolines of the Mach number at $t = 10$ for P_3, 3-BDF and $\tau^\ell = 0.05/2^{\ell-1},\ \ell = 1, 2, 3, 7$.

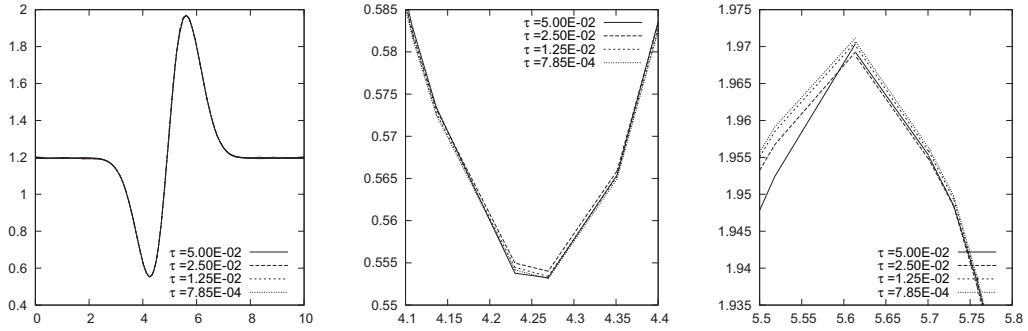

Fig. 2: Isentropic vortex propagation: distribution of the Mach number along the diagonal cut $\{(x_1, x_2) \in \mathbb{R}^2,\ (x_1, x_2) = (s, 10 - s),\ s \in (0, 10)\}$ at $t = 10$ for P_3, 3-BDF and $\tau^\ell = 0.05/2^{\ell-1},\ \ell = 1, 2, 3, 7$, total view (left) with two details.

converges with respect to h in the same order as the space-time residual error estimate η_{ST}. Hence, this estimate gives a reasonable information about the computational error when meshes are refined.

Comparison of different solution strategies We study the computational performance of the proposed BDF-DGFE method, namely the algebraic stopping criterion (67). For the isentropic vortex propagation, we compare several solution strategies for the solution of the nonlinear algebraic systems (24):

new : the strategy proposed in this paper, i.e., the iterative solver (31) – (32) with the algebraic stopping criterion (67),

imp7 : the iterative solver (31) – (32) with the "standard" algebraic stopping criterion (33), where we prescribe the tolerance TOL $:= \epsilon N_h$, where N_h is the number of equations in (24) and $\epsilon = 10^{-7}$,

imp9 : as *imp7* with $\epsilon = 10^{-9}$,

imp11 : as *imp7* with $\epsilon = 10^{-11}$,

semi : the semi-implicit time discretization developed in [17, 23].

We performed the computation by these techniques on two grids with the average element size $h_1 = 0.587$ and $h_2 = 0.284$ and with the aid of P_1 and P_3 polynomial approximation and the 3-steps BDF method. Table 3 shows this comparison, namely the total number of the time steps, the total number of the Newton iterations, the total number of the GMRES iterations and the total computational time.

We observe that the technique *semi* is inefficient due to an insufficient solution of the nonlinear algebraic systems on each time level which causes a significant decrease of the size of the time step in comparison to the approach *new*. Similarly, *imp7* is inefficient for all cases, only one Newton iteration (and one GMRES iteration) is performed at each time step since the tolerance TOL $= 10^{-7} N_h$ in the stopping criterion (33) is too high. Consequently, the insufficient solution of the nonlinear algebraic systems causes a significant decrease of the size of the time step.

The strategy *imp9* is faster than *new* for pairs (h_1, P_1) and (h_2, P_1) but inefficient for P_3 approximation. Moreover, *imp11* is comparable with *new* for (h_1, P_1), (h_2, P_1) and even for the pair (h_1, P_3) but not for

method	h	P_k	time steps	Newton iters	GMRES iters	CPU time
new	5.87E-01	1	178	807	807	162.3
imp7	5.87E-01	1	1267	1267	1267	564.9
imp9	5.87E-01	1	178	668	668	161.0
imp11	5.87E-01	1	177	845	845	190.0
semi	5.87E-01	1	1267	1267	1267	718.4
new	5.87E-01	3	535	2446	2446	1855.1
imp7	5.87E-01	3	13867	13867	13867	24650.7
imp9	5.87E-01	3	13867	13867	13867	24547.8
imp11	5.87E-01	3	535	1793	1793	2308.2
semi	5.87E-01	3	13867	13867	13867	26896.1
new	2.84E-01	1	232	1149	1149	985.8
imp7	2.84E-01	1	3156	3156	3156	6068.3
imp9	2.84E-01	1	233	832	832	872.3
imp11	2.84E-01	1	232	1102	1102	1089.4
semi	2.84E-01	1	3156	3156	3156	7691.6
new	2.84E-01	3	1106	5008	5008	16278.5
imp7	2.84E-01	3	38964	38964	38964	296222.1
imp9	2.84E-01	3	38979	38979	38979	295667.2
imp11	2.84E-01	3	38924	38924	38924	296329.4

Tab. 3: Isentropic vortex propagation: comparison of techniques *new*, *imp7*, *imp9*, *imp11* and *semi*.

$(h_2,\ P_3)$. Therefore, we conclude that for each pair $(h_i,\ P_k)$ we are able to find $\epsilon > 0$ such that the strategy with the stopping criterion (33) with TOL $:= \epsilon N_h$ is more efficient than (or comparable to) *new*. However, the problem is that we have to seek the optimal value ϵ experimentally. This is advantage of our approach that the stopping criterion (67) works universally.

Finally, let us note that the technique *semi* is more efficient than the others for steady-state solutions when we do not need to take a care about the accuracy with respect to the time.

Computational costs of the algorithm We study the relative computational costs for several parts of the proposed BDF-DGFE technique. Namely, we distinguish the following individual parts of our code:

- *matrices-vectors*: the evaluation of the flux matrices $\mathbb{C}_h$ and the vectors $\boldsymbol{F}_h$ in (32) for all Newton iterations in all time steps,
- *linear systems*: solution of the linear algebraic systems (32) for all Newton iterationss in all time steps,
- *reconstructions*: the piecewise polynomial and constant reconstruction $\boldsymbol{w}_{h\tau}$ and $\bar{\boldsymbol{w}}_{h\tau}$ including the evaluation of form $\boldsymbol{A}_{h\tau}$ in (62) for the basis function $\boldsymbol{\psi}_h$ of space $X_h^{K,k},\ K \in \mathscr{T}_h,\ k = 1, \ldots, r$,
- *constrained extrema*: the computing of the constrained extrema in (62) for each $K \in \mathscr{T}_h,\ k = 1, \ldots, r$,
- *others*: the pre-processing, the post-processing, saving of files, etc.

Table 4 shows the relative computational times for these parts of the code obtained by the 3-steps BDF-DGFE method with P_1 and P_3 polynomial approximation on the grids with $h_1 = 0.587$ and $h_2 = 0.284$. We simply observe that the computation of the constrained extrema does not essentially increase the total computational time and its relative computational costs are lower for the higher degrees of polynomial approximation and in fact independent of h. This supports the assertion from Section 5.5 that the seeking of the constrained extrema is relatively fast in comparison to the other parts of the computational process.

6.2 Inviscid low Mach number flow

In the second example, we deal with an inviscid low Mach number flow around the NACA 0012 airfoil. This case was solved in [5] where a preconditioned technique was developed. The experiments presented here were obtained by the standard (unpreconditioned) Vijayasundaram numerical flux.

h	P_k	matrices vectors	linear systems	reconstructions $\boldsymbol{w}_{h\tau}$, $\bar{\boldsymbol{w}}_{h\tau}$	constrained extrema	others	total
5.87E-01	1	25.7 %	21.5 %	29.7 %	8.2 %	14.9 %	100.0%
5.87E-01	3	24.6 %	41.0 %	16.7 %	6.0 %	11.7 %	100.0%
2.84E-01	1	26.4 %	22.1 %	29.8 %	8.2 %	13.6 %	100.0%
2.84E-01	3	24.5 %	40.8 %	16.9 %	6.0 %	11.8 %	100.0%

Tab. 4: Relative computational costs for each individual part of the BDF-DGFE algorithm.

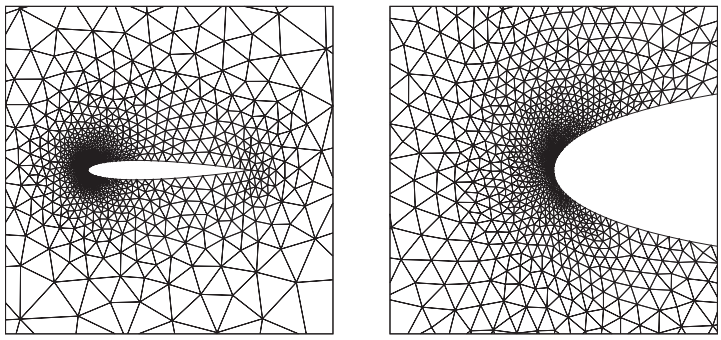

Fig. 3: The grid used, details of the profile (left) and leading edge (right).

The angle of attack is equal to zero and the far-field Mach number M_∞ is set to 10^{-1}, 10^{-2}, 10^{-3} and 10^{-4}. The computations were carried out on a grid having 3 587 elements (see Figure 3) with the aid of the n-BDF-DGFE method (20) with $n = 1, 2, 3$ and $P_p,\ p = 1, 2, 3, 4$ polynomial approximation. We employ the algorithm from Section 4 with the adaptive choice of the time step and the stopping criterion (67) presented in Section 5.6.

Obviously, the data settings of this example lead to a steady state solution. In this case, when a steady state solution is sought, we can employ our algorithm with no care about the size of the temporal error. In practice, we employ condition (68) with, e.g., $c_T = 10$. Figure 4 shows the comparison of the convergence of the solution to the steady state solution with $M = 0.1$ for P_3 approximation, the 3-steps BDF and with $c_T = 10$ and $c_T = 10^{-1}$ in (68). Similarly as in [23], the convergence to the steady-state is measured by $\left\|\boldsymbol{F}_h^{\mathrm{SS}}(\boldsymbol{\xi}_k)\right\|_{\ell^2}$, where (according to (23)) we set

$$\boldsymbol{F}_h^{\mathrm{ss}}\left(\boldsymbol{\xi}_k\right) := \left\{\boldsymbol{c}_h(\boldsymbol{w}_h^k, \boldsymbol{\varphi}_i)\right\}_{i=1}^{N_h} \in \mathbb{R}^{N_h},\ k = 1, \ldots, r, \tag{75}$$

where $\boldsymbol{\xi}_k \in \mathbb{R}^{N_h}$ is the algebraic representation of $\boldsymbol{w}_h^k \in \boldsymbol{S}_h^p$. We simply found that with $c_T = 10$ we achieve the steady state in significantly shorter computational time than with $c_T = 10^{-1}$ since in the computation with $c_T = 10$ we do not care about the accuracy with respect to the time. Moreover, if the accuracy with respect to the time is unimportant then it is sufficient to use the one-step BDF method (the backward Euler method).

Now, we present the performance of the iterative Newton-like method (31) – (32) with the stopping criterion (67) where we set $c_A := 0.1$ and the adaptive choice of the time step given following from condition (68) where we set $c_T := 10$.

Table 5 shows the results for the far field Mach numbers $M_\infty = 10^{-1}$, 10^{-2}, 10^{-3} and 10^{-4}, with the aid of the $P_p,\ p = 1, \ldots, 4$ polynomial approximation and the one step BDF method, namely the ratios $(\mathrm{p}_{\max} - \mathrm{p}_{\min})/\mathrm{p}_{\max}$, $(\rho_{\max} - \rho_{\min})/\rho_{\max}$, the drag coefficient c_D and the lift coefficient c_L. The pressure as well as the density fluctuations are of order M_∞^2 which is in agreement with the theoretical results in the incompressible limit. A small violation of the order of the density fluctuations is observed for $M_\infty = 10^{-4}$. This and the degradation of c_D and c_L values for $M_\infty = 10^{-4}$ are caused most likely due to not using a preconditioned numerical flux, see [5].

Moreover, Table 5 (the last three columns) shows the number of the time steps necessary to reach the steady state solution, the total number of the Newton iterations and the total number of the GMRES iterations in all time steps. We observe that these values do not change dramatically neither for increasing degree of approximation nor for the decreasing inlet Mach number.

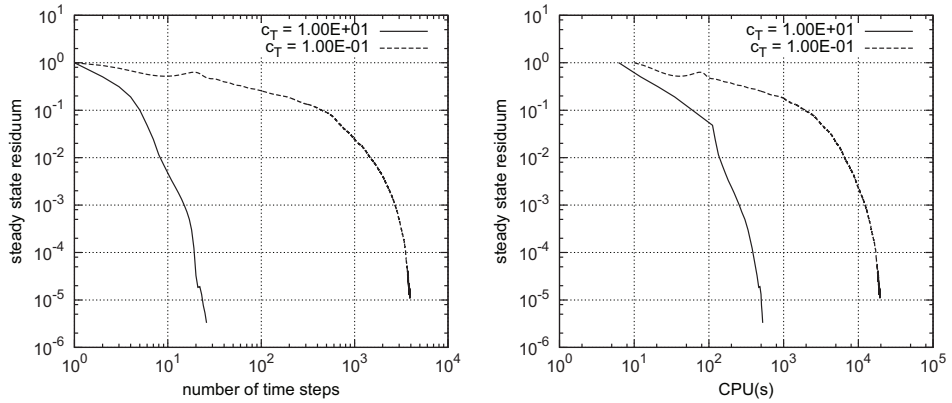

Fig. 4: Low Mach number flow with $M = 0.1$ for P_3 approximation, 3-steps BDF and with $c_T = 10$ and $c_T = 10^{-1}$, steady-state residuum with respect to the number of time steps (left) and with respect to the computational time (right).

M_∞	P_k	$\frac{\mathrm{p}_{\max}-\mathrm{p}_{\min}}{\mathrm{p}_{\max}}$	$\frac{\rho_{\max}-\rho_{\min}}{\rho_{\max}}$	c_D	c_L	time steps	Newton iters	GMRES iters
1.00E-01	1	9.89E-03	7.08E-03	3.53E-04	1.44E-03	25	50	478
1.00E-01	2	9.87E-03	7.09E-03	1.03E-04	1.04E-03	26	67	506
1.00E-01	3	9.87E-03	7.06E-03	5.35E-05	7.66E-04	31	95	583
1.00E-01	4	9.86E-03	7.06E-03	3.21E-05	6.55E-04	34	122	719
1.00E-02	1	9.93E-05	7.10E-05	3.63E-04	1.99E-03	24	32	630
1.00E-02	2	9.91E-05	7.11E-05	8.23E-05	1.40E-03	29	50	914
1.00E-02	3	9.90E-05	7.07E-05	5.13E-05	9.55E-04	35	79	955
1.00E-02	4	9.90E-05	7.07E-05	2.18E-05	8.56E-04	38	83	912
1.00E-03	1	9.92E-07	7.35E-07	3.50E-04	1.76E-03	26	30	1147
1.00E-03	2	9.92E-07	7.26E-07	6.68E-05	7.19E-05	27	33	1162
1.00E-03	3	9.90E-07	7.58E-07	4.55E-05	5.37E-04	37	64	1948
1.00E-03	4	9.91E-07	9.80E-07	9.67E-05	3.50E-04	37	65	1778
1.00E-04	1	9.91E-09	4.33E-08	2.56E-04	3.09E-04	19	19	905
1.00E-04	2	9.90E-09	2.04E-08	-6.60E-05	5.48E-04	24	24	999
1.00E-04	3	9.90E-09	2.51E-08	-9.56E-06	5.02E-04	33	35	1484
1.00E-04	4	9.90E-09	9.69E-08	-2.03E-05	5.11E-04	35	36	1440

Tab. 5: Low Mach number flows for far field Mach number $M_\infty = 10^{-1}$, 10^{-2}, 10^{-3} and 10^{-4}, with the aid of $P_p,\ p = 1, \ldots, 4$ polynomial approximation: ratios $(\mathrm{p}_{\max} - \mathrm{p}_{\min})/\mathrm{p}_{\max}$, $(\rho_{\max} - \rho_{\min})/\rho_{\max}$, drag coefficient c_D, lift coefficient c_L, the number of time steps and the total number of the Newton and GMRES iterations.

P_k	n	matrices vectors	linear systems	reconstructions $\boldsymbol{w}_{h\tau}$, $\bar{\boldsymbol{w}}_{h\tau}$	constrained extrema	others	total
1	1	16.3 %	56.7 %	16.0 %	5.3 %	5.8 %	100.0%
1	2	13.2 %	61.8 %	16.5 %	4.4 %	4.1 %	100.0%
1	3	11.5 %	60.7 %	19.5 %	4.1 %	4.2 %	100.0%
2	1	11.6 %	75.4 %	7.9 %	2.7 %	2.4 %	100.0%
2	2	12.9 %	68.0 %	12.4 %	3.3 %	3.3 %	100.0%
2	3	8.4 %	78.8 %	9.0 %	2.1 %	1.7 %	100.0%
3	1	11.5 %	79.3 %	5.2 %	2.4 %	1.6 %	100.0%
3	2	15.0 %	72.3 %	7.9 %	3.0 %	1.8 %	100.0%
3	3	8.5 %	81.9 %	6.4 %	1.9 %	1.4 %	100.0%
4	1	12.7 %	80.8 %	3.4 %	2.2 %	1.1 %	100.0%
4	2	12.2 %	79.9 %	4.5 %	2.3 %	1.1 %	100.0%
4	3	12.8 %	77.3 %	6.1 %	2.5 %	1.3 %	100.0%

Tab. 6: Relative computational costs for each individual part of the n-step BDF-DGFE algorithm.

Furthermore, Figures 5 and 6 demonstrate the performance of the iterative Newton-like method (31) – (32) for two selected cases ($M_\infty = 10^{-2}$ and $M_\infty = 10^{-4}$ using the P_4 polynomial approximation). Namely, Figures 5 – 6 (top) show the size of the algebraic, space and time residual error estimators η_{A}^k, η_{S}^k and η_{T}^k at time level $k = 1, \ldots, r$ and the corresponding size of the CFL_k number which reflect the size of the time step τ_k. The CFL_k number is defined as the ratio of the actual time step τ_k and the time step size following from the stability condition of an "explicit" time discretization, namely

$$\mathrm{CFL}_k := \frac{\tau_k}{\max_{K\in\mathscr{T}_h}\left(|K|^{-1}\max_{\Gamma\in\partial K}\lambda(\boldsymbol{w}_h^k|_\Gamma)|\Gamma|\right)},$$

where $\lambda(\boldsymbol{w}_h^k|_\Gamma)$ is the spectral radius of the matrix $\boldsymbol{P}(\boldsymbol{w}_h^k|_\Gamma, \boldsymbol{n}_\Gamma)$ given by (3). We observe the balance between the error estimators η_{A}^k, η_{T}^k and η_{S}^k at each time level. The size of the time step is adaptively increased up to very large values when the computation approaches to the steady-state.

Moreover, Figures 5 – 6 (centre) show the typical behaviour of the convergence of the nonlinear algebraic residuum $\|\boldsymbol{F}_h(\boldsymbol{\xi}_k^n)\|_{\ell^2}$ in the Newton-like iterative process, where $\boldsymbol{F}_h(\boldsymbol{\xi}_k^n)$ is given by (23). We found that only few Newton iterations are carried out at each time step. Moreover, the marked dots denote updates of the flux matrix $\mathbb{C}_h(\boldsymbol{\xi}_k^n)$ in the corresponding Newton step discussed in Section 4.3.2. We observe that only a few updates of $\mathbb{C}_h(\boldsymbol{\xi}_k^n)$ are carried out during the whole computation. Furthermore, Figures 5 – 6 (bottom) show the corresponding numbers of Newton iterations and the numbers of GMRES iterations at each time step for this example.

Finally, Table 6 (similarly as Table 4) shows the relative computational times for the individual parts of the code for $M_\infty = 0.01$, P_k, $k = 1, \ldots, 4$ polynomial approximation and n-step BDF, $n = 1, 2, 3$. We simply observe that the computation of the constrained extrema does not essentially increase the total computational time and its relative computational costs are lower for the higher degrees of polynomial approximation.

6.3 Viscous subsonic flow

Similarly as in [23, 35, 17], we consider a laminar viscous subsonic flow around the NACA 0012 profile with inlet Mach number $M_{\mathrm{inlet}} = 0.5$, angle of attack $\alpha = 2^\circ$ and Reynolds number $\mathrm{Re} = 5\,000$. In [23, 17], we presented steady-state solutions for this flow regime with several degrees of polynomial approximation with several grids. We employ an unstructured grid similar to the grid from Figure 3. Similarly as in Section 6.2, we carry out computations with the P_4 polynomial approximation, 3-steps BDF and with $c_T = 10$ and $c_T = 10^{-2}$ in (68). Figure 7 shows the convergence of the steady-state residuum and the corresponding value CFL_k for both settings $c_T = 10$ and $c_T = 10^{-2}$. Obviously, for $c_T = 10$ we obtain the steady-state solution. On the other hand, for $c_T = 10^{-2}$, where we have a much higher resolution in time, we obtain a non-steady solution. Moreover, Figure 8 shows the dependence of the lift coefficient c_L of the dimensionless physical time for P_3 and P_4 polynomial approximation with the 3-steps BDF-DGFE method with $c_T = 10^{-2}$ in (68). The constant value c_L-"steady" was obtained with the same method but with $c_T = 10$. Finally, Figure 9 shows the isolines of the Mach number for both P_3 and P_4 polynomial approximations and both $c_T = 10$ and $c_T = 10^{-2}$.

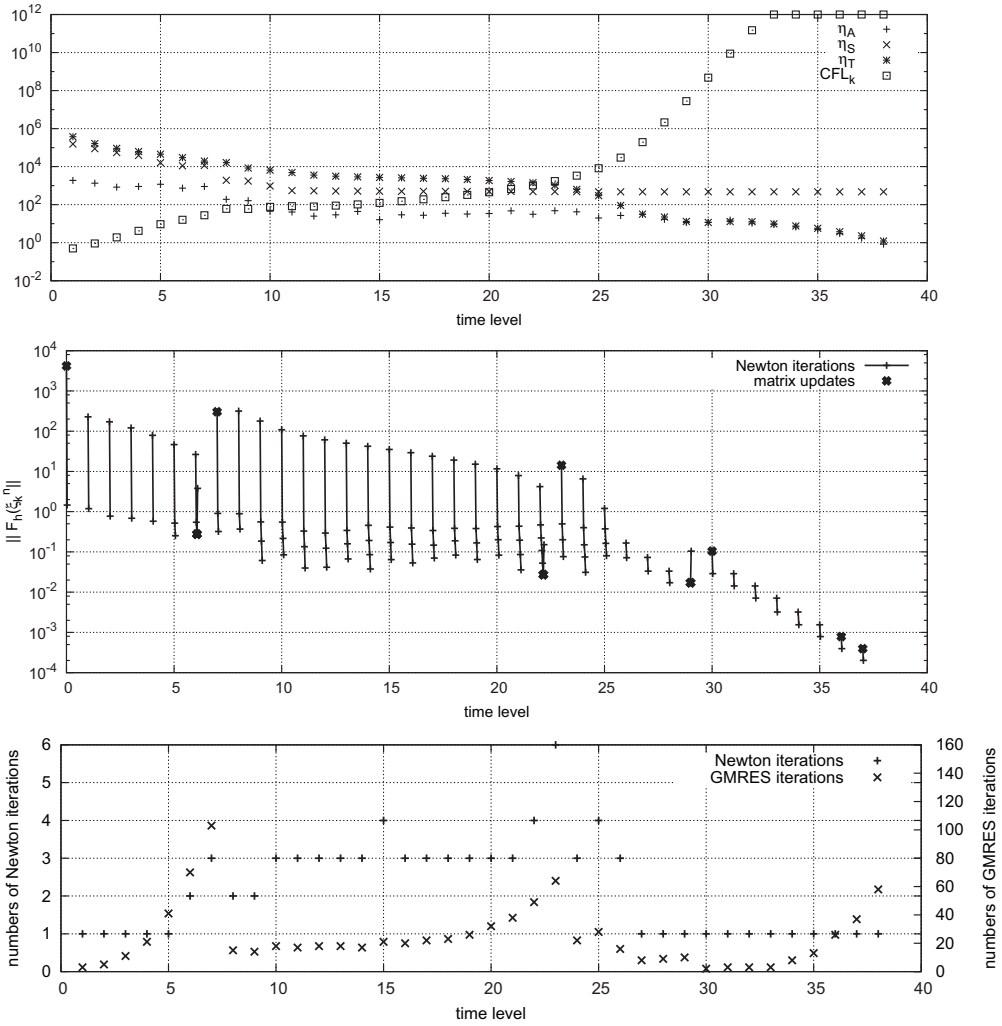

Fig. 5: Low Mach number flow with $M = 0.01$ for the P_4 approximation and the 1-step BDF: size of the algebraic, space and time residual error estimators η_{A}^k, η_{S}^k, η_{T}^k and the CFL_k number at time level $k = 1, \ldots, r$ (top), the convergence of the nonlinear algebraic residuum $\|\boldsymbol{F}_h(\boldsymbol{\xi}_k^n)\|_{\ell^2}$ in the Newton-like iterative process at each time level, marked dots denote updates of the flux matrix $\mathbb{C}_h(\boldsymbol{\xi}_k^n)$ in the corresponding Newton step (centre) and the number of Newton iterations and GMRES iterations at each time level (bottom).

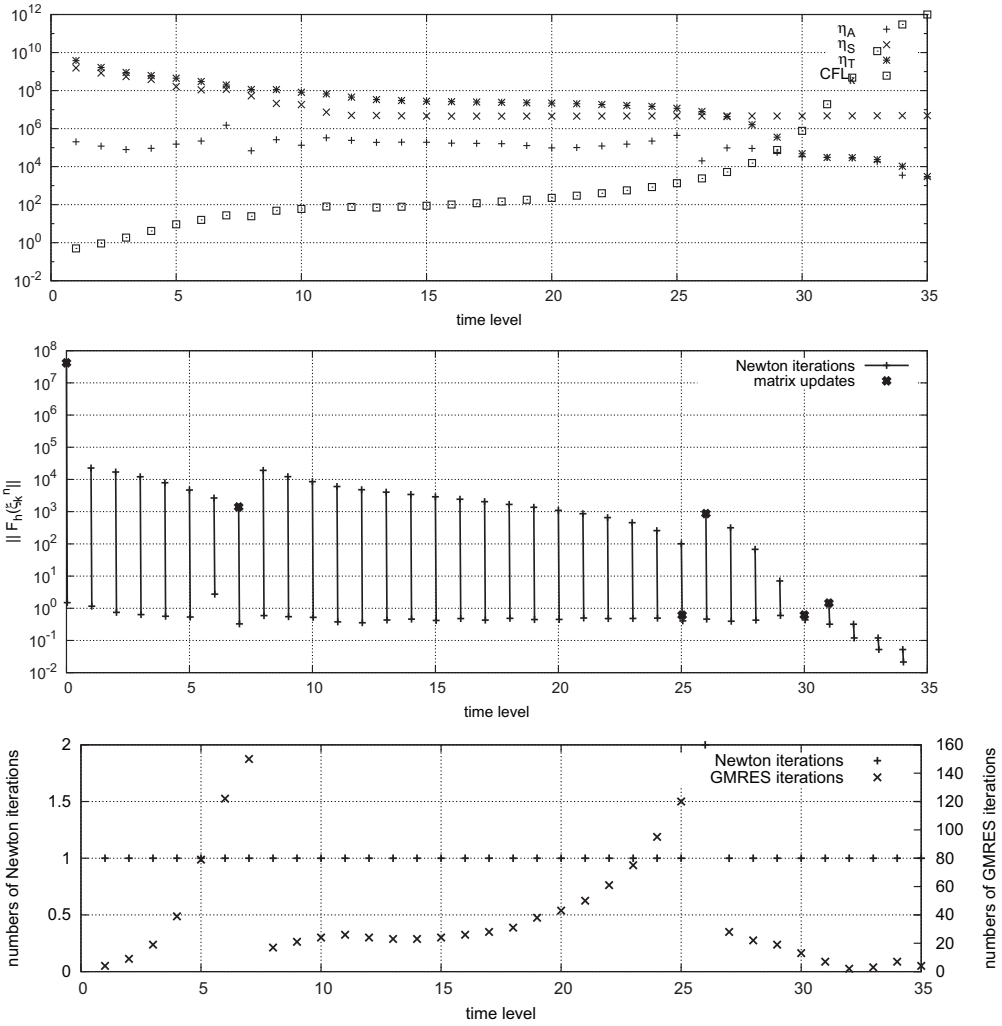

Fig. 6: Low Mach number flow with $M = 0.0001$ for the P_4 approximation and the 1-step BDF: size of the algebraic, space and time residual error estimators η_{A}^k, η_{S}^k, η_{T}^k and the CFL_k number at time level $k = 1, \ldots, r$ (top), the convergence of the nonlinear algebraic residuum $\|\boldsymbol{F}_h(\boldsymbol{\xi}_k^n)\|_{\ell^2}$ in the Newton-like iterative process at each time level, marked dots denote updates of the flux matrix $\mathbb{C}_h(\boldsymbol{\xi}_k^n)$ in the corresponding Newton step (centre) and the number of Newton iterations and GMRES iterations at each time level (bottom).

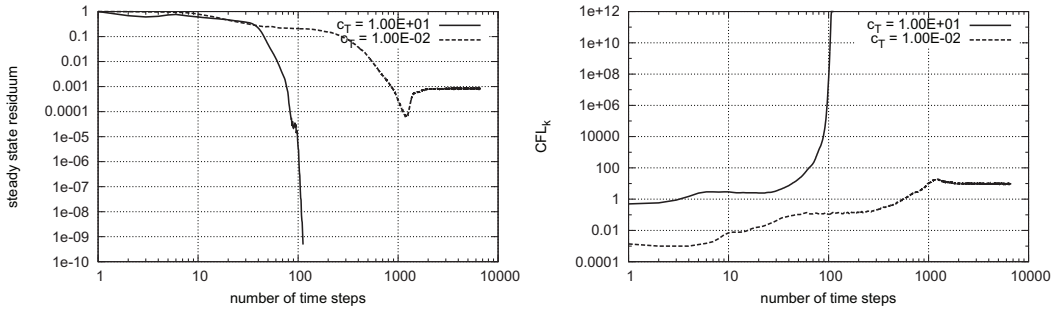

Fig. 7: Viscous subsonic flow for P_4 approximation, $c_T = 10$ and $c_T = 10^{-2}$, steady-state residuum (left) and the value CFL_k (right) with respect to the number of time steps.

These experiments indicate that an insufficiently accurate resolution with respect to the time may lead a different flow regime (steady-flow vs. non-steady). These results are in agreement with [35], where this example was solved by several research groups. They achieved mostly the steady state solution using steady-state solvers or implicit time discretizations (with large time steps). Only a sufficiently accurate (explicit) time discretization (carried out by University of Stuttgart) gave the unsteady flow regime, see [35, Chapter 5]. We suppose that the same observation can be obtained by the (standard) adaptive time methods based on estimates of the local discretization error of the resulting ODE systems. However, the tolerance for the local discretization error depends on the user and the tolerance, which is able to capture the unsteadiness of this flow regime and which is not too small, has to be found empirically. On the other hand, in our approach, we simply set the tolerance from the comparison of the space and time error estimates.

6.4 Viscous shock-vortex interaction

This example represents a challenging unsteady viscous flow simulation. Similarly as in [14, 44, 27], we consider the viscous interaction of a plane weak shock with a single isentropic vortex. During the interaction, acoustic waves are produced, and we investigate the ability of the numerical scheme to capture these waves. The computational domain is $\Omega = (0,2) \times (0,2)$ with the periodic extension in the x_2-direction. A stationary plane shock is located at $x_1 = 1$. The prescribed pressure jump through the shock is $\mathtt{p}_R - \mathtt{p}_L = 0.4$, where $\mathtt{p}_L$ and $\mathtt{p}_R$ are the pressures from the left and the right of the shock wave, respectively, corresponding to the inlet (left) Mach number $M_L = 1.1588$. The reference density and velocity are those of the free uniform flow at infinity. Particularly, we define the initial density, x_1-component of velocity and pressure by

$$\rho_L = 1,\ u_L = M_L \gamma^{1/2},\ \mathtt{p}_L = 1, \quad \rho_R = \rho_L K_1,\ u_R = u_L K_1^{-1},\ \mathtt{p}_R = \mathtt{p}_1 K_2, \tag{76}$$

where

$$K_1 = \frac{\gamma+1}{2} \frac{M_L^2}{1 + \frac{\gamma-1}{2} M_L^2}, \quad K_2 = \frac{2}{\gamma+1} \left(\gamma M_L^2 - \frac{\gamma-1}{2} \right). \tag{77}$$

Here, the subscripts $_L$ and $_R$ denote the quantities at $x < 1$ and $x > 1$, respectively, $\kappa = 1.4$ is the Poisson constant. The Reynolds number is 2000. An isolated isentropic vortex centred at $[0.5, 1]$ is added to the base flow. The tangential velocity in the vortex is given by

$$v_\theta = c_1 r \exp(-c_2 r^2), \quad c_1 = u_c / r_c, \quad c_2 = r_c^{-2}/2, \quad r = ((x_1 - 0.5)^2 - (x_2 - 1)^2)^{1/2}, \tag{78}$$

where we set $r_c = 0.075$ and $u_c = 0.5$. The computations are stopped at the dimensionless time $T = 0.7$.

We apply the space-time adaptive BDF-DGFE method from Section 5.7. The initial coarse grid has 192 element and it was priori refined in the vicinity of the stationary shock wave. We employ the 2-steps BDF-DGFE method with the P_4 polynomial approximation in space since it is more stable than the 3-steps BDF method in situation when the size of the time step has to be decreased after mesh refinement. In the space-time adaptive algorithm we set $c_A = 10^{-2}$ in (67), $c_T = 0.1$ in (68) and $\omega = 0.08$ in (70). Figure 10 shows the mesh adaptively refined in the first step. This figure shows also the initial setting of the shock and the isentropic vortex with their details.

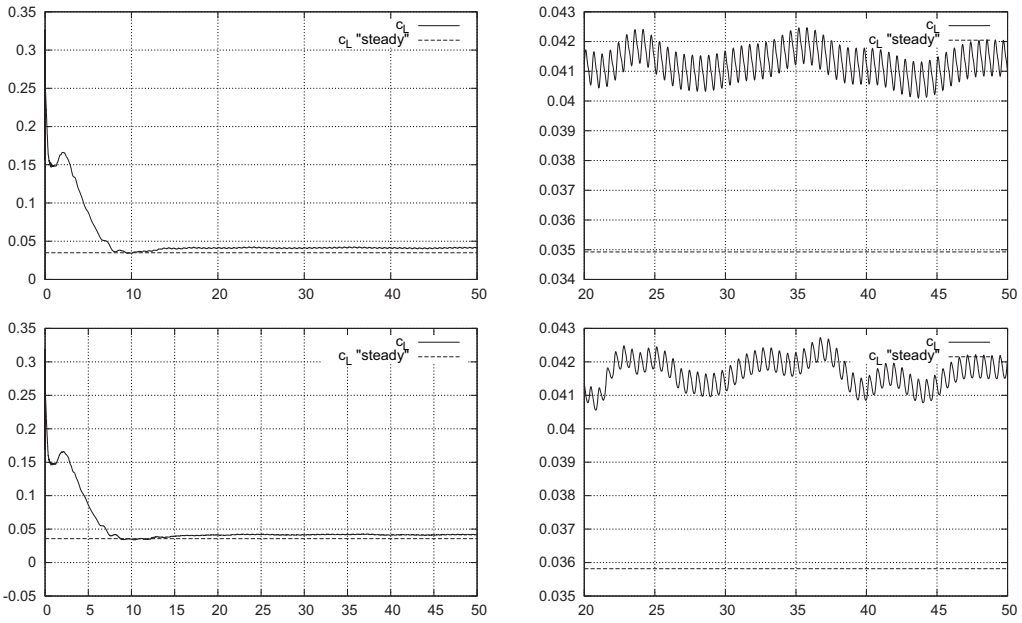

Fig. 8: Viscous subsonic flow for P_3 (top) and P_4 (bottom) approximation, time evolution of the lift coefficient c_L with respect the physical time for the setting $c_T = 10^{-2}$ (left) and its detail (right), the value c_L "steady" was obtain with $c_T = 10$.

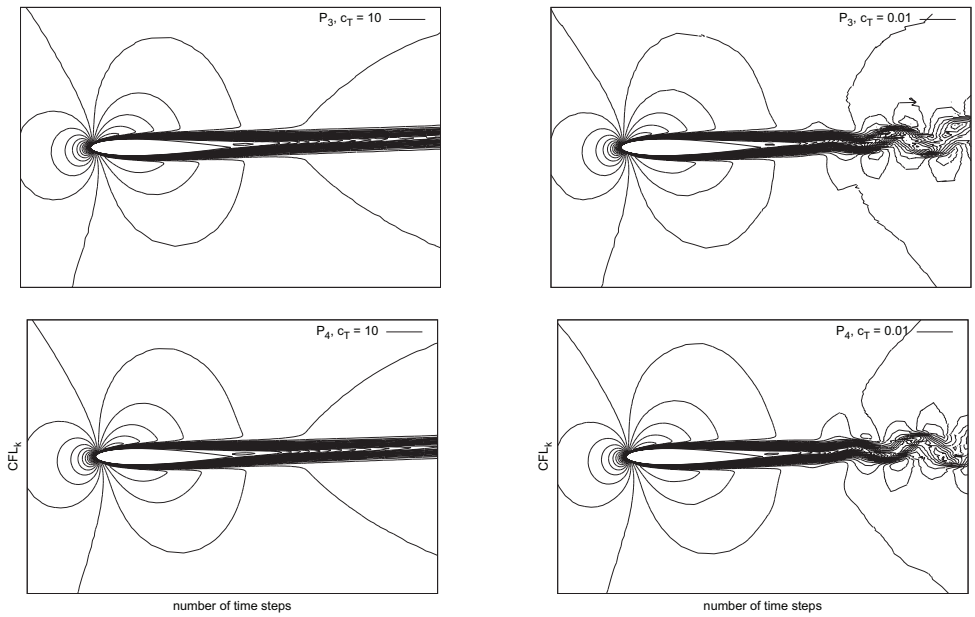

Fig. 9: Viscous subsonic flow for P_3 and P_4 polynomial approximations, isolines of the Mach number for the setting $c_T = 10$ and $c_T = 10^{-2}$.

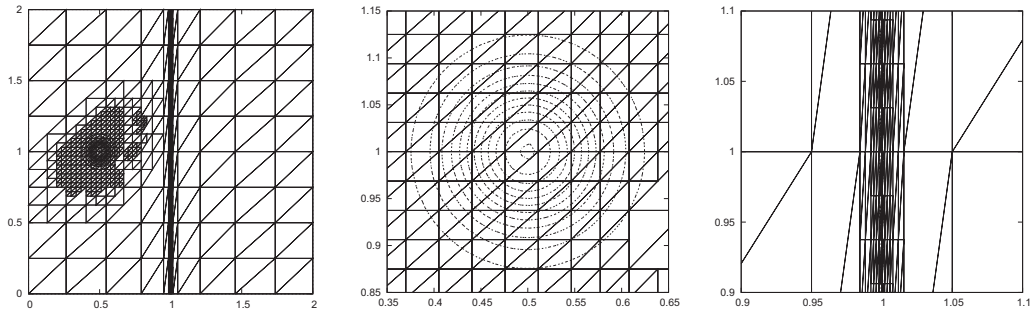

Fig. 10: Viscous shock-vortex interaction: the initial adaptively refined grid and the pressure isolines at $t = 0$, the total view (left) and its details near the vortex (centre) and the shock (right).

Figures 11 and 12 show the results of the simulation of the viscous shock-vortex interaction, namely adapted meshes with the corresponding isolines of the pressure and the pressure distribution along $x_2 = 1$ at several time instants. We observe a nice capturing of the interaction with the appearance of the incident and the reflected acoustic waves. Moreover, we observe that there is only a small refinement in the vicinity of the vortex. It is due to the fact, that the P_4 polynomial approximation is already very accurate and hence the h-adaptation is not further required.

7 Conclusion

We presented a higher order numerical method for the solution of the non-stationary compressible Navier-Stokes equations, which is based on the discontinuous Galerkin discretization and the high backward difference formulae. We developed the residual error estimates technique, which is based on the approximation of the error measured in the dual norm. This approach is able to identify the several ingredients of the total error, namely its algebraic, spatial and temporal parts. Based on them we defined an adaptive algorithm which

i) solves the corresponding algebraic systems untill the algebraic error estimate does not influence the spatial error estimate,

ii) chooses the time step such that the temporal error estimate is controlled by the spatial error estimate,

iii) adapts the mesh in such a way that space-time error estimate is under a given tolerance.

Several numerical experiments confirmed these items and demonstrated the ability of the presented method.

However, there are several open questions, namely a theoretical background of this approach. Another challenge is a use of these residual error estimates for an (possible anisotropic) hp-adaptive method.

References

[1] G. Akrivis, C. Makridakis, and R. H. Nochetto. A posteriori error estimates for the Crank-Nicolson method for parabolic equations. *Math. Comp.*, 75(254):511–531, 2006.

[2] Georgios Akrivis, Charalambos Makridakis, and Ricardo H. Nochetto. Galerkin and Runge-Kutta methods: unified formulation, a posteriori error estimates and nodal superconvergence. *Numer. Math.*, 118(3):429–456, 2011.

[3] L. El Alaoui, A. Ern, and M. Vohralík. Guaranteed and robust a posteriori error estimates and balancing discretization and linearization errors for monotone nonlinear problems. *Comput. Methods Appl. Mech. Engrg*, 200:2782–2795, 2011.

[4] T. J. Barth. Space-time error representation and estimation in Navier-Stokes calculations. In *Complex Effects in Large Eddy Simulations*, volume 56 of *Lecture Notes in Computational Science and Engineering*, pages 29–48. Springer-Verlag, 2007.

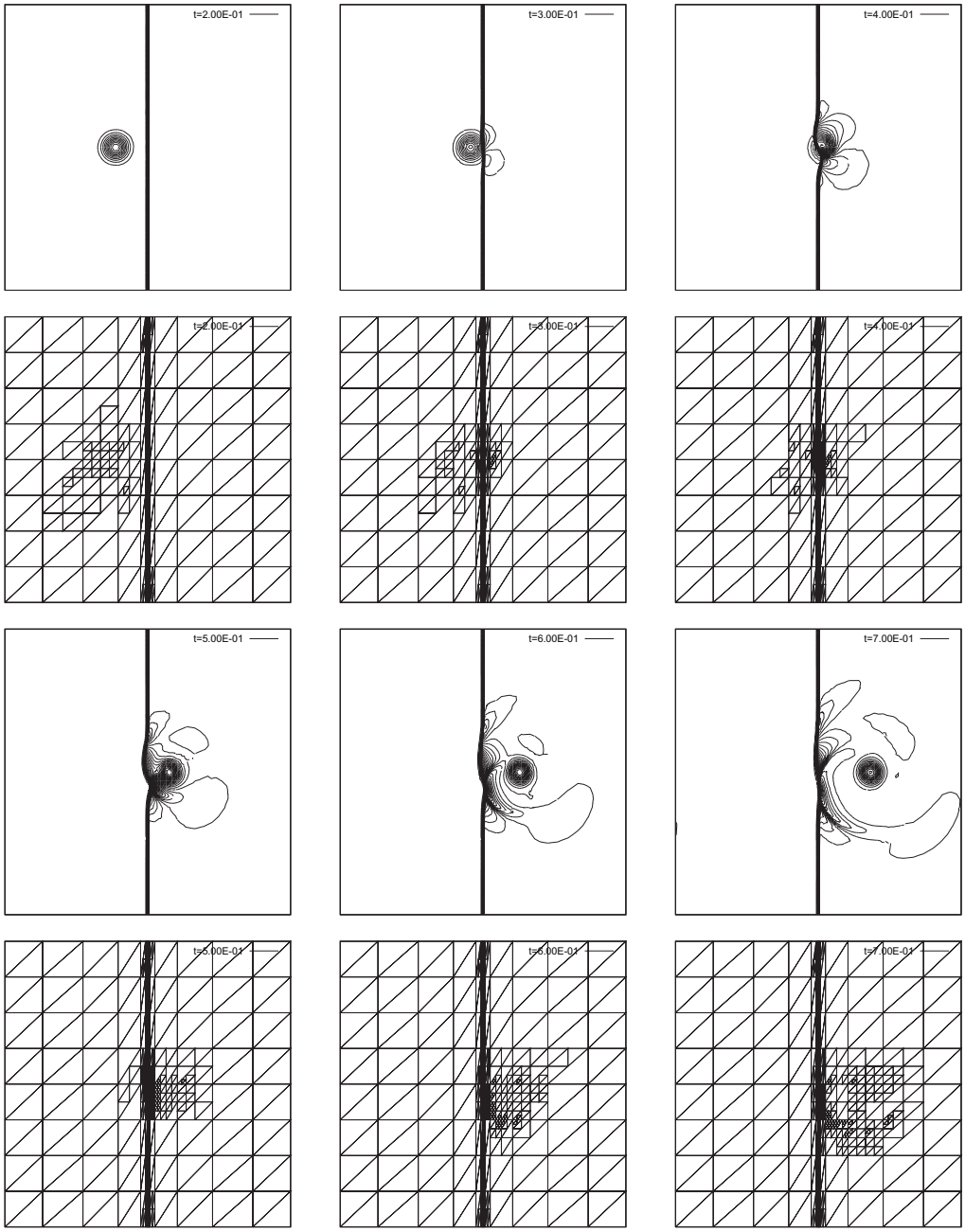

Fig. 11: Viscous shock-vortex interaction: isolines of the pressure and the adapted meshes at $t = 0.2,\ 0.3,\ 0.4,\ 0.5,\ 0.6$ and 0.7.

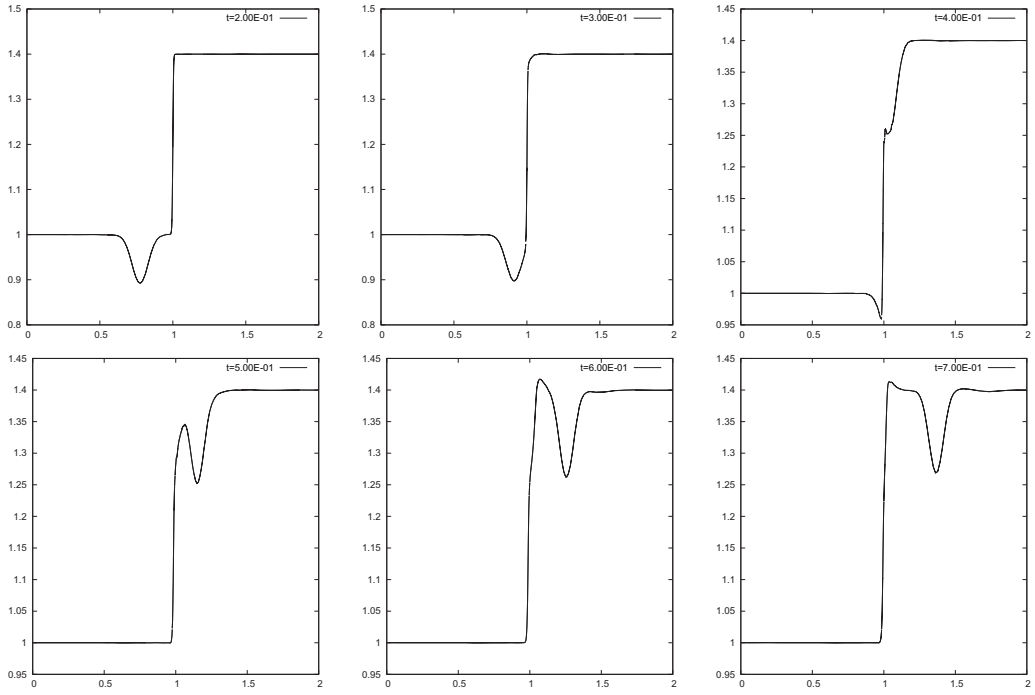

Fig. 12: Viscous shock-vortex interaction: isolines of the pressure at $t = 0.2,\ 0.3,\ 0.4,\ 0.5,\ 0.6$ and 0.7.

[5] F. Bassi, C. De Bartolo, R. Hartmann, and A. Nigro. A discontinuous Galerkin method for inviscid low Mach number flows. *J. Comput. Phys.*, 228:3996–4011, 2009.

[6] F. Bassi, L. Botti, A. Colombo, and S. Rebay. Agglomeration based discontinuous Galerkin discretization of the Euler and Navier-stokes equations. *Computers and Fluids*, 61:77–85, 2012.

[7] F. Bassi and S. Rebay. A high-order accurate discontinuous finite element method for the numerical solution of the compressible Navier–Stokes equations. *J. Comput. Phys.*, 131:267–279, 1997.

[8] F. Bassi and S. Rebay. A high order discontinuous Galerkin method for compressible turbulent flow. In B. Cockburn, G. E. Karniadakis, and C.-W. Shu, editors, *Discontinuous Galerkin Method: Theory, Computations and Applications*, Lecture Notes in Computational Science and Engineering 11, pages 113–123. Springer-Verlag, 2000.

[9] C. E. Baumann and J. T. Oden. A discontinuous *hp* finite element method for the Euler and Navier-Stokes equations. *Int. J. Numer. Methods Fluids*, 31(1):79–95, 1999.

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

[11] E. Burman and A. Ern. Discontinuous Galerkin approximation with discrete variational principle for the nonlinear Laplacian. *Comptes Rendus Mathematique*, 346(17-18):1013–1016, 2008.

[12] J. Česenek, M. Feistauer, J. Horáček, V. Kučera, and J. Prokopová. Simulation of compressible viscous flow in time-dependent domains. *Applied Mathematics and Computation*, 2011. Article in Press.

[13] A. Chaillou and M. Suri. A posteriori estimation of the linearization error for strongly monotone nonlinear operators. *J. Comput. Appl. Math.*, 205(1):72–87, 2007.

[14] V. Daru and C. Tenaud. High order one-step monotonicity-preserving schemes for unsteady compressible flow calculations. *J. Comput. Phys.*, 193(2):563–594, 2004.

[15] P. Deuflhard. *Newton Methods for Nonlinear Problems*, volume 35 of *Springer Series in Computational Mathematics*. Springer, 2004.

[16] V. Dolejší. On the discontinuous Galerkin method for the numerical solution of the Navier–Stokes equations. *Int. J. Numer. Methods Fluids*, 45:1083–1106, 2004.

[17] V. Dolejší. Semi-implicit interior penalty discontinuous Galerkin methods for viscous compressible flows. *Commun. Comput. Phys.*, 4(2):231–274, 2008.

[18] V. Dolejší. *hp*-DGFEM for nonlinear convection-diffusion problems. *Math. Comput. Simul.*, 87:87–118, 2013.

[19] V. Dolejší, A. Ern, and M. Vohralík. A framework for robust a posteriori error control in unsteady nonlinear advection-diffusion problems. *SIAM J. Numer. Anal.*, 51(2):773–793, 2013.

[20] V. Dolejší and P. Kůs. Adaptive backward difference formula – discontinuous Galerkin finite element method for the solution of conservation laws. *Int. J. Numer. Methods Eng.*, 73(12):1739–1766, 2008.

[21] V. Dolejší. Discontinuous Galerkin method for the numerical simulation of unsteady compressible flow. *WSEAS Transactions on Systems*, 5(5):1083–1090, 2006.

[22] V. Dolejší and M. Feistauer. Semi-implicit discontinuous Galerkin finite element method for the numerical solution of inviscid compressible flow. *J. Comput. Phys.*, 198(2):727–746, 2004.

[23] V. Dolejší, M. Holík, and J. Hozman. Efficient solution strategy for the semi-implicit discontinuous Galerkin discretization of the Navier-Stokes equations. *J. Comput. Phys.*, 230:41764200, 2011.

[24] A. Ern and I. Mozolevski. Discontinuous Galerkin method for two-component liquid-gas porous media flows. *Computational Geosciences*, 16(3):677–690, 2012.

[25] M. Feistauer, J. Felcman, and I. Straškraba. *Mathematical and Computational Methods for Compressible Flow*. Oxford University Press, Oxford, 2003.

[26] M. Feistauer and V. Kučera. On a robust discontinuous Galerkin technique for the solution of compressible flow. *J. Comput. Phys.*, 224(1):208–221, 2007.

[27] J. Fürst. *Modélisation numérique d'écoulements transsoniques avec des schémas TVD et ENO*. PhD thesis, Universite Mediterranee, Marseille and Czech Technical University Prague, 2001.

[28] S. Giani and P. Houston. Anisotropic *hp*-adaptive discontinuous Galerkin finite element methods for compressible fluid flows. *Int. J. Numer. Anal. Model.*, 9(4):928–949, 2012.

[29] M. Giles and E. Süli. Adjoint methods for PDEs: a posteriori error analysis and postprocessing by duality. *Acta Numerica*, 11:145–236, 2002.

[30] E. Hairer, S. P. Norsett, and G. Wanner. *Solving ordinary differential equations I, Nonstiff problems*. Number 8 in Springer Series in Computational Mathematics. Springer Verlag, 2000.

[31] E. Hairer and G. Wanner. *Solving ordinary differential equations II, Stiff and differential-algebraic problems*. Springer Verlag, 2002.

[32] R. Hartmann and P. Houston. Symmetric interior penalty DG methods for the compressible Navier-Stokes equations I: Method formulation. *Int. J. Numer. Anal. Model.*, 1:1–20, 2006.

[33] R. Hartmann and P. Houston. Symmetric interior penalty DG methods for the compressible Navier-Stokes equations II: Goal-oriented a posteriori error estimation. *Int. J. Numer. Anal. Model.*, 3:141–162, 2006.

[34] F. Hindenlang, G. J. Gassner, C. Altmann, A. Beck, M. Staudenmaier, and C.-D. Munz. Explicit discontinuous Galerkin methods for unsteady problems. *Comput. Fluids*, 61:86–93, 2012.

[35] N. Kroll, H. Bieler, H. Deconinck, V. Couallier, H. van der Ven, and K. Sorensen, editors. *ADIGMA A European Initiative on the Development of Adaptive Higher-Order Variational Methods for Aerospace Applications*, volume 113 of *Notes on Numerical Fluid Mechanics and Multidisciplinary Design*. Springer Verlag, 2010.

[36] A. Kufner, O. John, and S. Fučík. *Function Spaces*. Academia, Prague, 1977.

[37] P. Ladevèze and N. Moës. A new a posteriori error estimation for nonlinear time-dependent finite element analysis. *Comput. Methods Appl. Mech. Engrg.*, 157(1-2):45–68, 1998.

[38] I. Lomtev, C. B.Quillen, and G. E. Karniadakis. Spectral/*hp* methods for viscous compressible flows on unstructured 2d meshes. *J. Comput. Phys.*, 144(2):325–357, 1998.

[39] Charalambos Makridakis and Ricardo H. Nochetto. A posteriori error analysis for higher order dissipative methods for evolution problems. *Numer. Math.*, 104(4):489–514, 2006.

[40] J. Nečas. *Les Méthodes Directes en Thèorie des Equations Elliptiques*. Academia, Prague, 1967.

[41] R. H. Nochetto, A. Schmidt, and C. Verdi. A posteriori error estimation and adaptivity for degenerate parabolic problems. *Math. Comput.*, 69(229):1–24, 2000.

[42] Y. Saad and M. H. Schultz. GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM J. Sci. Stat. Comput.*, 7:856–869, 1986.

[43] C.W. Shu. Essentially non-oscillatory and weighted essentially non-oscillatory schemes for hyperbolic conservation laws. In A. Quarteroni et al, editor, *Advanced numerical approximation of nonlinear hyperbolic equations*, Lect. Notes Math. 1697, pages 325–432. Berlin: Springer, 1998.

[44] C. Tenaud, E. Garnier, and P. Sagaut. Evaluation of some high-order shock capturing schemes for direct numerical simulation of unsteady two-dimensional free flows. *Int. J. Numer. Meth. Fluids*, 126:202–228, 2000.

[45] R. Verfürth. A posteriori error estimates for nonlinear problems: $L^r(0,T;L^\rho(\Omega))$-error estimates for finite element discretizations of parabolic equations. *Math. Comput.*, 67(224):1335–1360, 1998.

[46] R. Verfürth. A posteriori error estimates for nonlinear problems: $L^r(0,T;W^{1,\rho}(\Omega))$-error estimates for finite element discretizations of parabolic equations. *Numer. Meth. Part. Diff. Eqs*, 14:487–518, 1998.

[47] G. Vijayasundaram. Transonic flow simulation using upstream centered scheme of Godunov type in finite elements. *J. Comput. Phys.*, 63:416–433, 1986.

[48] F. Vilar, P. Maire, and R. Abgrall. Cell-centered discontinuous Galerkin discretizations for two-dimensional scalar conservation laws on unstructured grids and for one-dimensional lagrangian hydrodynamics. *Computers and Fluids*, 46(1):498–504, 2011.

[49] M. Vohralík. *A posteriori error estimates, stopping criteria and inexpensive implementation*. Habilitation thesis, Université Pierre et Marie Curie – Paris 6, 2010.